Patent Value Models
Partial Least Squares Path Modelling with Mode C and Few Indicators

Alba Martínez Ruiz
Supervisor Dr. Tomás Aluja Banet

Department of Statistics and Operations Research
Technical University of Catalonia
Barcelona, Spain

A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor by the Technical University of Catalonia
Technical and Computer Applications of Statistics, Operational Research and Optimization Program

· December 2010 ·

Abstract

This thesis is two-fold. Firstly, investigating several model specifications, structural equation models of patent value are formulated. An initial definition is made involving model specification supported by a strong theory. Variations were aimed to study nonlinearities among constructs and the longitudinal nature of patent value. Secondly, robustness of Partial Least Squares (PLS) Path Modelling for estimating structural equation models is addressed and further developed. Three situations are investigated: PLS with Mode C, the case of few indicators per construct, and nonlinearities between constructs.

Keywords: Patent value, Partial Least Squares path modelling, structural equation models.

New tools have appeared, new areas of application have been explored.

Ludovic Lebart

Principle 4: The researcher must be aware and beware of all assumptions underlying a method of analysis, the mathematical consequences of these assumptions, and their relations to the hypotheses pursued, data collected, and outcomes of statistical modelling in order to perform a meaningful analysis.

Keith F. Widaman

Contents

List of Figures

List of Tables

Acknowledgments

This research was funded by grants from the Universidad Católica de la Ssma. Concepción (UCSC) and the Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), Chile. I am grateful for the support provided by these institutions. Additionally, I have received funding from the program Ajut per Estada de Curta Durada, of the Department of Statistics and Operations Research (DEIO) at the UPC as well as from the program Financiamiento Especial de Actividades Académicas at the UCSC, which allowed me to do the three-month research stay in 2008.

I am deeply grateful to my supervisor, Professor Tomás Aluja Banet, and I want to express my sincere gratitude to him. This thesis would not have been possible without him and his outstanding knowledge about statistics, models, and data. Professor Aluja introduced me to dimensionality reduction and multivariate data analysis, and throughout the various stages of this thesis he guided this research with unerring scientific intuition, patience and wisdom. He welcomed the patents and me with sincere generosity, showed me new procedures for analyzing data, funded the presentation of many of the papers presented at conferences, and made me embrace statistics as a discipline for life. I am grateful for the life opportunity acquired from having known and worked with him.

I also wish to thank Professor Samuel Kaski for allowing me the privilege to do a three-month research stay in his research group and to Professor Michael Thelwall for his trust and encouragement, and for asking me to talk about my work in his research group. That was also a privilege.

I am also very grateful to Professor Guadalupe Gómez who welcomed me in the DEIO; Òscar Carbó and Eva Villalon from the UPC Patents and Licenses Office for guiding us in protecting some results of this thesis; the administration staff: Toni Font, Marta Llabres, Carme Macias, Pia Margarit, and Celia Martin, because they always helped me with care and efficiency; and Matt Elmore for giving me advice on improving my scientific writing in English. I would also like to thank my colleagues and friends, especially Lesly Acosta, Nancy Jacome, Jeaneth López, Nuncia Medina, Cristina Montañola, Sonia Navarro, Jordi Cortes, Francisco López, and Gastón Sanchez, for helpful comments and all the unforgettable moments we shared, and all the people in the Department who in one way or another contributed to the successful completion of this thesis.

The roots of this thesis originate from many years ago, when I was in the Faculty of Engineering at the UCSC. The UCSC accepted the challenge of turning their students into technology entrepreneurs in 2000. The university not only believed that almost all of the students "live the experience" of Business Incubator and Entrepreneurship at the UCSC, but also that technology entrepreneurs should be trained in Academic Programs. In those years, I started to look for areas of technological development for the Faculty and the Incubator. But which technological areas were the most valuable and how could we identify them? Through the research project PID01/04, we began by analyzing the current state by characterizing the creation of advanced human capital and scientific production in Chile. Using the information contained in patent documents, publications and projects, we aimed to identify scientific communities, knowledge clusters, potential poles of technological developments and to develop strategies for action. At the same time, I implemented the first university courses in Entrepreneurial Technology Business, Intellectual Property and Innovation at the UCSC and possibly in the Bío Bío Region, and I proposed and supervised undergraduate final projects related to industrial activity, such as monitoring the technology of biotechnological applications for wine based on patent indicators, the characterization of patent applications of Chilean universities, technology transfer in the wood sector for the VIII Region of Chile, and so on. However, the concept of technological/patent value seemed very difficult to me to quantify. I hope that this thesis makes a small contribution in this regard and I would like to acknowledge and thank all who worked with me at the UCSC for helpful discussions in those years, especially Jorge Beyer, María Teresa Bull, Claudia Carrasco, Enrique Fernández, Jorge Galleguillos, Mauricio González, Mariella Gutiérrez, Ana Narvéz, Rodrigo Rebolledo and Marcos Vergara.

Finally, I want to make a special acknowledgment to my “close” family, my mother and my Chilean clan of 37 members, because everything I am I owe to them. This thesis is dedicated to them. Without their encouragement and support the development of this project simply would not have been possible. I am also very grateful to my friends, because they were like my family in Barcelona, and all the people I met during these years. Without them, this project would have been a doctoral thesis without the life experience that I have been fortunate to have.

List of Publications

Journal Papers

- Martinez-Ruiz, A., Aluja-Banet, T. (2010) Two-Step PLS Path Modelling Mode B: Nonlinear and Interaction Effects between Formative Constructs (submitted).

- Martinez-Ruiz, A., Aluja-Banet, T. (2010) PLS Path Modelling with Mode B and Mode C: Monte Carlo simulations and computational experiments (submitted).

- Martinez-Ruiz, A., Thelwall, M. (2010) The importance of technology and the visibility of the firms on the web: An exploratory study. Cybermetrics, 14(1):2.

- Martinez-Ruiz, A., Aluja-Banet, T (2009). Toward the definition of a structural equation model of patent value: PLS path modelling with formative constructs. REVSTAT Statistical Journal, 7(3):265-290.

Articles in Edited Books

- Martinez-Ruiz, A., Aluja-Banet, T. (2010) PLS Path Modelling with Mode C: Computational experiments. Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2010, WCE 2010, 30 June - 2 July, London, U.K., pp:1987–1992. Recommended for the Best Paper Award. ISBN: 978-988-18210-8-9, ISSN: 20780958 (print), ISSN: 2078-0966 (online).

- Martinez-Ruiz, A., Aluja-Banet, T. (2009) Two-step procedure for PLS path modelling with interaction effects between formative constructs and time-dependent variables. In: Esposito Vinzi, V., Tenenhaus, M. and Guan, R. (Eds.) Proceedings of the 6th International Conference on Partial Least Squares and Related Methods, Beijing: Publishing House of Electronics Industry, pp. 39-43. ISBN 978-7-121-09342-5.

- Martinez-Ruiz, A., Aluja-Banet, T. (2008) Second-order model of patent and market value. In: Brito, P. (Ed.) Proceedings on Computational Statistics, Heidelberg: Physica Verlag, pp. 1445-1452. ISBN: 978-3-7908-2083-6.

Conference Proceedings

- Martinez-Ruiz, A., Aluja-Banet, T. (2010) Longitudinal nature of patent value and technological usefulness: Exploring structural equation models. Patent Statistics for Decision Makers 2010, a conference organized by the European Patent Office and the Organisation for Economic Cooperation and Development, November 17-18.

- Martinez-Ruiz, A., Thelwall, M. (2009) The importance of technology and the visibility of the firms on the web, 5th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & Tenth COLLNET Meeting, September 13-16.

- Martinez-Ruiz, A., Aluja-Banet, T (2009) Predictive dynamic model of patent value, EURISBIS'09 European Regional Meeting of the International Society for Business and Industrial Statistics, May 30 - June 3.

- Martinez-Ruiz, A., Aluja-Banet, T. (2008) Predictive dynamic model of patent value for renewable energy. 25th DRUID Celebration Conference 2008 on Entrepreneurship and Innovation, June 17-20.

- Martinez-Ruiz, A., Aluja-Banet, T. (2007) Partial least squares path modelling in patent value analysis: A proposal, 5th International Symposium on PLS and Related Methods, September 5-7.

PhD Conferences

- Martinez-Ruiz, A., Aluja-Banet, T. (2008) Structural model of patent and market value: An application in energy patents. DRUID-DIME Academy Winter PhD Conference on Economics and Management of Innovation and Organizational Change, January 17-19.

- Martinez-Ruiz, A., Aluja-Banet, T. (2007) Partial least squares path modelling for technological information valuation. Doctoral Forum of 11th International Conference on Scientometrics and Informetrics, June 24-27.

Research Stay

- Three months research stay at Statistical Machine Learning and Bioinformatics Research Group at the Laboratory of Computer and Information Science at the Helsinki University of Technology, September-November, 2008.

Patents

- Multidimensional method and computer system for patent and technology rating and related database (US122205136, applied for in August 2008, published in March 4, 2010); Método multidimensional y sistema implementado por ordenador para clasificación de carteras de patentes y tecnologías y bases de datos asociadas (P200802549); Inventors: Alba Martinez-Ruiz, Tomás Aluja-Banet; Assignees: Universidad Católica de la Ssma. Concepción (Chile) and Technical University of Catalonia (Spain).

Chapter 1

Introduction

Abstract. This chapter presents an introduction to the thesis. The topics are on the general research approach –including objectives and research scope– and the document's structure.

1.1 General Research Approach

Within the field of technology management and technological change, determinants of patent value have aroused interest. Patents are an important intangible asset for companies. In the 90's, intangible assets represented three-quarters of the market value of the main companies of the world (Rivette & Kline, 1999). Intellectual property management, which is closely related to technology management, can help a company gain a competitive advantage, e.g. through technological leadership or by helping to form an industry standard.

Estimating ("predicting") patent value is a complicated issue. Patents can represent legal instruments, technologies, intangible assets, innovations, barriers to entry into a market, the main results of R&D activities, the driving force of technological change, and so forth. Thus, the problem can be approached from different perspectives. Interpretation is also affected by the interests of different units of companies, such as R&D, marketing or production units. Therefore, different value definitions are possible. Some approaches have been proposed by the scientific community, focusing on the private or social value of patents. For instance, Lanjouw *et al.* (1998, p. 407) defined the private value of a patent in terms of "the difference in the returns that would accrue to the innovation with and without patent protection." The magnitude of this difference would be crucial in applying or renewing the protection. In this thesis, we do not seek to determine the value of an individual patent or obtain a monetary value of the intangible assets. Rather, here patent value deals with the technological usefulness of the inventions for development of other inventions. Thus, we are interested in identifying and relating the variables that may determine the patent value in terms of technology.

Reitzig (2004a) has suggested modelling patent value as a construct, but little progress has been made along these lines and, to the best of our knowledge, a causal model with a latent variables approach for modelling patent value has not been addressed before in the literature. Hence, this thesis project seeks to propose structural and measurement models for patent valuation. This is exploratory and prediction-oriented research, where the prior theoretical knowledge is scarce. Therefore, PLS Path Modelling is used to explore and estimate the proposed models. Unobservable variables are sought and identified through a comprehensive literature review. Propositions are also made. Observable variables (or indicators) are mainly built from data contained in patent documents. Given the specific characteristics of patent indicators, such as high heterogeneity, asymmetry and large variances, a multivariate normality distribution assumption is not recommended. This also drives us to consider a component-based approach for structural equation models (SEMs) so as to analyze and compare different representations looking for the best patent value model. One of the main rewards of using a SEM approach is that we can compare the performance [value] of different companies or specific technological areas, and to systematize this comparison.

Structural equation models aim to explain relationships among different types of variables. Those that are directly measurable in an individual or object, even those that are unobservable, and represent an abstract conception of something. For this purpose, two main approaches have been developed. In 1970, Karl Jöreskog proposed the Linear Structural Relation Model (LISREL) procedure based on analysis of covariance. Later, in 1979, Herman Wold introduced Partial Least Squares (PLS) Path Modelling as an alternative to Jöreskog's approach, which makes strong assumptions about data distribution.

After having been forgotten for some years, in the middle eighties the PLS Path Modelling procedure was rediscovered by researchers such as Wynne Chin, Michel Tenenhaus and Vincenzo Esposito-Vinzi. Their contributions in conjunction with the conferences on PLS and Related Methods have promoted theoretical and practical research in this area, also supported by the recent availability of software including PLS procedures as a statistical solution. The main advances are related to two active research fronts: PLS Path Modelling applications for multi-group analysis –where several algorithms have been recently proposed; and PLS Path Modelling applications for investigating nonlinearities in inner models. Mainly using Monte Carlo simulations, research has also been concerned with the behavior of the PLS algorithm under certain specific conditions, such as multicollinearity, skewed distributions of observed variables and misspecification of models. These investigations have mostly studied SEMs with reflective outer models. Recently, some debate has arisen about the distinction between reflective and formative measurement models. Interestingly, it has been noted how many measurement models are theoretically far better supported by formative relationships, even though reflective outer models have been reported in the literature. According to the theory developed by economist of technological change and statistical evidence (which we provide below) in regards to patent value models, formative relationships suitably represent the links between manifest variables and some constructs. In a component-based approach, formative relationships are usually modeled with PLS Path Modelling with Mode B, and Mode A is used if the SEM includes reflective outer models. "The algorithm is called PLS Mode C if each of Modes A and B is chosen at least once in the model" (Wold, 1982, p. 10). Nevertheless, robustness and performance of PLS Path Modelling with Mode C has been studied little. Thus, contributions of this thesis are also deeply concerned with this topic. Additionally, nonlinear and interaction effects among formative constructs have not been investigated before. Hence, a Two-Step PLS Path Modelling procedure is implemented and guidelines are suggested in this regard. This procedure is also applied to the investigation of nonlinearities among formative constructs of patent value models.

In this thesis, contributions are made to the valuation of patents and to the understanding of PLS Path Modelling. Consequently, two general aims are posed:

- A To investigate causality relationships among variables that determine the patent value, considering data contained in patent documents.
- B To investigate the robustness and performance of PLS Path Modelling with Mode C for estimating structural equation models with formative and reflective outer models.

For each general objective, the following specific goals are established:

- A.1 To define a theoretical framework, primary contributions and advances in the field of patent valuation.
- A.2 To formulate a structural and measurement model for estimating the patent value.
- A.3 To estimate and to validate the proposed models using PLS Path Modelling and validation techniques.
- B.1 To define a theoretical framework, primary contributions and advances in the field of PLS Path Modelling.
- B.2 To determine the robustness and performance of PLS Path Modelling with Mode C for estimating SEMs with formative constructs and reflective latent variables.

B.3 To determine the robustness and performance of a Two-Step PLS Path Modelling with Mode C procedure for estimating nonlinear and interaction effects in SEMs with formative and reflective outer models.

Finally, data quality and choice of manifest variables are key factors in SEM research (Balachandra & Friar, 1997; Astebro, 2004). Consequently, data were selected and retrieved from patent databases to have useful data to seed the models. Building our own database, patent indicators were computed. The analyzed sample comprises a set of 2,901 patents in the renewable energy field.

1.2 Thesis Structure

This thesis is structured as follows:

- Chapter 2 gives an overview of the different approaches for patent value from the perspective of technological change. Definitions related to patent documents and patent indicators are provided.

- Chapter 3 reports on patent sample descriptions. We present criteria to retrieve data, the procedure for calculating patent indicators, and a statistical data description.

- Chapter 4 provides an introduction to structural equation models including origins, basic background and recent developments. In addition, it provides guidelines for model specification and modelling process for structural equation models. Special emphasis is placed on determining the reflective or formative nature of measurement models.

- Chapter 5 puts forward the main partial least squares algorithms: nonlinear estimation by iterative partial least squares (NIPALS), PLS regression, and PLS Path Modelling. We present two path modelling implementations: Lohmöller and Wold's procedures. Additionally, insights are given on procedure sensitivity to starting weight values and weighting schemes; algorithm properties, such as consistency and consistency at large; and convergence. We briefly review some PLS Path Modelling extensions and relationships with other procedures. The chapter ends by describing validation techniques.

- Chapter 6 provides evidence about the accuracy and precision of PLS Path Modelling with Mode C to recover true values in SEMs with few indicators per construct. Monte Carlo simulations and computational experiments are carried out to study the performance of the algorithm.

- Chapter 7 addresses the formulation and estimation of patent value models. This entails the identification and definition of observable and unobservable variables, the determination of blocks of manifest variables and structural relationships, the specification of a first- and a second-order model of patent value, and the models' estimation by PLS Path Modelling.

- In Chapter 8, the evolution of patent value over time using longitudinal structural equation models is investigated. Two set-ups are explored. The first longitudinal model includes time-dependent manifest variables and the second includes time-dependent unobservable variables. The structural equation models are estimated using PLS Path Modelling.

- In Chapter 9, there is a description of a Two-Step PLS Path Modelling (TsPLS) with Mode C procedure to study nonlinear and interaction effects among formative constructs. Monte Carlo simulations are performed to generate data and to determine the accuracy and precision of this approach to recover true values. This chapter includes an application of the TsPLS algorithm to patent value models.

- Finally, in Chapter 10, we provide a summary of conclusions, the author's contributions and future research.

Chapter 2

Patents as a Proxy for Technology

Abstract. This chapter gives a background on patents as a proxy for technologies in their early stage of development. That is, when they are created and protected. We begin by specifying what constitutes a patent and a patent document. A second goal is to define patent indicators and to describe their relationships with some constructs studied in the field of technological change. Some approaches for patents count and citations count as well as examples showing the role of patent indicators are also addressed. We examine different patent value concepts and the relationship between patent value, R&D and market value.

2.1 Introduction

We live in an era of rapid technological change. Every day new technologies appear and change the lives of millions. It was during the 80s and 90s through the globalization of markets, the arrival of information technologies, the Internet boom, and the transformation of society into a culture based on knowledge, that technologies took on a significant role. There is no doubt that technological development has been the promoter of long-term economic growth and has increased social welfare for years. Maybe, these are the main reasons why many researchers have focused on the study of all aspects related to this development. Technologies emerge, grow, mature and decline in direct relation to their environment. So, the value of technologies throughout their life cycle is influenced by the social, cultural and economic context in which they develop.

Studying the value of technologies is a complex issue, because research can be approached from different perspectives depending on how the problem is addressed. From an economic standpoint, the scientific community has studied issues related to innovation and its relationship to companies and countries. From a social science perspective, the scientific community has investigated the development of indicators through bibliometric studies and information retrieval techniques. The scientific community which concerns itself with technology management has emphasized technological forecasting, technology watch and data analysis to be competitive. In this research we focus on some developments made by these scientific communities; but our approach is mainly related to technological change and the development of indicators. We are interested in investigating the value of technologies in their early stages of development. That is, when they are newly invented and protected.

Patents are the results of innovation processes in an institution or company, and innovation processes are key to the economic growth and competitiveness of a country. Almost all the results of the R&D efforts of companies and institutions –those that are ultimately exploited– are protected by patents. But only some of these patents are truly valuable. We distinguish two approaches in relation to patent value. Firstly, patents can be valuable because they protect important technologies for the development of other technologies in the future; this is discussed in purely technological terms. Secondly, patents can be valuable because they yield substantial benefits to companies or society in general. We are interested in both types of values, but in this research the first approach is addressed.

Furthermore, there are two important aspects that need to be mentioned in relation to patents. (1) Patent documents are an important source of technological information and its use has many advantages. There is much information in a patent document, not only technical information, but also information about who, when and where technology is produced. There is a significant flow of patents every year and this information is available in databases. Even though there are still some difficulties in processing these large volumes of information, it is presumed to be consistent. Patent information has proved to be useful for selecting research and development (R&D) portfolios, beginning R&D or engineering projects, developing new products and new markets, the acquisition of intellectual property and the exploitation of intellectual assets, developing technological collaboration, evaluating organizational competence, forecasting opportunities and threats, planning technological strategies and making technological roadmapping. (2) Patents are related to a company's intangible assets management, i.e. patents, trademarks and copyrights. In the 90s, intangible assets represented three-quarters of the market value of the main companies of the world (Rivette & Kline, 1999). Intellectual property management, which is closely related to technology management, can help a company gain a competitive advantage, e.g. through technological leadership or by helping to form an industry standard (Reitzig, 2004b).

This chapter presents a background on patents because they are a proxy for new technologies. Besides the definition of patent and patent indicators, the structure of patent documents and the main source of information, this section introduces the relationship between patents and constructs, such as knowledge and value.

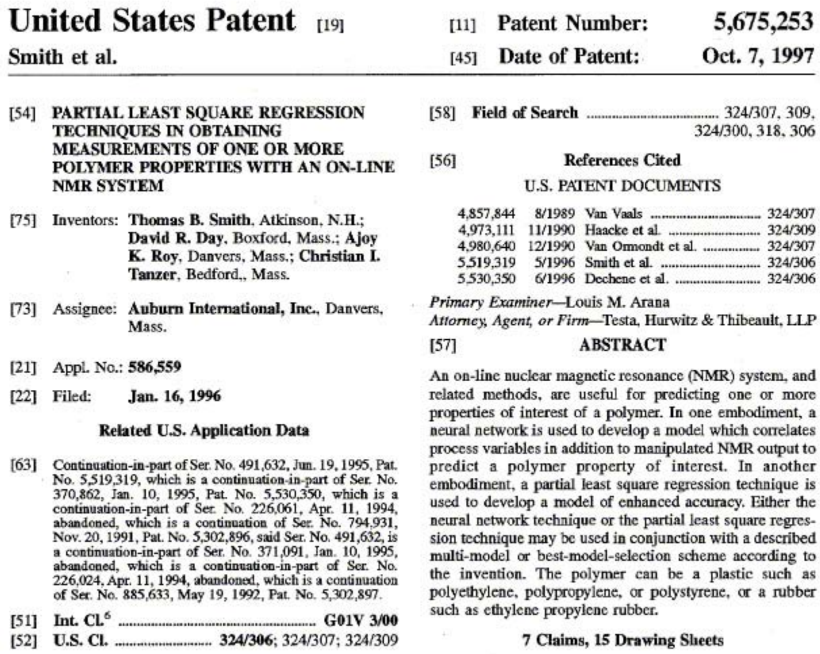
2.2 Defining a Patent Concept

A patent "is the right granted to an inventor by a state, or by a regional office acting for several states, which allows the inventor to exclude anyone else from commercially exploiting his invention for a limited period, generally 20 years" (WIPO, n.d.). A patent may protect a product or a process that is a new way of doing something. It is granted when the invention fulfils three basic requirements: the invention is new (novelty), involves an inventive activity and it is useful for industry. These conditions are called patentability conditions and they affect the value of patents (Nordhaus, 1967; Green & Scotchmer, 1997).

To patent an invention, the inventor must comprehensively describe the invention by delivering the technical details that must be known in order for it to be replicated. This description should be compared to the existing technological developments and it should provide details on the characteristics that are new in order to prove the novelty of the invention. In a formal way, novelty describes the technological distance between the patented invention and the previous state of the art. By the same token, the invention must show an inventive step "that could not be deduced by a person with average knowledge of the technical field" (WIPO, n.d.). The U.S Patent and Trademark Office talks about non-obviousness. It describes the technological distance between the invention patented and current technology in terms of evidence. Lastly, the industrial applicability is related with the utility of the invention. This must be of practical use or capable of some kind of industrial application.

On the other hand, an invention must be patentable according to the applicable law in the country where protection is sought. For example many countries exclude from the patent right: scientific theories, mathematical methods, variety of animals or plants, natural substances, methods for medical treatments and any invention where –in order to prevent its commercial exploitation– it is necessary to protect the public order, the moral or the public health. Generally, patents are first applied for in a national patent office. However, patents can also be applied for in a regional office according to a special treaty[1]. Interested readers can visit the website of the World Intellectual Property Organization (www.wipo.int),

[1]The main agreements are: the Patent Cooperation Treaty (1970), known as PCT; the Strasbourg Agreement Concerning the International Patent Classification (1971); the Microorganisms for the Purposes of Patent Procedure (1977); the Paris Convention for the Protection of Industrial Property (1883); the Berne Convention for the Protection of Literary and Artistic Works (1886); and the Budapest Treaty on the International Recognition of the Deposit of Patent Law Treaty (2000).



United States Patent [19]
Smith et al.

[11] **Patent Number:** **5,675,253**
[45] **Date of Patent:** **Oct. 7, 1997**

[54] **PARTIAL LEAST SQUARE REGRESSION TECHNIQUES IN OBTAINING MEASUREMENTS OF ONE OR MORE POLYMER PROPERTIES WITH AN ON-LINE NMR SYSTEM**

[75] Inventors: **Thomas B. Smith**, Atkinson, N.H.; **David R. Day**, Boxford, Mass.; **Ajoy K. Roy**, Danvers, Mass.; **Christian I. Tanzer**, Bedford,, Mass.

[73] Assignee: **Auburn International, Inc.**, Danvers, Mass.

[21] Appl. No.: **586,559**

[22] Filed: **Jan. 16, 1996**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 491,632, Jun. 19, 1995, Pat. No. 5,519,319, which is a continuation-in-part of Ser. No. 370,862, Jan. 10, 1995, Pat. No. 5,530,350, which is a continuation-in-part of Ser. No. 226,061, Apr. 11, 1994, abandoned, which is a continuation of Ser. No. 794,931, Nov. 20, 1991, Pat. No. 5,302,896, said Ser. No. 491,632, is a continuation-in-part of Ser. No. 371,091, Jan. 10, 1995, abandoned, which is a continuation-in-part of Ser. No. 226,024, Apr. 11, 1994, abandoned, which is a continuation of Ser. No. 885,633, May 19, 1992, Pat. No. 5,302,897.

[51] **Int. Cl.**^{6} **G01V 3/00**
[52] **U.S. Cl.** **324/306**; 324/307; 324/309

[58] **Field of Search** 324/307, 309, 324/300, 318, 306

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,857,844	8/1989	Van Vaals	324/307
4,973,111	11/1990	Haacke et al.	324/309
4,980,640	12/1990	Van Ormondt et al.	324/307
5,519,319	5/1996	Smith et al.	324/306
5,530,350	6/1996	Dechene et al.	324/306

Primary Examiner—Louis M. Arana
Attorney, Agent, or Firm—Testa, Hurwitz & Thibeault, LLP

[57] **ABSTRACT**

An on-line nuclear magnetic resonance (NMR) system, and related methods, are useful for predicting one or more properties of interest of a polymer. In one embodiment, a neural network is used to develop a model which correlates process variables in addition to manipulated NMR output to predict a polymer property of interest. In another embodiment, a partial least square regression technique is used to develop a model of enhanced accuracy. Either the neural network technique or the partial least square regression technique may be used in conjunction with a described multi-model or best-model-selection scheme according to the invention. The polymer can be a plastic such as polyethylene, polypropylene, or polystyrene, or a rubber such as ethylene propylene rubber.

7 Claims, 15 Drawing Sheets

Figure 2.1: Patent document of the United States Patent and Trademark Office (USPTO)

which offers a comprehensive and detailed overview about patents and intellectual property.

2.2.1 Patent Document

A patent document is a legal document. Figures 2.1, 2.2 and 2.3 show the front page of patent documents applied for in the United States Patent and Trademark Office (USPTO), in the European Patent Office (EPO) and in the Spanish Patent and Trade Mark Office (SPTO). Generally, patent documents contain standard information, although this can vary depending on whether the patent is published by a national or regional office. All data in the patent document is identified according to the Internationally Agreed Numbers for the Identification of (bibliographic) Data, or INID codes. The best known are[2]:

- 11: Number of the patent.
- 12: Plain language designation of the kind of document, for instance European Patent Application.
- 19: WIPO Standard ST.3 code[3], or other identification, of the office or organization publishing the document.

[2] Source of information: http://www.wipo.int/standards/en/pdf/03-09-01.pdf.

[3] This is a recommended standard on two-letter codes for the representation of states, other entities and intergovernmental organizations. See http://www.wipo.int/standards/en/pdf/03-03-01.pdf.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 600 386 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
30.11.2005 Bulletin 2005/48

(51) Int Cl.7: **B65D 1/34**, B65D 1/46

(21) Application number: **05253240.5**

(22) Date of filing: **26.05.2005**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR
Designated Extension States:
AL BA HR LV MK YU

(30) Priority: **28.05.2004 US 857268**

(71) Applicant: **Sealed Air Corporation**
Saddle Brook New Jersey 07663 (US)

(72) Inventor: **Owensby, Joseph E**
Spartanburg, SC 29301 (US)

(74) Representative: **Jackson, Martin Peter et al**
J.A. Kemp & Co.,
14 South Square,
Gray's Inn
London WC1R 5JJ (GB)

(54) **Tray for case-ready meat**

(57) An optimized tray for case-ready meat products is disclosed. The tray comprises a base, a plurality of sidewalls connected at a plurality of corners, a flange

is varied in width. The ribs can also vary in thickness along at least a partial length of at least one of the sidewalls. The interaction of the at least one bowed sidewall,

Figure 2.2: Patent document of the European Patent Office (EPO)

- 21: Number(s) assigned to the application(s).
- 22: Date(s) of filing the application(s).
- 30: Data relating to priority under the Paris Convention to the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement).
- 43: Date of making available to the public by printing or similar process of an unexamined patent document, on which no grant has taken place on or before the said date.
- 45: Date of making available to the public by printing or similar process of a patent document on which grant has taken place on or before the said date.
- 51: International Patent Classification (IPC) codes.
- 52: Domestic or national classification.
- 54: Title of the invention.
- 56: List of prior art documents, if separate from descriptive test.
- 57: Abstract or claim.
- 58: Field of search.


(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: 2 186 549

(21) Número de solicitud: 200101549

(51) Int. Cl.^{7}: G01N 21/35

G01N 33/02

(12) SOLICITUD DE PATENTE A1

(22) Fecha de presentación: 03.07.2001

(43) Fecha de publicación de la solicitud: 01.05.2003

(43) Fecha de publicación del folleto de la solicitud: 01.05.2003

(71) Solicitante/s: **UNIVERSIDAD DE VIGO**
C/ Oporto 1
36201 Vigo, Pontevedra, ES

(72) Inventor/es: **Tojo Suárez, Emilia;**
Prado Fernández, Juan y
Rodríguez Vázquez, José Antonio

(74) Agente: **No consta**

(54) Título: **Análisis cuantitativo de carrageninas mediante espectroscopia de IR aplicando el método PLS.**

(57) Resumen:
Análisis cuantitativo de carrageninas mediante es-

Figure 2.3: Patent document of the Spanish Patent and Trade Mark Office (SPTO)

- 63: Number and filing date of the earlier application of which the present patent document is a continuation.
- 71: Name(s) of applicant(s), for example in the European Patent Office.
- 72: Name(s) of inventor(s) if known to be such, for example in the European Patent Office.
- 73: Name(s) of grantee(s), holder(s), assignee(s) or owner(s), for example in the U.S.
- 74: Name(s) of attorney(s) or agent(s).
- 75: Name(s) of inventor(s) who is (are) also applicant(s) and grantee(s), for example in the U.S.
- 84: Designated Contracting States under regional patent conventions.

2.2.2 Patent Databases

Today there are a number of databases that provide information about patents. Usually each national patent office has a patent database freely available on the Internet. The most extended databases are: the Patent Full-Text and Full-Page Image Databases of the United States Patent and Trademark Office; esp@cenet, co-ordinated by the European Patent Office; PatentScope administered by the WIPO; and the Industrial Property Digital Library (IPDL) of the Japan Patent Office and the National Center for Industrial Information and Training. Among applications offered by companies, the most commonly used databases are: Derwent World Patents Index (DWPI), Chemical Abstracts Plus (CAS) and Delphion. A comparison among these databases and others can be found in González-Albo & Zulueta (2007). Recently, Google implemented a service for reading the full text and downloading U.S. patents and patent applications.

2.3 Conceptual Constructions and Patent Indicators

Patents have been used to answer many research questions. We are particularly interested in studying the value of technologies when they are at an early stage of development. So, patents are used as a proxy for new technologies.

The most general patent indicator is the number of patents per company or country, per application or granted year. This indicator has been used by a number of researchers to study technological change. Nevertheless, a simple patent count has a temporal dimension and considers that all patents are equally important. To resolve the latter problem, Trajtenberg (1990b) proposed a weighted patent count based on citations received by patents and showed that this weighted indicator is highly correlated with the social benefits of innovations. Regarding the temporal dimension of patent count, Hall *et al.* (2001, 2005) called attention to the intrinsic bias of the data due to its truncation problem (for both patent and citations count). For instance, if we consider a sample of patents granted between 1991 and 1995, it should be noted that the sample will include (a) patents applied for before 1991 and granted between 1991 and 1995, (b) patents applied for and granted between 1991 and 1995, and (c) patents applied for between 1991 and 1995, and not granted before 1995. As expected, patents granted in the last years will account for fewer citations because some of them will have been made outside of the study period. Hence, Hall et al. have proposed a way to eliminate the truncation problem using "the application-grant empirical distribution to compute weight factors" for simple patent count.

Researchers have shown that data contained in patent documents, and those linked to them, are useful for making inference in statistical models. Section 2.2 showed that patent documents have a standard structure and they are available through a number of databases. Retrieving the right information, patent indicators can be built to be used later in a model. Examples of patent indicators are: the number of inventors, the number of applicants, the number of cited patents, references or backward citations (some researchers have made the distinction between citations made to other patents or made to scientific papers), the number of claims (dependent or independent), the number of international patent classification codes, type of priority (national or regional), the number of countries where the patent is protected and the number of designated states, the number of citations received by patents or forward citations, the number of words describing the state of the art, the number of words describing the technical problem, and the number of references to the technical advantages of an invention.

The first to use the information contained in patent documents was Jacob Schmooker in 1966[4]. The researcher considered the number of applicants per industry in order to match technological subclasses with the standard industrial classification (SIC) code of each company in the U.S.[5]. Later Williams Nordhaus suggested that the characteristics of inventions should contribute to its economic value. Examining the lifetime of patents, he concluded that value increases monotonically over the life of patents (Nordhaus, 1967; Reitzig, 2003) and that if the returns do not decrease over time, the optimal life of patents tends to be very large. The most important run-of-the-mill inventions tend to have a shorter life. Nordhaus presented some very interesting ideas as the concepts of drastic inventions, global interdependence or global spillover, optimal breadth of a patent, competing patents and inventing around patents.

From the work of these researchers, patent indicators have been widely used as measurement variables. In many cases the problem of interest can be represented by an idea, concept or construct. In what follows we provide a description of the constructs and patent indicators found in the literature.

2.3.1 International Scope, Patenting Strategy and Family Size

An application is usually first filed in a local country, receiving a filing/application number and an application date. The latter is also called priority date. Under the “Paris Convention priority right”, an applicant has a period of 12 months (priority period) from the priority date to protect its invention in other countries. For each additional country in which the application is made, the application will receive a national application number unique to that country. Family size of a patent is the set/number of countries –designated states– where the protection is sought for the same invention. After the priority year it is still possible to apply for a patent if the invention has not been made public. The prior art in a patent document does not include patent applications pending in other offices that have not been published.

Family size was first investigated by Grefermann *et al.* (1974) and Schmoch

[4]F. M. Scherer made an investigation a year earlier, but we do not have this paper: Firm size, market structure, opportunity and the output of patented innovations, America Economic Review, 1965, 55:1097-1125.

[5]The SIC code indicates the primary line of business filed by a company at the time of registration, www.osha.gov.

et al. (1988)[6], but it was introduced as a patent indicator by Putnam (1996)[7]. The size of patent family has been related to the international scope of patent protection. A larger number of countries would entail a broader international scope. Thus, it seems reasonable to think that a company protects an invention in many countries because it believes that the invention is, in some way, valuable.

The scope of protection is also related to the patenting strategy of the applicant (Guellec & van Pottelsberghe, 2000; Harhoff & Reitzig, 2004; Reitzig, 2004a). That is, where, how, and why an applicant protects an invention. For instance, the patenting strategy of a company may be reflected in whether a patent is first filed in a national patent office –giving priority to a local market– or in a regional office –looking for a broader protection. Guellec & van Pottelsberghe (2000) –using a probit model– attempted to determine the probability that a patent filed in the EPO is granted. They showed that it is more likely that a patent is granted if it is filed at the EPO via the Patent Cooperation Treaty (PCT) Chapter 2, or if the applicants are from different countries. Gallini (2002) pointed out that the patenting strategy is also associated with the company's intention to hinder the R&D of its competitors. This is what Reitzig (2003) calls "invent around."

2.3.2 Patent Breadth and the Number of Claims

The number of claims has been presented as an indicator by Tong & Frame (1994) and also used by other researchers, such as Lanjouw & Schankerman (2001) and Reitzig (2004a). Claims are made in a special section in the patent document, where the thing that is being protected is specified. The claims section consists of a numbered list. This contains all those aspects that are protected by the patent; thus, the claims reveal all the new technologies and knowledge. Therefore, the number of claims is in fact the number of inventions protected (Tong & Frame, 1994, p. 134).

Tong & Frame (1994) found that the conclusions derived from simple patent count may differ substantially from those obtained using the number of claims. The number of claims is a better indicator of the technological inventiveness of a country, and it is better correlated with other indicators of science and technology, such as the cost in R&D or the number of scientists and engineers of a country. On average there are countries that tend to specify their inventions in a larger number of claims than others. For example, Japan favors a smaller number of claims, unlike the U.S. that uses 35% more inventive units than Japan. There

[6] Grefermann *et al.* (1974) and Schmoch *et al.* (1988) are in German, they are cited by Reitzig (2004a).

[7] Putnam's work is a PhD thesis at Yale University. We could not access this information. However, the work is cited given the fact that most of the authors attribute this researcher with the introduction of family size as an indicator.

are no formal reasons to explain this; but perhaps there are cultural reasons. Although, some researchers have mentioned that it may be because it is cheaper to include many claims in a single patent. This fact may also reflect an increase in technological complexity, i.e. more complex systems require a larger number of related patents. The number of claims by patent tends to increase over time.

Klemperer (1990) and Gilbert & Shapiro (1990) have introduced the concept of patent breadth. Gilbert & Shapiro (1990, p. 6-7) give some definitions for breadth. They identified the patent breadth with "the flow rate of profit available to the patentee while the patent is in force" and also "as the ability of the patentee to raise price." For example, "a larger patent breadth in his model corresponds to a larger region of the product space that is included in the patent grant." We interpret from these definitions that the breadth of a patent is the potential of the patent to produce profits for a company –the company must be smart enough to produce these profits. It is assumed that a wider breadth will produce higher profits. So, the breadth concept is related to the exploitation of patent rights and the new opportunities provided by these rights. This concept is closely related to the technological scope in the sense that an invention may be useful in several technological fields. From our point of view, there is a subtle difference between both constructs. While technological scope refers only to the ability of the invention to be useful in several technological areas, the breadth also deals with the ability of the applicant to generate profits. The breadth may be reflected in the way that inventive units are described. This sometimes has to do with the ability of the agent who drafted the patent.

2.3.3 Technological Scope and the Number of IPC Codes

The WIPO manages four treaties that establish international classifications for inventions: the Strasbourg Agreement regarding to International Patent Classification (IPC), the Nice Agreement relating to International Classification of Goods and Services for the Purposes of the Registration of Marks, the Locarno Agreement concerning an International Classification for Industrial Designs, and the Vienna Agreement establishing an International Classification of the Figurative Elements of Marks[8].

The IPC system establishes a common classification for patents. National or regional patent offices are responsible for classifying an invention into a hierarchical system of sections, classes, subclasses, and groups. The classification system is comprised of eight sections with at least 70.000 branches, each with a particular symbol. Currently, it is available the ninth edition of the IPC system. Table 2.1 shows the IPC sections. For example, section H is divided into six classes: H01

[8]See http://www.wipo.int/classifications/en/.

Table 2.1: Sections of the International Patent Classification System

Section	**Description of Section**
A	Human necessities
B	Performing operations; transporting
C	Chemistry; metallurgy
D	Textiles; paper
E	Fixed constructions
F	Mechanical engineering; lighting; heating; weapons; blasting
G	Physics
H	Electricity

(basic electric elements), H02 (generation, conversion or distribution of electric power), H03 (basic electronic circuitry), H04 (electric communication techniques), H05 (electric techniques not otherwise provided for) and H99 (subject matter not otherwise provided for in this section). The class H02 includes subclasses and groups. Table 2.2 shows an example of the class code H02. Here only three subclasses are presented: H02B, H02G and H02H with their corresponding groups.

Even though countries have their own system for patent classification, the IPC system is used in more than 100 countries. This system allows document retrieval in the search for establishing the novelty of an invention or determining the state-of-the-art technology in a specific field. The IPC system facilitates the search for pertinent information. But it is important to define in a suitable fashion the search terms and to determine the technical terms that are of interest. WIPO has an official Catchword Index that contains key words for technical terms used in patent documents. In addition, the TACSY system provides support for searching through natural language processing.

On the other hand, technological scope or the level of technological protection of patents has to do with the applicability of the invention in different technological areas –for example as part of a product, or as a product itself in one or more industries. Harhoff *et al.* (2003) have emphasized that this variable is not easy "to operationalize and measure", so researchers have followed Lerner's approach (Lerner, 1994), which estimated the technological scope using the number of different four-digit IPC codes. Other researchers have suggested that the number of claims may also be a suitable indicator of technological scope.

2.3.4 Opposition and Indicators from Patent Text

Harhoff *et al.* (2003) introduced the outcome of opposition proceedings as a patent value indicator. The researchers said that the opposition is "a kind of first-instance challenge suit attacking the patent's validity" (p. 1345) and patents that survive this procedure are more valuable. In addition, Reitzig (2003) and Reitzig

Table 2.2: Example of classes, subclasses and groups of section H

Class Code	Description	Subclass Code	Description	Group Code
H02	Generation, conversion or distribution of electric power	H02B	Boards, substations or switching arrangements for the supply or distribution of electric power	H02B 1/01 ...
		H02G	Installation of electric cables or lines, or of combined optical and electric cables or lines	
		H02H	Emergency protective circuit arrangements	

(2004a) have studied almost all patent indicators that have been proposed in the literature. However, to the best of our knowledge, in the area of technological change, only Reitzig (2004a) has worked with indicators constructed from the technical description, the claims, or the abstract of the patent[9]. For example, the author used the number of words describing the state of the art and the number of words describing the technical problem. The former variable was used as a proxy for the novelty of the patents and the latter as an indicator of the degree of inventive step. For the particular case of Reitzig, these variables were found to be significantly correlated with the probability of an opposition when correcting the variables for heteroscedasticity. The author infers that "the number of words describing the technical problem mainly correlates with the potential profits from protecting the invention" (p. 954). The number of inventors and applicants have also been used as manifest variables for predicting the occurrence of an opposition.

2.3.5 Novelty and the Number of Inventors and Applicants

The novelty of patents is a difficult construct to model. As seen above, novelty describes the technological distance between the patented invention and the previous state of the art. Some researchers have linked this variable to the number of inventors and applicants, backward citations and number of claims (Reitzig, 2004a). A better indicator seems to be the number of words describing the state of the art, a variable also used by Reitzig (2004a).

2.3.6 Renewal Process and Renewal Information

Renewal information of patents refers to information about the patent renewal process. Renewing a patent, at the EPO for instance, is expensive. For this

[9]We know that computer science researchers have worked in the semantic processing of patent texts. However, the aim has been primarily to improve information retrieval in the information search process.

reason, the fact that a company renews their patent right each year is indicative of patent value. Information about the renewal process is not contained in the patent document. Although these data are more difficult to retrieve, researchers such as Pakes & Schankerman (1984), Pakes (1986), Schankerman & Pakes (1986), Pakes *et al.* (1989), Lanjouw (1998) and Lanjouw *et al.* (1998) have used indicators related to this process.

Pakes *et al.* (1989) studied the renewal process of patents because the patent holders are willing to pay fees for renewal only if the value of having a patent is greater than the cost of maintaining it. The hypothesis of Pakes et al. was "observations on the proportion of patents renewed at different ages, along with the relevant renewal fee schedules, will thus contain information on the distribution over the life span of the patents" (p. 331). The aim was to identify the "stochastic process generating the returns to patent protection" (p. 332).

2.3.7 Inventive Activity and Disclosure

Inventive activity is the technological distance between the protected invention and existing technology in terms of obviousness. Green & Scotchmer (1997) have studied this aspect of patents and have introduced the concept of technical non-obviousness. The term "disclosure" was also presented by Green & Scotchmer (1997), who mentioned that the technical information disclosed by the patent provides a positive externality to competitors, a fact that applicants wish to avoid. Reitzig (2004a) also related the number of words describing the technical problem to the inventive step.

2.3.8 Constructs and Citations

There are a number of contributions that make use of citations as manifest variables to study innovation (Trajtenberg, 1990b; Hall *et al.*, 2001), spillovers or knowledge diffusion (Jaffe *et al.*, 1993; Caballero & Jaffe, 2002), patent and market value (Hall *et al.*, 2005) and so forth. Each of these concepts or constructs is a complex multidimensional variable that can be approached from different perspectives. In this section we attempt to give a brief but thorough overview of citations and some of these constructs. In subsequent sections, we present in more detail the relationship between citations and patent value; as well as between the latter, R&D and market value.

Two types of citations can be distinguished: backward and forward citations. Backward citations are the number of citations made by a patent (Carpenter *et al.*, 1981; Narin *et al.*, 1997; Reitzig, 2004a). They can be made by applicants and patent examiners (Jaffe *et al.*, 1993; Jaffe & Trajtenberg, 1999) and may include references to other patents and to scientific articles (Harhoff *et al.*, 2003). Forward citations are the number of citations received by a patent (Trajtenberg, 1990b; Albert *et al.*, 1991; Harhoff *et al.*, 1999; Lanjouw & Schankerman, 2001; Harhoff *et al.*, 2003; Harhoff & Reitzig, 2004; Hall *et al.*, 2005). The number of citations received by patent A can be computed accounting for all citations made by patents granted after the granted date of patent A (Hall *et al.*, 2001). Trajtenberg (1990b) found that the distribution of citations depends on the patent's age; that is, the older patents will receive more citations than the new ones. Moreover, the researcher showed that the smaller the difference between the cited and citing year (the lag), the lower the effect of the citation lag distribution on citation counts.

Citations as measure of impact were introduced by the information scientist Eugene Garfield for scientific articles and journals in 1955. Citations are a measure of impact or importance because it is assumed that if a scientific paper or patent is cited, it is because this prior knowledge has been necessary to drive discussions –at least, on a given problem. Thus, this reflection enables us to propose alternative ways to address a problem and find new solutions. It was Carpenter *et al.* (1981) and Narin *et al.* (1997) who introduced the citations as indicators in the patent field and Trajtenberg (1990b) in economic research[10]. Carpenter *et al.* (1981) used the citations to examine whether the patents that protect the most important technological advances receive more citations. The authors were able to confirm this hypothesis and suggested citations as a suitable indicator for technological policy analysis. In a similar type of investigation, Narin *et al.* (1997) used backward citations to study the relationship between public science[11] and industrial technology. Narin *et al.* (1997) analyzed what proportion of the industrial patents cite scientific papers. These researchers coined the term "the patent-to-science linkage" (p. 318) and showed that three quarters of the publications cited by U.S. patents were papers from public science, and moreover they mostly correspond to results obtained in the country itself.

From an economic standpoint, Trajtenberg (1990b) linked patent count weighted by citations to the value of innovations. We discuss Trajtenberg's contributions of 1990 in section 2.4. Additionally, Trajtenberg *et al.* (1997) have used patent citations to study "basicness"[12] and appropriability, comparing results of universities and companies. As expected, universities have a lower measure of appropriability than companies. These researchers also developed measures of importance and generality of basic innovations. Generality is related to the utility and impact of an invention. The impact is evident when the invention is cited

[10]Trajtenberg (1990b, p. 173) said that "up to now, though, virtually the only patent measures used in economic research have been simple patent counts (henceforth SPC), that is, the number of patents assigned over a certain period of time to firms, industries, countries, etc."

[11]In this case, public science was represented by papers authored at academic, governmental and other public institutions (Narin *et al.*, 1997, p. 317).

[12]Basicness refers to the quality of basic research that can be carried out in universities.

by patents that belong to different technological fields. This generality index is proposed as follows: $Generality_i = 1 - \sum_j^{n_i} S_{ij}^2$ where S_{ij} represents the percentage of citations received by a patent i belonging to class j and n_j is the number of classes of patent i. This value will be high if the patent is cited by patents that belong to a broad range of technological fields and small if the citations are concentrated in a few fields. Originality index can be understood in the same way, but with citations made. Originality will be low if a patent cites patents which belong to a small group of technological fields. This approach does not necessarily work for all types of technologies. For example "if a nanotechnology patent is invented based on conventional miniaturization technologies, the patent will refer to a non-nanotechnology patent and eventually have a high score in the originality indicator" (Igami & Okazaki, 2007, p. 27). In general, these measures are biased and depend on the technological classification system that is used.

On the other hand, Jaffe *et al.* (1993) compared the geographical origins of the citing and cited patents to study how knowledge flows occur. Understanding location as geography, institutional and technology space and how this interacts with time, the researchers showed that patent citations are geographically located. This means that "citations to domestic patents are more likely to be domestic, and more likely to come from the same state." Moreover, Jaffe & Trajtenberg (1999) showed that the frequency of citations depends on: the grant year of the cited patent, the location of the cited inventor, the technological field of the cited patent, the grant year of the citing patent, and the location of the citing patent. Jaffe & Trajtenberg (1999) studied what percentage of citations received by patents, are citations made by the same applicant, i.e. correspond to self-citations. Self-citations are more common in the U.S. than in other countries and occurr more quickly. According to Jaffe *et al.* (1993), self-citations cannot be considered for studying spillover effects; rather, self-citations are an indicator of the accumulation of useful knowledge by companies and how they appropriate this knowledge. In the words of Trajtenberg *et al.* (1997), "self-citations are an important indicator of the cumulative nature of technology, and firms' ability to appropriate the returns to their inventions." Moreover, Jaffe & Trajtenberg (1999, p. 119) remark that "Putnam (1997) finds that the number of self-citations is a good predictor of firms' decision to pay renewal fees for patents that would otherwise expire." Jaffe & Trajtenberg (1999, p. 129-130) summarize their main results as follows: (1) patents assigned to the same company are more likely to cite each other and come sooner than other citations, (2) patents in the same patent class are approximately 100 times as likely to cite each other as patents from different patent classes, but there is not a strong time pattern to this effect, (3) "citations within the same patent class have a slight tendency to geographic localization, but, not surprisingly, much less so than citations within the same organization", (4) patents whose inventors reside in the same country are typically 30 to 80% more likely to cite each other than inventors from other countries and these citations come sooner, (5) there are clear country-specific citation tendencies; for instance, Japanese citations typically come sooner than those of other countries. There does not appear to be much interaction between the self-citation and technological proximity effects, and (6) there is strong symmetry between citing and cited intensities. Jaffe & Trajtenberg (1999) found that the probability –except for the U.S.– that a country cites to another one after 20 years is higher than after the first year.

Other results of Jaffe & Trajtenberg (1999, p. 108-109) are: (1) the probability that a given inventor will know of a given antecedent increases as the time lag between them grows while the probability that the antecedent will actually be helpful declines, on average; (2) patent citations are a proxy for a given bit of knowledge that is useful in the development of a descendent bit; (3) the citation frequency rises rapidly in the first few years after the cited patent; (4) a US-invented patent is much more likely to be cited by a US-invented patent than it is by a foreign-invented patent; (5) raw citation frequencies are afflicted by numerous theoretical and actual biases that make their interpretation dangerous; (6) we interpret the citation frequency as an estimate of a probability that a randomly drawn patent in the citing group will cite a randomly drawn patent in the cited group.

2.4 Patent Value

In recent years patent indicators have been used to study the economical value of patents. Not all researchers have worked with all patent indicators to estimate value. We can clearly distinguish two approaches. The first one focuses on the relationship between citations, patent value and market value of companies (Griliches, 1981; Griliches *et al.*, 1988; Griliches, 1990; Trajtenberg, 1990b; Hall *et al.*, 2001; Hall & Ziedonis, 2001; Hall *et al.*, 2005). The second one focuses on the relationship between patent indicators and patent value (Harhoff *et al.*, 1999, 2003; Reitzig, 2003; Harhoff & Reitzig, 2004; Reitzig, 2004a).

Patents are intellectual assets that do not necessarily have an immediate return. A patent may protect a product that can be manufactured and sold. But a patent may also protect technologies which, together with other technologies, enable the manufacture of a final product. In both cases, obtaining an economic value from patents may be extremely difficult. In studying patent value, different approaches have been taken throughout the literature. Some studies focus on the private value of a patent while others concentrate on the patent's social value.

2.4.1 Social Value of Innovations

Trajtenberg (1990b) studied the social value of innovations using as manifest variables a patent count weighted by citations in a multinomial logit model. The researcher defined the social value of innovations as follows:

> By value I mean the social benefits generated by the innovation in the form of the additional consumer surplus and the profits stemming from the innovation. The "value", "output", and "magnitude" of innovations are taken to mean exactly the same thing (p. 173).

It is true that the innovation concept is wider and also involves the commercialization process of a product. But innovations make reference to new products. It is assumed that these new products have been protected by a patent(s). In his book, Trajtenberg (1990a) refers to patents and innovations as interchangeable concepts. The author said that the value of a patent comprises three aspects:

> (a) The value of the property rights (VPR) conferred by the patent, which is that fraction of the profits generated by the innovation exclusively attributable to the extra monopoly power traceable to the legal exclusion of potential competitors, (b) the private value of the innovation/patent (PV), which is the present discounted value of the stream of additional profits to the assignee, brought about by the innovation disclosed in the patent –clearly, the private value of a patent is inclusive of the value of the property rights, and may actually be much larger than that– and (c) the social value of innovation/patent (SV), which consists, as repeatedly stated, of the extra surplus the innovation generates in the form of incremental consumer surplus and profits. In sum VPR ⊂ PV ⊂ SV; and recalling that $\triangle W$ comprises as measured only the incremental surplus, $SV = \triangle W + (PV - VPR) + VPR$, where $\triangle W$ are gains from innovation (quoted almost verbatim from Trajtenberg (1990a, p. 185)).

2.4.2 Private Value of Patents

Schankerman & Pakes (1986) and Lanjouw *et al.* (1998) have studied the private value of patents. Schankerman & Pakes (1986) proposed a model of patent renewal. The researchers assume that "since the renewal decision is based on the value of patent protection to the patentee, our procedure directly estimates the private value of the benefits derived from the patent laws" (p. 1052). On the other hand, Lanjouw *et al.* (1998) defined the private value of a patent in terms of "the difference in the returns that would accrue to the innovation with and without patent protection" (p. 407). The magnitude of this difference would be crucial in applying or renewing the protection. In a parallel type of investigation, Harhoff *et al.* (2003) have also focused on the private value of patents. The researchers defined the value "as the price for which the original inventor would be willing to sell the patent right" (p. 1344). Harhoff *et al.* (2003) said that the private value of patent comprises two values: (a) the value of renewed patent protection, and (b) the asset value of the patent right. "A third value concept could be considered as well: the value of the patent right to a 'stand alone' inventor who compares her profit in the case of technical leadership to the profit gained in some ex ante state of the industry" (p. 1246). Using a probit model, Harhoff *et al.* (2003) considered patent indicators as manifest variables for modelling the patent value. These researchers found that backward and forward citations are positively correlated with the patent value, the number of four-digit IPC codes is not an informative variable, and those patents with a larger family size and with opposition –or long litigation processes– tend to be more valuable.

Reitzig (2003) also studied the variables that determine the economic value of patents. The researcher found that for patents acting as bargaining chips, novelty and inventive activity are the important variables related to value, and inventing around and disclosure have less importance. Technical and marketing experts were surveyed and asked about the determinants of patent value. According to them, factors that determine patent value are: state of the art (existing technologies), novelty, inventiveness, breadth, difficulty of inventing, disclosure and dependence on complementary assets[13]. This author has made explicit the need to model the patent value as a construct or latent variable, Schankerman & Pakes (1986) also pointed out the unobservable nature of this variable.

Lanjouw & Schankerman (2004) used multiple indicators to estimate patent value. Value or quality index is modeled as an unobservable construct by using Factor Analysis. They use the term "quality to emphasize both the technological and value dimensions of an innovation" (p. 443). Among others, they found that patent quality "does not appear to have a strong impact on research productivity at the firm level, however patent quality is strongly associated with variations in market value firms." They also remarked that the use of a composite index reduces the variability in the unobservable construct, and that the latter is most useful when "one averages –either the mean over time for a given firm or the mean over firms for a given year." The indicators used by the researchers were backward citations, forward citations, patent family size and the number of claims[14]. Finally,

[13] We attempt to consider these variables as constructs in the proposed structural patent value models. However, recall that in this research, the manifest variables are mainly obtained from the patent documents. So, latent and manifest variables are subject to this constraint.

[14] In this paper, it is interesting to analyze the value of the obtained loadings in the relationships of each indicator and the corresponding construct.

Lanjouw & Schankerman (2004) recommended that: (1) family size should be directly related to the expected (private) value of protecting an innovation and this to the value of the innovation itself; (2) forward citations are most directly related to technological importance, this is also true for backward citations; (3) the number of claims is an indication that an innovation is broader and of greater potential profitability (p. 448).

2.4.3 Intrinsic and Potential Value of Technology

Guellec & van Pottelsberghe (2000) presented a value scale proposing that technology increases its own value as it passes through different stages: from invention to application, examination, publication and decision to grant, and finally to the high value stage if the patent is granted. The distinction is made between the intrinsic value of the patent simply for being granted (and thereby having proven novelty, inventive activity and applicability) and the potential value of technology (dependent on its potential for generating future returns). Even though Guellec & van Pottelsberghe (2000) develop a theory that supports their value scale, the researchers do not define exactly which value is meant when talking about patent value.

On the other hand, Pakes *et al.* (1989) distinguished between two value concepts: the value of the protection provided by patents and the value of the ideas underlying the patents. They said that for example "renewal data allow us to construct more accurate measures of the value of patented ideas than the measures obtained from the patent count indexes currently in use" (p. 332).

It is well-known that many elements may affect the invention and the protection process. Nevertheless, although the proposed models have proved useful, we believe that many of the concepts presented here can be analyzed as latent variables or constructs and can be related to in a structural model. What benefit is obtained with a SEM approach? A multidimensional view of the analyzed problem.

2.5 Market Value, R&D, Patents and Citations

Some researchers have studied the relationship between patents and R&D (Hausman *et al.*, 1984; Hall *et al.*, 1986; Trajtenberg, 1990b; Jaffe & Lerner, 2001), and between these variables and market value (Griliches, 1981; Pakes, 1985, 1986; Connolly & Hirschey, 1988; Griliches *et al.*, 1988; Megna & Klock, 1993; Lerner, 1994; Hall *et al.*, 2001; Hall & Ziedonis, 2001; Hall *et al.*, 2005).

Griliches (1981) was the first to study the relationship between market value, R&D and patents. Basing the research on a time-series cross-section analysis of U.S. companies, Griliches (1981) found a significant relationship among the market value of a company, the book value of the R&D expenditures and the number of patents. Later on, Pakes (1985) investigated the dynamic relationship between the number of successful applications/innovations, R&D expenditures (measure of the inventive activity of companies) and market value of companies (indicator of its inventive output). The researcher posed a dynamic factor analysis and found that “the events that lead the market to reevaluate the firm are indeed significantly correlated with unpredictable changes in both the R&D and the patents of the firm” (p. 406). This also happens vice-versa. Pakes (1985) also pointed out that “there is a large variance to the increases in the value of the firm that are associated with a given increase in its patents. This may reflect an extremely dispersed distribution of the values of patented ideas” (p. 406-407). As expected, the researcher reported that “most of the variance in the stock market rate of return has little to do with the firm’s inventive endeavors, at least as measured by its R&D input and its patent output” (p. 407). Pakes (1986) delved into these subjects. Because patents are protecting new technologies, it is difficult for a company to know at an early stage, whether this technology or innovation will generate revenues in the future. Pakes (1986) proposed a model to estimate the flow of future returns for an innovation.

Hall *et al.* (1986) analyzed the lag between patents and R&D expenditures on companies from the U.S. in the manufacturing sector in the 70s. The underlying issue in this research was that R&D expenditures are an investment in the knowledge stock of a company. This knowledge reduces its value as time passes. Using patent applications in a year t as a proxy of the acquired knowledge –and therefore added to the existing stock in the company– the researchers studied the relationship between R&D lag distribution and patents lag distribution. Some of the main results are that the relationship between R&D and patents remains, despite controlling for the effect of company size. They also remarked that the almost constant nature of R&D expenditures over time makes it quite complicated to use as input data in a model.

Presenting an extension of the works of Pakes (1985) and Pakes (1986), Connolly & Hirschey (1988) developed a Bayesian approach to relate market value and patents. Connolly & Hirschey (1988) found evidence to consider patents as “economically relevant information” proving a significant effect of patents on market value. Griliches *et al.* (1988) also analyzed the relationship between patents, R&D and market value. As expected, they found that “changes in patenting rates can account for only an infinitesimal fraction of the changes in the stock market value of the firm, and hence provide essentially no additional information to the estimation procedure.” Trajtenberg (1990b) showed that there is a strong association between simple patent count and R&D expenditures, and that “R&D explains a great deal of the cross-sectional variance in patenting but not much of the variation over time" (p. 183). Megna & Klock (1993) studied the relationship between patents and Tobin's q[15]. The researchers found that patents contribute to the change of q but they do not fully explain the variation. Lerner (1994) found that the patent scope –measured as the number of four-digit IPC codes– positively affect the market value of companies. Finally, Hall *et al.* (2001, 2005) have investigated the trend in U.S. patenting activity over the last 30 years. The researchers found that ratios of R&D to assets stock, patents to R&D and citations to patents, significantly affect market value. However, R&D stock appears correlated with the market value more than patent stock and more related to citation stock than to patent stock.

[15]The ratio q is defined as "the market's valuation of the financial claims on a firm to the cost of replacing that firm's assets" (Megna & Klock, 1993).

Chapter 3

Descriptive Analysis of the Renewable Energy Sample

Abstract. The purpose of this chapter is to describe the patent sample used in the thesis. The foci is on the criteria used to retrieve the data, how patent indicators were computed and the statistical description of the data. The sample includes 2,901 patents on renewable energy technologies, applied for and granted in the United States. Data were retrieved from a specialized database in October, 2007.

3.1 Introduction

Patent data, as data, involve a multidimensional complexity. The diversity of data contained in a patent documents allows the analysis under different perspectives. Depending on the objective, they may provide insights into the technological aspects of the protection, its geographic or temporal scope, about what kind of applicant creates the inventions or knowledge, and so on. In addition, patent data have an inherent legal nature, because inventions are protected according to the procedures and laws of each country. More important, the complexity of analyzing patent data lies in the meaning of the data and how this may be used to meet certain targets.

Years ago it was difficult to access to patent documents, and much more to the data contained in them in a simple way. To obtain useful results, researchers had to spend many hours extracting, organizing and cleaning data. But the situation has changed and advances in information technologies and the Internet have facilitated the data cleaning and preprocessing; although, depending on where one retrieve the data, is still a work-intensive task.

As this research attempts to propose and test a new patent value model, several strategies were used in order to avoid bias in the patent sample used in the analysis. It is well known that there is a large variability in the value or importance of technologies developed by companies. Likewise, a macro analysis by technological area is desired. Because it is an interesting field for Chile, the sample considers renewable energy patents. One might think that an easy way to identify inventions protected by companies in this area is to look for a companies' directory in the renewable energy field and then look for their patents in a database. However, many of the companies listed in directories under some criteria, such as market capitalization, are not necessarily producers of technologies, and often they do not appear in the patent databases. The energy field is particularly sensitive to this problem, because there are a lot of companies whose business is the exploitation of natural resources, or electric power generation, transmission and distribution. Hence, to facilitate the analysis and interpretation of results, all inventions protected in the renewable energy field were retrieved, and later it was identified if patents belong to a company, other institutions or individuals. In addition, it was only considered patents applied for and granted in the U.S. –however, this does not mean that patents owners are based on the U.S. or the priority country is the U.S. To obtain a random sample of patents under the aforementioned conditions, we arbitrarily chose four time-periods where patents were published. Since "whenever possible, the application data should be used as the relevant time placer for patents" (Hall *et al.*, 2001, p.10), we reorganized the data by application year and the indicators were computed.

This approach has three implications. First, we obtained a homogenous sample in terms of technology field and country. Second, it was found that there are companies from different industries that are developing renewable energy technologies and this heterogeneity could affect the results. Third, we are considering the renewable patent portfolio of companies, and not their complete patent portfolios. Therefore, if an analysis by company is done, only a part of the relationships that may exist between the studied variables is captured.

3.2 A Main Criterion to Retrieve the Data

Identification of renewable energy patents is not trivial. To the best of our knowledge, there are few studies that have used patent data related to energies, and they have used rather macro patent indicators, as the number of patent applications. None has considered a set of patent indicators as in this research. However, Johnstone *et al.* (2007) studied the effect of the environmental policies on technological innovation in renewable energies. In their work, they identified the international patent classification (IPC) codes related to renewable energies based on a set of keywords. These keywords were identified by the researchers making an extensive literature review of technological developments in this field.

The OECD Compendium of Patent Statistics (2007) also provides information relating to the IPC codes on renewable energies, but they referenced the work of Johnstone *et al.* (2007). So, in this research, the code list given by these authors was used to identify the appropriate set of patents.

The IPC codes include the section, class, sub-classes, main groups and subgroups related to wind, solar, geothermal, wave-tide, biomass and waste energies. Some examples are wind motors rotation axis substantially in wind direction (F03D 1/00-06), devices for producing mechanical power from solar energy (F03G 6/00-08), devices for producing mechanical power from geothermal energy (F03G 4/00-06), liquid carbonaceous fuels - organic compounds (C10L 1/14) and manufacture of fuel cells - combined with treatment of residues (H01M 8/06). Table 3.1 shows the technological field and the IPC codes compiled by Johnstone *et al.* (2007) in its study.

To retrieve patents granted in the U.S., some commercial database were tested. Finally, we used Delphion database, a product of Thomson Reuters, because offers many advantages –advanced search options are available– and it is widely recognized and used in patent search.

3.3 Comments about Delphion and Data Retrieval Process

The Delphion database allows visitors to seek a set of patents in at least two collections: granted or applied patents in the U.S. Data were retrieved by selecting the "U.S. (Granted)" option. So, patents have been granted and therefore may have been cited. Delphion database –in its advanced search– allows the patents' search combining different options. The following criteria were introduced to retrieve data:

- Type of patents: granted in the U.S.
- IPC Code: according to Johnstone *et al.* (2007).
- Publication Year: 1990-1991, 1995-1996, 2000-2001 and 2005-2006.

It is important to recall that, until recently, U.S. patents were only published after grant. This has changed since November 29, 2000[1]. Nowadays, U.S. patent applications are published 18 months after "the earliest effective filing date or priority data claimed by an application." According to the United States Patent and Trademark Office, USPTO[2], when the published patent is a granted patent,

[1]See http://www.uspto.gov/web/offices/pac/doc/general/index. html#pub.
[2]See http://www.uspto.gov/ web/offices/pac/mpep/documents /0900_901_05_b .htm.

Table 3.1: Application fields and IPC codes on renewable energies

Field name	IPC Codes	Definition in IPC (8th edition)
Wind	F03D 1/00-06	Wind motors with rotation axis substantially in wind direction
	F03D 3/00-06	Wind motors with rotation axis substantially at right angle to wind direction
	F03D 5/00-06	Other wind motors
	F03D 7/00-06	Controlling wind motors
	F03D 9/00-02	Adaptations of wind motors for special use
	F03D 11/00-04	Details, component parts or accessories not provided for in, or of interest apart from, the other group of this subclass
	B60L 8/00	Electric propulsion with power supply from force of nature, e.g. sun, wind
	B63H 13/00	Effecting propulsion by wind motors driving water-engaging propulsive elements
Solar	F03G 6/00-08	Devices for producing mechanical power from solar energy
	F24J 2/00-54	Use of solar heat, e.g. solar heat collectors
	F25B 27/00	Machine plant or systems using particular sources of energy - sun
	F26B 3/28	Drying solid materials or objects by processes involving the application of heat by radiation, e.g. sun
	H01L 31/042	Semiconductor devices sensitive to infrared radiation, including a panel or array of photoelectric cells, e.g. solar cells
	H02N 6/00	Generators in which light radiation is directly converted into electrical energy
	E04D 13/18	Aspects of roofing for the collection of energy, i.e. solar panels
	B60L 8/00	Electric propulsion with power supply from force of nature, e.g. sun, wind
Geothermal	F24J 3/00-08	Other production or use of heat, not derived from combustion, using natural or geothermal heat
	F03G 4/00-06	Devices for producing mechanical power from geothermal energy
	H02N 10/00	Electric motors using thermal effects
Wave/tide	F03B 13/12-24	Adaptations of machines or engines for special use, characterized by using wave or tide energy
	F03G 7/05	Mechanical-power producing mechanism, ocean thermal energy conversion
	F03G 7/04	Mechanical-power producing mechanism, using pressure differentials or thermal differences
	F03B 7/00	Water wheels
Biomass	C10L 5/42-44	Solid fuels based on materials of non-mineral origin, animal or vegetable
	F02B 43/08	Engines operating on gaseous fuels from solid fuel, e.g. wood
	C10L 1/14	Liquid carbonaceous fuels, organic compounds
	B01J 41/16	Anion exchange, use of materials, cellulose or wood
Waste	C10L 5/46-48	Solid fuels based on materials of non-material origin, refuse or waste
	F25B 27/02	Machine plant or systems using particular sources of energy-waste
	F02G 5/00-04	Hot gas or combustion, profiting from waste heat of exhaust gases
	F23G 5/46	Incineration of waste, recuperation of heat
	F01K 25/14	Plants or engines characterized by use of industrial or other waste gases
	C10J 3/86	Prod. of combustible gases, combined with waste heat boilers
	F23G 7/10	Incinerators or other apparatus consuming waste, field organic waste
	H01M 8/06	Manufacture of fuel cells, combined with treatment of residues

the INID (Internationally agreed Numbers for the Identification of Data) code 45 (date of patent) refers to "date of making available to the public by printing or similar process of a patent document on which grant has taken place on or before the said date." In this case, the patent document has a "B1 or B2" next to the patent number[3]. On the other hand, when a published patent is an application, the patent does not have the INID code 45, but the INID code 43 (publication date), that refers to "date of making available to the public by printing or similar process of an unexamined patent document, on which no grant has taken place on or before the said date." In this case the patent document has an "A1, A2 or A9" next to the publication number. The sample used in this research was retrieved using the "U.S. granted" option in Delphion. So, the publication year of

[3] See http://www.uspto.gov/web/patents/authority/kindcode.htm.

the patents corresponds to the granted year.

The following data were retrieved for each patent (the data mentioned below were available for all US patents; otherwise, the procedure to recover the missing data is specified):

- Original title of the patent.
- Maintenance status, that is the USPTO status regarding fee maintenance for the patent (the options are R1: reinstated, E1: expired, CC: certificate of correction issued, XT: term extended); there are missing values in this field and it was not possible to recover the information from another source.
- Number of claims.
- Number of pages of the patent document.
- Publication number, date and country (in the sample the country is the U.S.).
- Application number, date and country (in the sample the country is the U.S.).
- Priority number, date and country.
- Field of search, that is, the US class codes that represent the fields (by U.S. class) that were examined prior to the granting of the patent.
- Applicants' name and country, patents have missing data in this field, some data were recovered from the USPTO database.
- Inventors' name and country.
- Number of backward U.S. references, that is, U.S. patents and applications cited as references by this patent or application.
- Number of forward U.S. references, that is, U.S. patents or applications that cite this patent as a reference; this field does not include the foreign patents that have cited these patents.
- Foreign references, that is, the codes of non-US patents and application cited as references by the patent.
- Other citations, that is, non-patent prior art that patents reference; these data were not used in the analysis because it is textual information and is still being structured; data are not available.

- IPC codes; Delphion provides the following subfields: advanced and core codes, according to IPC reform of January 2006 (IPC-R or IPC8), and IPC-7 codes (codes before the reform). The IPC-7 data recovered from Delphion database contain 584 missing values. So, in this research it was considered the IPC-R data.

- Family information, that is, the set of patents filed with different patenting authorities that refer to the same invention.

3.4 Characterizing Patents and Computing Patent Indicators

Table 3.2 presents the number of patents retrieved under the selected criteria. We retrieved all available data under aforementioned conditions. The sample contains a 39.86% of patents on solar technologies, a 30.34% on waste technologies, a 15.74% on wind technologies, a 5.52% on wave/tide technologies, a 5.4% on geothermal technologies and a 3.14% on biomass technologies. A total of 3,349 patents were retrieved. If from this sample it is left out those patents that have a missing value in the applicant field, a final sample of 2,901 patents is obtained.

The 2,901 patents belong to 1,581 applicants. These correspond in a 69% to companies, 25% to individuals and 6% to universities, research centres or governmental institutions. Table 3.3 shows the number of patents and the percentage of patents by priority country. As can be seen, the vast majority of patents were first applied for in the U.S., but there is also a significant number of patents that were first filed in Japan, Germany, Great Britain and France. It is worth noting that only 31 patents were first sought through the European Patent Office (EPO) and 23 through the Patent Cooperation Treaty of the World International Patent Office (WIPO). This result follows the same tendency that the number of U.S. patents distributed by country of origin and by calendar year of grant. That is, the countries that more patent in the U.S. are Japan, Germany, United Kingdom, France, Canada, and so on. See the Patent Counts by Country/State and Year, Utility Patents Reports 2008 available on http://www.uspto.gov/web/offices/ac/ido/oeip/taf/ reports.htm#by_geog; in this case the patent origin is determined by the residence of the first-named inventor[4].

As a random way to recover patents, data were retrieved using the publication year. However, the series of grants tend to fluctuate more than the number of patens applied for and it is recommended to use application date for the analysis (Griliches, 1990; Hall *et al.*, 2001). Likewise, it has been showed that patents

[4]Recall that according to the U.S. laws, in this country, it is the inventor who apply for a patent.

Table 3.2: Number of patents retrieved under the selected criteria

Year	**Wind**	**Solar**	**Geothermal**	**Wave/tide**	**Biomass**	**Waste**	**Total**
1990	26	159	14	16	11	74	300
1991	27	138	23	16	6	54	264
1995	31	118	17	26	19	64	275
1996	39	154	14	23	12	60	302
2000	52	217	40	16	15	129	469
2001	60	211	24	33	16	172	516
2005	127	172	28	26	14	223	590
2006	165	166	21	29	12	240	633
Total	527	1335	181	185	105	1016	3349

statistics ordered by application year correlated better with economic indicators (Paci *et al.*, 2004, p. 27). So, data were reorganized by application year. Figure 8.2 shows the number of patents by priority, application and publication year. Although the retrieved sample does not include the full set of patent applied for by year on the renewable energy filed in the U.S., it is possible to observe as the number of patents applied for has gradually increased over time. Three peaks are clearly observed in the number of applications (1989 with 230, 1994 with 237 and 1999 with 358 applied for patents) and two periods of "drought" (1991-92 with 15 and 32 applied for patents and 1996 with 23 applied for patents). The "drought" periods coincided with a general decrease in the total number of application in the United States (see http://uspto.gov/go/taf/us_stat.htm for the U.S. patent statistics, calendar years 1963-2008). Recall that, in these years, there were economic crisis. Especially between 1996 and 1998, the so-called Asian crisis. On the other hand, the final decrease of the curves, number of patents by priority and application year, it is because the patent sample was retrieved by publication year. So, there are few patents applied for between 2004 and 2006 and granted in the U.S. before 2006.

Contrary to expectations, in all the most productive years in terms of number of patents, solar energy inventions have been more protected. One would expect that the wind energy industry create and protect more inventions.

From the retrieved data, the following indicators were computed for the subsequent analysis:

- The number of inventors, it was counted the number of names in inventors' field.
- The number of applicants, it was counted the number of names in assignees' field.

Table 3.3: Number and percentage of patents by priority country

Priority Country	Total	%	Priority Country	Total	%
United States	1705	58.77	Austria	9	0.31
Japan	566	19.51	Spain	9	0.31
Germany	233	8.03	Norway	9	0.31
Great Britain	58	2.00	South Africa	4	0.14
France	44	1.52	Belgium	3	0.10
EPO	31	1.07	Sri Lanka	3	0.10
Sweden	23	0.79	New Zealand	3	0.10
WIPO	23	0.79	Russian Federation	3	0.10
Korea	22	0.76	China	1	0.03
Denmark	20	0.69	Czechoslovakia	1	0.03
Australia	19	0.65	Greece	1	0.03
Canada	19	0.65	Ireland	1	0.03
Switzerland	18	0.62	Malaysia	1	0.03
Israel	16	0.55	Poland	1	0.03
Italy	15	0.52	Yugoslavia, Serbia and Montenegro	1	0.03
Netherlands	15	0.52	Zimbabwe	1	0.03
Finland	12	0.41	Total	2901	100.00
Taiwan	11	0.38			

- The number of cited patent, it was calculated as the number of backward US references plus the number of foreign references.
- The number of claims, it is provided directly by Delphion.
- The number of IPC codes, it was counted the number of four-digit IPC codes in the IPCs' field.
- The number of U.S. forward citations, it is provided directly by Delphion.
- The patent family size, it was counted the number of countries in which protection was sought.
- A set of dummy variables was computed from the priority country field to indicate where patents were first filed; the variable was called "prior country."
- A set of dummy variables indicating whether the patents were filed in other countries besides the U.S.; this indicator was computed from the data contained in the patent family field; the major producers or markets of renewable energy were considered; the variable was called "PN country."
- The time-lag of the patents, it was computed as the difference between the granted and application years.

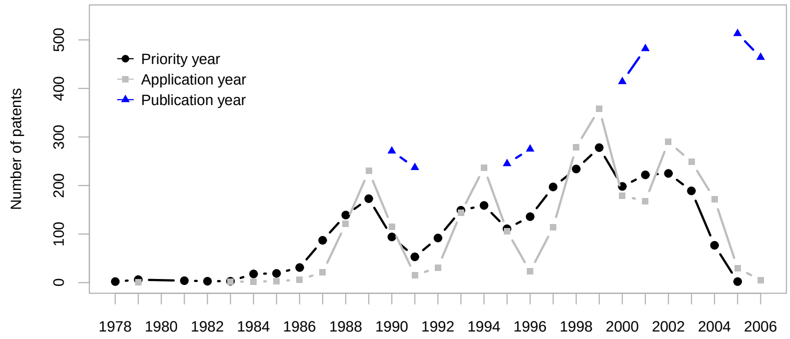


Figure 3.1: Number of patents by priority, application and publication year

3.5 Statistical Description of the Sample

Table 3.4 shows the descriptive statistics for the patent indicators (manifest variables) used in the analysis. The count data indicates that some variables are very heterogeneous and asymmetric and they also exhibit large variances. This is the case of the cited patents, the number of claims, the forward citations and the size of the patent family. Positive values of skewness indicate positive/right skew. Notice that the backward and forward citations and the family size have very similar skewness index. Likewise, positive kurtosis indexes show distributions sharper than normal peak. The histograms of each variables are shown in Figure 3.5, they confirm the skewed nature of data. In this patent sample, the inventions are created on average by two inventors (mean = 2.22, median =2) and the vast majority have one applicant. The priority dummy variables more informative are Prior JP and Prior DE; the same happens with PN JP and PN DE. On average, the time-lag between application and granted years is 2.28 years.

Tables 3.6 and 3.7 show the Pearson and Spearman correlations, respectively and significance (Bayer *et al.*, 1990; Ahlgren *et al.*, 2003). In general, correlations are small and medium correlations[5], but many of them are significant. This may be expected because despite indicators for each patent come from a same patent document, the variables are actually generated by very different phenomena. For instance, the number of inventors who create an invention is not necessarily correlated with the number of countries where the invention sought to be protected. The latter depends rather on the patenting strategy that may have the applicant and the decision will probably not depend on the inventors.

[5]Cohen (1988) suggests that correlations of 0.1, 0.3, and 0.5 express small, medium and large effect sizes, respectively.

Table 3.4: Descriptive statistics for patent indicators

Variable	Mean	Median	Minimum	Maximum	S.D.	Skewness	Kurtosis
$N°$ inventors	2.22	2	1	14	1.58	1.76	4.23
$N°$ applicants	1.05	1	1	9	0.29	12.85	260.81
$N°$ cited patents	15.36	11	0	327	18.97	5.54	50.79
$N°$ claims	17.03	14	1	279	15.08	4.29	43.65
$N°$ IPC codes	6.29	5	1	48	4.52	2.09	7.71
$N°$ forward citations	5.63	2	0	158	10.17	5.30	46.83
Family size	8.54	6	1	202	11.62	5.58	51.27
Prior US	0.59	1	0	1	0.49	-0.36	-1.87
Prior JP	0.20	0	0	1	0.40	1.54	0.37
Prior DE	0.08	0	0	1	0.27	3.09	7.55
Prior GB	0.02	0	0	1	0.14	6.86	45.12
Prior FR	0.02	0	0	1	0.12	7.94	61.05
Prior EP	0.01	0	0	1	0.10	9.52	88.75
Prior WO	0.01	0	0	1	0.09	11.10	121.35
Prior Nin	0.08	0	0	1	0.28	3.01	7.09
PN US	0.96	1	0	1	0.20	-4.68	19.87
PN JP	0.44	0	0	1	0.50	0.23	-1.95
PN DE	0.33	0	0	1	0.47	0.75	-1.44
PN GB	0.05	0	0	1	0.22	4.20	15.65
PN FR	0.03	0	0	1	0.18	5.22	25.30
PN EP	0.44	0	0	1	0.50	0.25	-1.94
PN WO	0.35	0	0	1	0.48	0.63	-1.60
PN Nin	0.01	0	0	1	0.09	10.86	116.09
Time lag	2.28	2	0	12	1.21	1.16	2.66

Some results are interesting. It is possible to differentiate between the correlations (1) between the indicators, (2) between the indicators and the dummy variables, and (3) the correlations between the dummy variables. The highest correlation is between the family size and the number of IPC codes (rho= 0.27, p-value < 0.01)[6]. This may mean that patents classified in a larger number of IPC codes tend to be filed in more countries. It is worth noting that the correlation between the number of claims and the number of cited patents is 0.20 (p-value $<$ 0.01). This may mean that patents with more claims tend to refer more previous U.S. and foreign references.

With regard to correlations between the dummy variables, it was found a significant negative correlation between the Prior U.S. and the family size (rho = -0.29, p-value < 0.01). The first variable indicates if the patent has been first applied for in the U.S. This seems to have a negative relationship with the patent family size, i.e. the number of countries where the protection is sought. Moreover, it was found a significant positive correlation between the Prior JP and the number of inventors (rho=0.25, p-value < 0.01). Patents that have been first applied for in Japan have a positive relationship with the number of inventors. Unlike the correlation between the Prior U.S. and the number of inventors that is significant but negative (rho=-0.15, p-value < 0.01). Likewise, and as expected, the correlations

[6]Spearman correlation

between the Prior U.S. and Prior JP, and between Prior U.S. and Prior DE are negative and significant (rho=-0.58 and rho=-0.35 respectively, p-value < 0.01).

The highest correlations are obtained between the family size and the PN EP (rho=0.71, p-value < 0.01), PN JP and PN DE (rho=0.57, p-value < 0.01), PN WO (rho=0.53, p-value < 0.01) and PN GB (rho=0.24, p-value < 0.01). Recall that the variables PN-country indicates if the patents have been filed in other countries besides in the U.S. (59% of the patents have been first applied for in the U.S.). Moreover, positive and significant correlations were found between the number of IPC codes and PN JP and PN EP (rho=0.25 and rho=0.19 respectively, p-value < 0.01). There are significant medium correlations between the Prior U.S. and PN JP, PN DE and PN EP, but all of them negative (rho=-0.35, rho=-0.30 and rho=-0.24 respectively, p-value < 0.01). On the contrary, there are significant positive correlations between PN EP and PN DE (rho=0.47, p-value < 0.01), PN EP and PN JP (rho=0.42, p-value < 0.01), PN EP and PN WO (rho=0.40, p-value < 0.01), and PN FR and PN GB (rho=0.23, p-value < 0.01).

Finally, there is a significant negative correlation between the time-lag of the patents and the citations received (rho=-0.35, p-value < 0.01). As expected, the larger is the difference between the granted and application years, the smaller is the number of citations received. On the contrary, significant positive correlations were found between the time-lag and the number of IPC codes, and between the time-lag and the number of cited patents (rho=0.22 and rho=0.19 respectively, p-value < 0.01).

3.6 Patent Indicators by Application Year

Figure 3.2 presents the number of patents by application year and technological field. The number of patents increases over time. The inventions related to solar energy are the most important in quantity, then those related to waste and wind energy. Figures 3.3 and 3.4 show the number of cited patents, the forward citations, the mean citations made and received by application year. The number of cited patents and the mean citations made by patent increase over time. The same applies to the number of citations received but, as expected, the growth over time is less strong. As in the case of the number of patents by application, priority and publication year, the final decrease of the curves in the recent years is due to patents applied for and granted in recent years have received few citations.

Table 3.5: Histograms of patent indicators

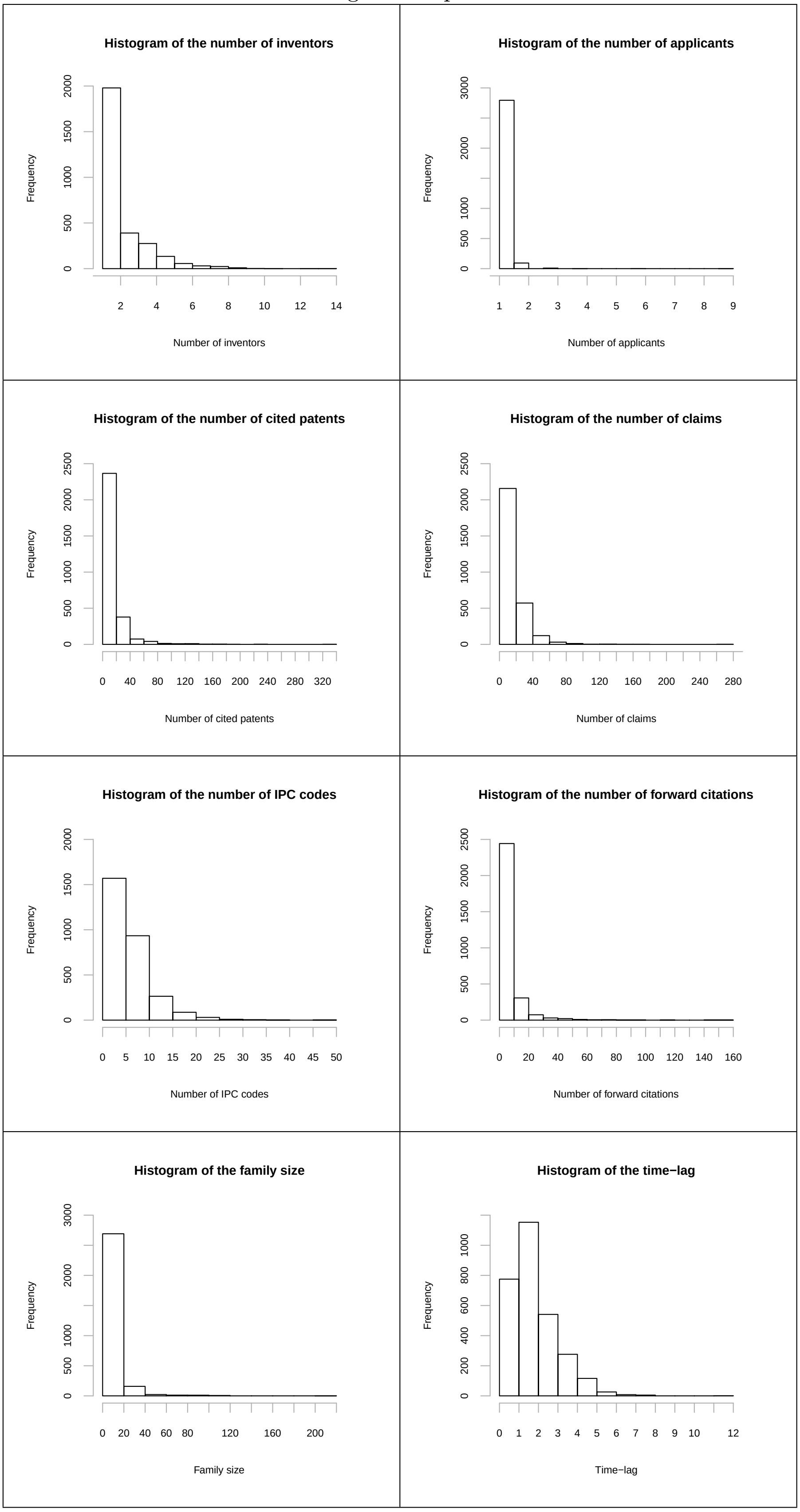


Table 3.6: Pearson correlations among patent indicators and significance (** p < 0.01, * p < 0.05)

Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
$N°$ inventors	(1)	1																							
$N°$ applicants	(2)	0.15**	1																						
$N°$ cited patents	(3)	0.04*	-0.00	1																					
$N°$ claims	(4)	0.08**	0.02	0.20**	1																				
$N°$ IPC codes	(5)	0.14**	0.03	0.14**	0.10**	1																			
$N°$ forward citations	(6)	0.01	-0.02	-0.02	0.06**	-0.03	1																		
Family size	(7)	0.04*	-0.01	0.18**	0.19**	0.22**	0.06**	1																	
Prior US	(8)	-0.17**	-0.04*	0.13**	0.15**	-0.09**	0.11**	-0.06**	1																
Prior JP	(9)	0.28**	0.04*	-0.12**	-0.11**	0.12**	-0.05**	-0.03*	-0.58**	1															
Prior DE	(10)	0.00	0.00	-0.05**	-0.05**	0.01	-0.03*	0.03*	-0.35**	-0.14**	1														
Prior GB	(11)	-0.04*	-0.00	0.01	-0.01	-0.00	0.01	0.06**	-0.17**	-0.07**	-0.04*	1													
Prior FR	(12)	-0.02	0.04**	-0.01	-0.02	0.00	-0.03	0.03	-0.14**	-0.06**	-0.03*	-0.01	1												
Prior EP	(13)	0.00	0.03	-0.02	-0.02	-0.00	-0.00	0.00	-0.12**	-0.05**	-0.03	-0.01	-0.01	1											
Prior WO	(14)	-0.01	-0.00	-0.00	-0.02	-0.01	-0.01	0.02	-0.10**	-0.04*	-0.02	-0.01	-0.01	-0.00	1										
Prior Nin	(15)	-0.08**	-0.02	-0.00	-0.02	-0.02	-0.07**	0.06**	-0.36**	-0.14**	-0.08**	-0.04**	-0.03*	-0.03	-0.01	1									
PN US	(16)	-0.03*	0.00	0.02	-0.01	0.02	-0.07**	0.04*	-0.01	-0.03	0.04**	0.01	0.01	0.00	-0.00	0.01	1								
PN JP	(17)	0.17**	-0.01	0.02	0.01	0.23**	0.05**	0.37**	-0.35**	0.38**	0.03*	0.01	0.00	0.02	-0.00	0.02	0.05**	1							
PN DE	(18)	0.08**	0.01	0.01	0.02	0.10**	0.05**	0.42**	-0.30**	0.02	0.31**	0.09**	0.08**	0.03*	-0.00	0.08**	0.09**	0.32**	1						
PN GB	(19)	-0.02	-0.02	0.00	-0.01	0.04*	0.01	0.20**	-0.10**	-0.00	-0.03*	0.50**	-0.01	-0.00	-0.00	-0.02	0.03	0.06**	0.14**	1					
PN FR	(20)	0.01	0.02	-0.03	-0.04*	0.02	-0.02	0.09**	-0.10**	0.02	-0.02	-0.02	0.51**	-0.01	-0.01	-0.03	0.02	0.11**	0.18**	0.23**	1				
PN EP	(21)	0.11**	0.00	0.10**	0.10**	0.17**	0.03*	0.45**	-0.24**	0.03	0.14**	0.10**	0.07**	0.07**	0.03*	0.11**	0.14**	0.43**	0.47**	0.01	-0.00	1			
PN WO	(22)	0.04*	-0.03*	0.17**	0.10**	0.15**	-0.02	0.34**	-0.04*	-0.11**	0.01	0.09**	0.04*	0.00	0.10**	0.12**	-0.02	0.20**	0.15**	0.05**	-0.01	0.40**	1		
PN Nin	(23)	-0.02	-0.01	-0.00	-0.01	0.00	0.05**	-0.05**	-0.00	-0.03*	-0.01	-0.01	0.02	-0.00	-0.00	0.06**	-0.44**	-0.08**	-0.06**	-0.02	-0.01	-0.08**	-0.06**	1	
Time-lag	(24)	0.11**	0.00	0.13**	0.11**	0.22**	-0.18**	0.06**	-0.18**	0.08**	0.07**	0.04*	0.04*	0.02	0.05**	0.04*	0.04*	0.08**	0.03*	-0.01	-0.01	0.14**	0.18**	-0.00	1

Table 3.7: Spearman correlations among patent indicators and significance (** $p < 0.01$, * $p < 0.05$)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
N° Inventors	(1)	1																							
N° applicants	(2)	0.14**	1																						
N° cited patent	(3)	0.06**	-0.04**	1																					
N° claims	(4)	0.11**	0.00	0.20**	1																				
N° IPC codes	(5)	0.17*	0.00	0.12**	0.08**	1																			
N° forward citations	(6)	-0.04*	-0.04*	-0.06**	0.00	-0.14**	1																		
Family size	(7)	0.16**	-0.02	0.15**	0.09**	0.27**	-0.05**	1																	
Prior US	(8)	-0.15**	-0.04**	0.11**	0.15**	-0.11**	0.16**	-0.29**	1																
Prior JP	(9)	0.25**	0.06**	-0.11**	-0.13**	0.14**	-0.08**	0.08**	-0.58**	1															
Prior DE	(10)	0.02	0.00	-0.01	-0.03*	0.02	-0.05**	0.11**	-0.35**	-0.14**	1														
Prior GB	(11)	-0.03*	-0.00	-0.02	0.00	0.00	-0.01	0.12**	-0.17**	-0.07**	-0.04*	1													
Prior FR	(12)	-0.01	0.02	0.01	-0.01	0.00	-0.03	0.09**	-0.14**	-0.06**	-0.03*	-0.01	1												
Prior EP	(13)	0.01	0.03	-0.01	-0.02	-0.00	-0.04*	0.04*	-0.12**	-0.05**	-0.03	-0.01	-0.01	1											
Prior WO	(14)	-0.01	0.00	-0.00	-0.01	-0.02	-0.02	0.05**	-0.10**	-0.04*	-0.02	-0.01	-0.01	-0.00	1										
Prior Nin	(15)	-0.08	-0.02	-0.01	-0.03	-0.02	-0.06**	0.16**	-0.36**	-0.14**	-0.08**	-0.04*	-0.03*	-0.03	-0.01	1									
PN US	(16)	-0.03*	0.00	0.04*	-0.02	0.03	-0.07**	0.12**	-0.01	-0.03	0.04*	0.01	0.01	0.00	-0.00	0.01	1								
PN JP	(17)	0.17**	-0.00	-0.00	-0.01	0.25**	0.00	0.57**	-0.35**	0.38	0.03*	0.01	0.00	0.02	-0.00	0.02	0.05**	1							
PN DE	(18)	0.08**	0.00	0.03	0.02	0.11**	0.03*	0.57**	-0.30**	0.02	0.31**	0.09**	0.08**	0.03	-0.00	0.08**	0.09**	0.32**	1						
PN GB	(19)	-0.01	-0.02	-0.03*	-0.03	0.04*	0.03	0.24**	-0.10**	-0.00	-0.03*	0.51**	-0.01	-0.00	-0.00	-0.02	0.03*	0.06**	0.14**	1					
PN FR	(20)	0.02	0.00	-0.03*	-0.06**	0.03	-0.00	0.17**	-0.09**	0.02	-0.02	-0.02	0.51**	-0.01	-0.01	-0.03	0.02	0.11**	0.18**	0.23**	1				
PN EP	(21)	0.11**	-0.00	0.13**	0.08**	0.19**	-0.02	0.71**	-0.24**	0.03	0.14**	0.10**	0.07**	0.07**	0.03*	0.11**	0.14**	0.42**	0.47**	0.01	-0.00	1			
PN WO	(22)	0.05**	-0.03	0.17**	0.10**	0.15**	-0.11**	0.53**	-0.0419*	-0.11**	0.01	0.09**	0.04*	0.00	0.10**	0.12**	-0.02	0.19**	0.15**	0.05**	-0.01	0.40**	1		
PN Nin	(23)	-0.02	-0.01	-0.03	-0.01	-0.00	0.03	-0.08**	-0.00	-0.03	-0.01	-0.01	0.01	-0.00	-0.00	0.06**	-0.44**	-0.08**	-0.06**	-0.02	-0.01	-0.08**	-0.06**	1	
Time-lag	(24)	0.13**	0.02	0.19**	0.13**	0.22**	-0.35**	0.15**	-0.16**	0.08**	0.07**	0.03	0.03*	0.02	0.04*	0.03	0.04*	0.08**	0.02	-0.01	-0.02	0.13**	0.17**	-0.01	1

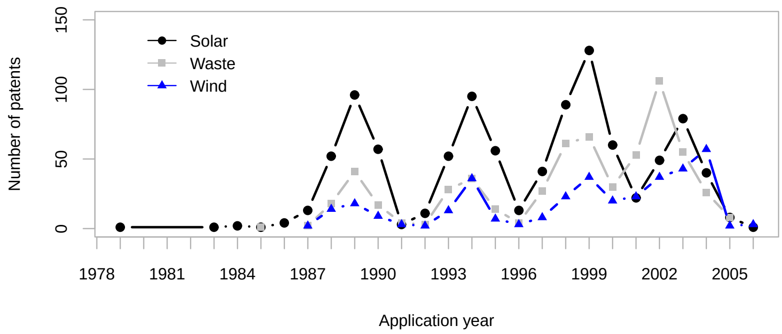


Figure 3.2: Number of patents by application year and technological field

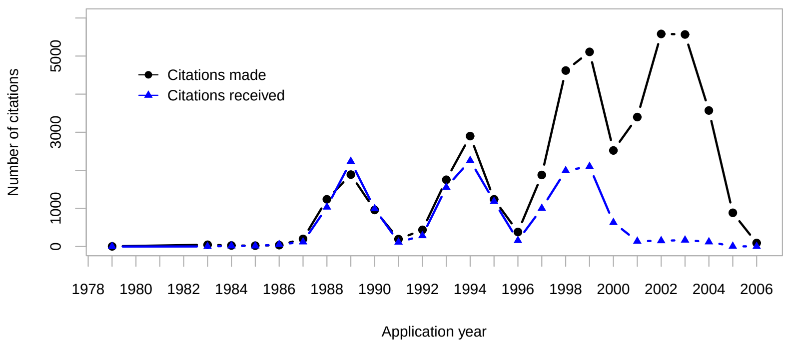


Figure 3.3: Number of cited patents and forward citations by application year

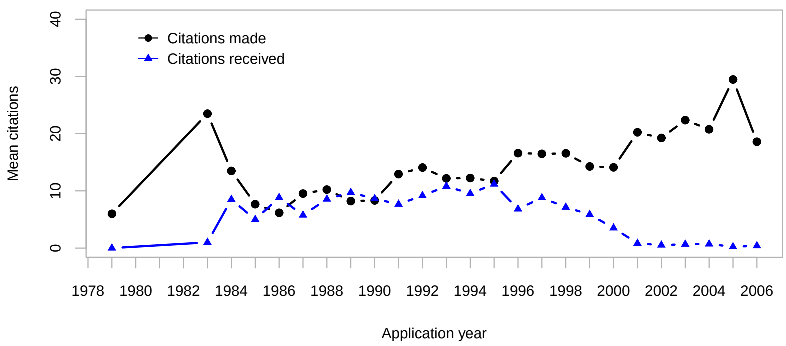


Figure 3.4: Mean citations made and received by application year

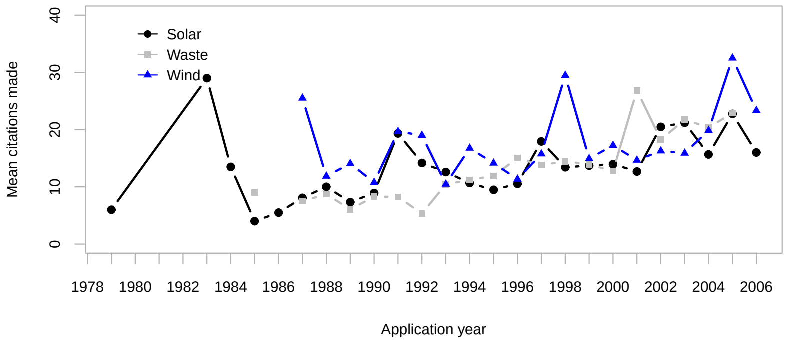


Figure 3.5: Mean citations made by application year and technological field

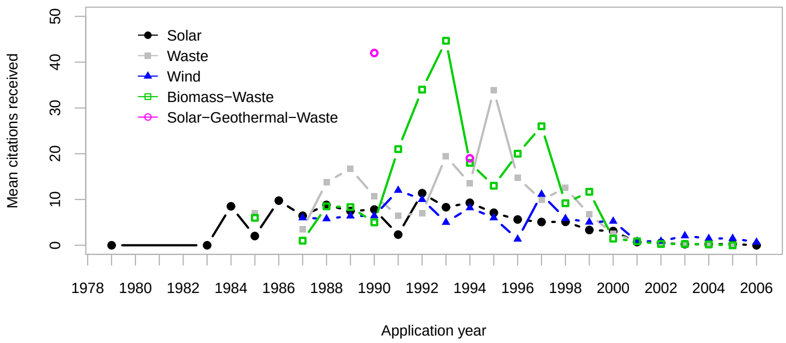


Figure 3.6: Mean citations received by application year and technological field

Figure 3.5 presents the mean citations made by application year in the case of solar, waste and wind technologies. All technology areas show an increase in mean citations made over time. It is note worthy that although there is a smaller number of patent in wind technologies in the sample, they has made, in average, more citations than patents in solar technologies. This may mean that wind technology companies have quickly appropriated new inventions. Figure 3.6 shows the mean citations received by application year and technological field. The mean citations received appear to be very similar for solar and wind technologies. However, striking the case of patent classified in IPC codes related to biomass and waste energies. The behavior is clearly different from other areas between 1991 and 1997.

Figures 3.7 and 3.8 presents the mean time-lag distribution by application year and by application year and technological field, respectively. As can be seen, the elapsed time between the granted and application years gradually decreased until

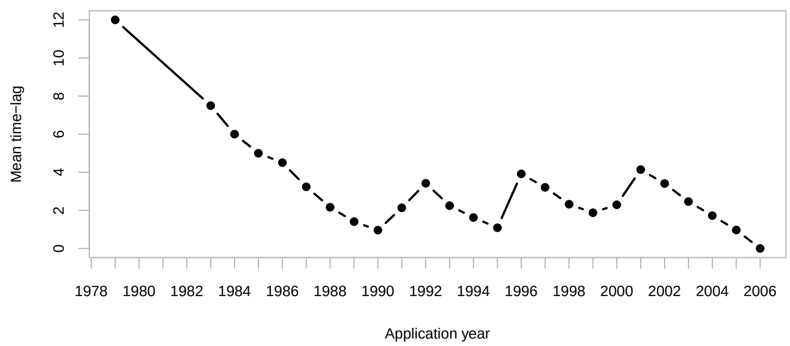


Figure 3.7: Mean time-lag distribution by application year

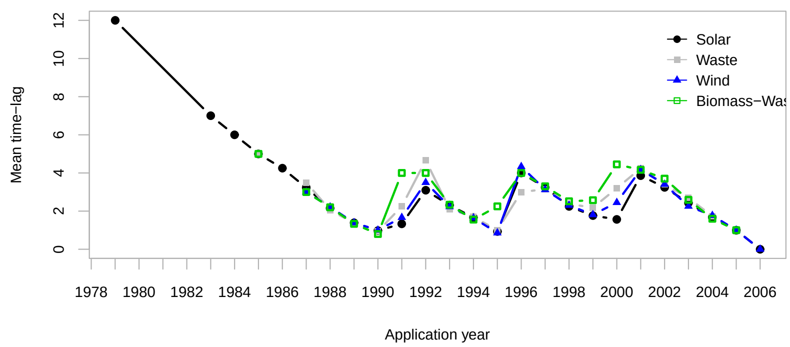


Figure 3.8: Mean time-lag distribution by application year and technological field

1990. In this sample, the first protected inventions are related to solar energy (the first patent was applied in 1978). In the mid-80s, inventions related to wind and waste energy are protected.

The mean number of claims by application year and by application year and technological field (Figures 3.9 and 3.10) seems to be stabilizing over the years. Most patents have between 15 and 20 claims. The situation is completely different for the mean number of IPC codes by application year, and by application year and technological field (Figures 3.11 and 3.12). Clearly, as time passes, inventions tend to be classified in more IPC codes. In particular, this happens especially for inventions related at the same time to waste and biomass energies. Inventions classified in IPC codes related to solar, waste and wind energy seem to aim at a number of IPC codes (between 5 and 10). It is not the case of inventions classified in IPC codes associated with wave/tide energies. These data features –the number of claims seems to be a more stable indicator over time than the number of IPC

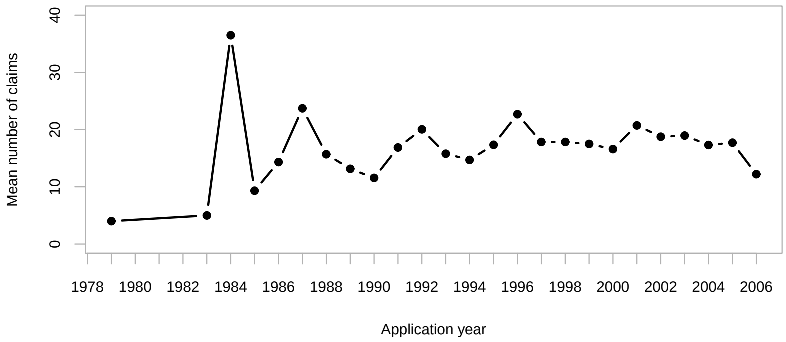


Figure 3.9: Mean number of claims by application year

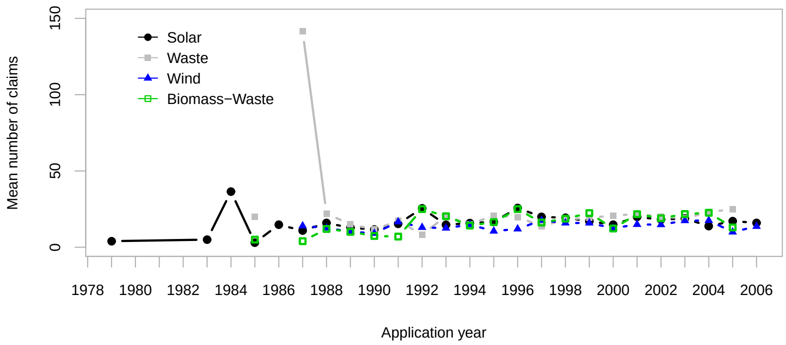


Figure 3.10: Mean number of claims by application year and technological field

codes– may explain why the former shows to be less related to the technological scope that the latter.

The mean family size by application year (Figures 3.13) did not show a tendency to increase or decrease over time, it rather seems to converge to a value between 5 and 10. Figure 3.14 presents the mean family size by application year and technological field. Again the inventions classified at the same time in IPC codes associated with the waste and biomass energies, show a mean family size larger than the rest of the inventions.

3.7 Longitudinal Nature of Forward Citations

It is important to emphasize that some of the patent indicators described above have a temporary nature. The number of inventors, applicants, cited patents, claims and IPC codes are determined at the time of filing of the patent applica-

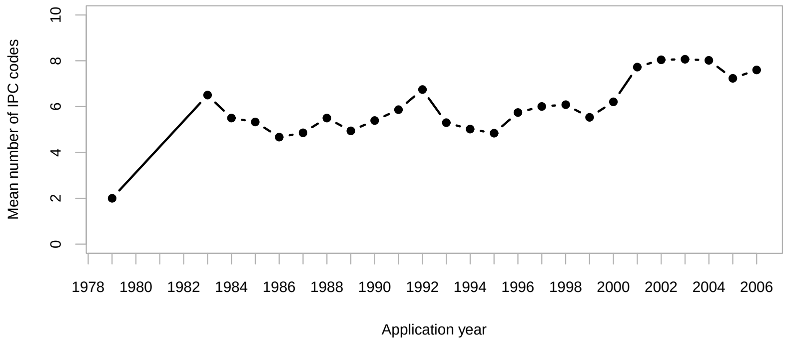


Figure 3.11: Mean number of IPC codes by application year

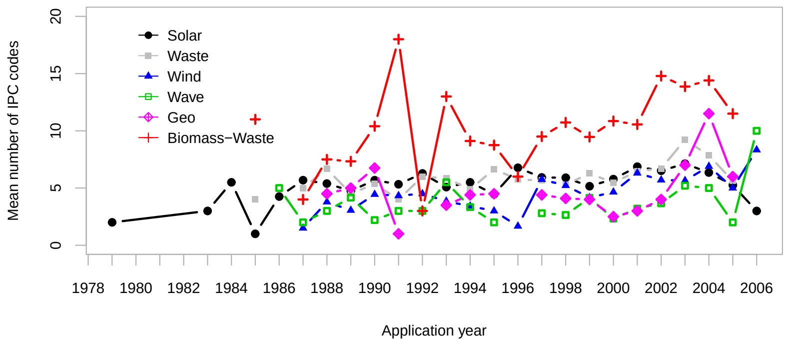


Figure 3.12: Mean number of IPC codes by application year and technological field

tion or during the patent examination process. One may assume that they are determined at the instant zero. However, this does not apply to citations received. The forward citations are received as time passes. Due to the temporary nature of this indicator and since this feature may have important implications on the results of the estimated models, it was retrieved the number of forward citations by year for each patent applied for in 1989, 1990, 1991, 1995, 1996 and 2000. Data were retrieved from USPTO database.

Figures 3.15, 3.16, 3.17 and 3.18 show the number of citations received by year, the accumulated citations received by year, the mean citations received by year and the mean accumulated citations received by year, respectively, for patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000. In all these figures, it is observed an increase in the number of citations over time. In the figure 3.15, it is possible appreciate that the number of citations reach a peak for then decrease. For patents applied for in 1989, 1990 and 1991, probably this is because patents

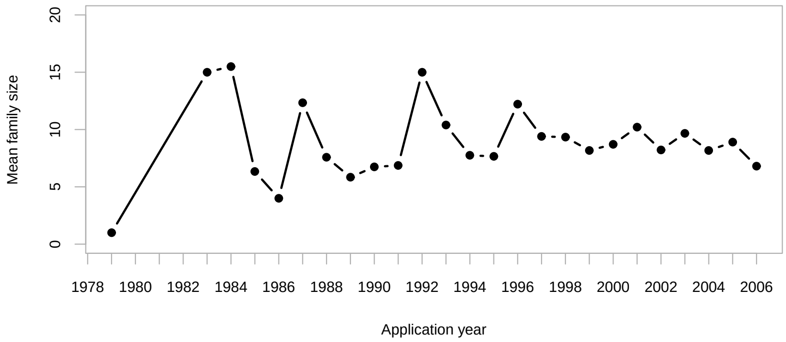


Figure 3.13: Mean family size by application year

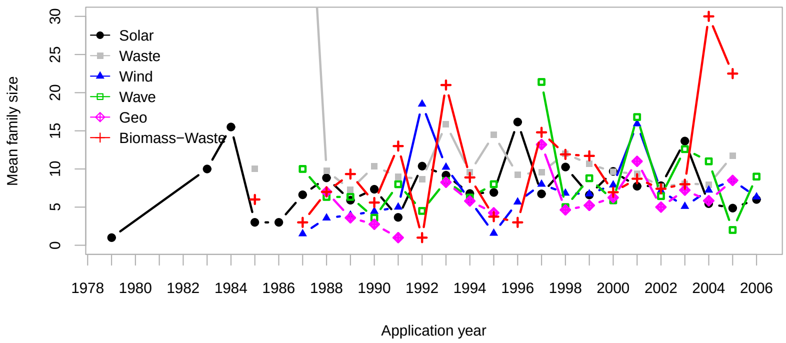


Figure 3.14: Mean family size by application year and technological field

are less cited in recent years. Perhaps inventions are less useful in technological terms or knowledge begins to become obsolete –between 1990 and 2006, there is a period of 16 years. However, for patents applied for in 2000, this decline in the curve is mainly due to patents applied for and granted in recent years have not received all the citations that probably they will receive.

Patents applied for in 1989 are the most cited. It is worth noting that patents applied for in 1995 have received more citations than patents applied for in 1990 and 1991. This is clearly shown in figure 3.16. The patents less cited are those applied for in 1991 and 1996. With regard to the mean citations received by year, the patents applied for in 1995 are clearly more cited (Figure 3.17). Even though the patents applied for in 2000 are recent, they have been rather cited. On average, patents applied for in 1995 have more accumulated citations than the rest of patents.

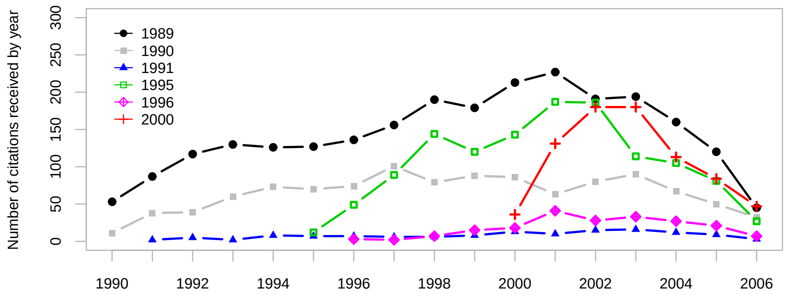


Figure 3.15: Number of citations received by year, patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000

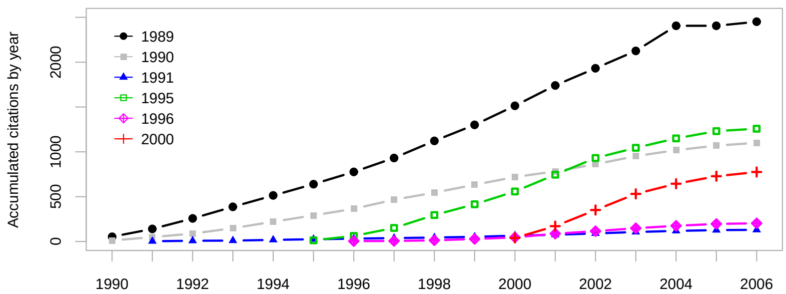


Figure 3.16: Accumulated citations received by year, patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000

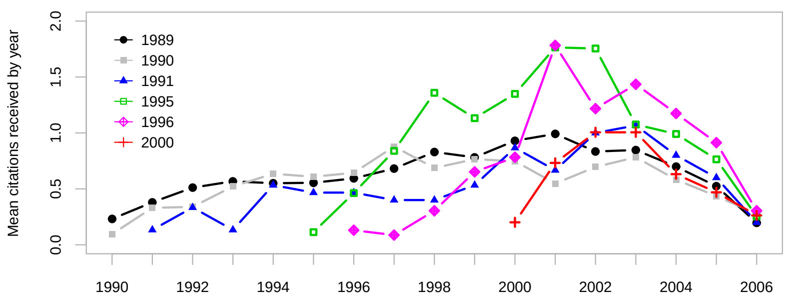


Figure 3.17: Mean citations received by year, patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000

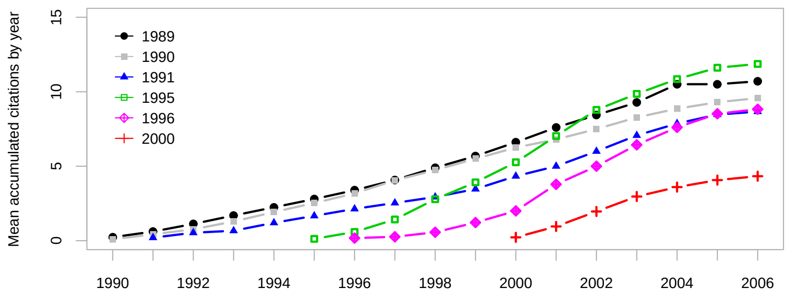


Figure 3.18: Mean accumulated citations received by year, patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000

Chapter 4

Structural Equation Models

Abstract. This chapter presents an overview of structural models with latent variables. Emphasizing a component-based approach, two approaches are described –covariance-based and partial least square path modelling– in terms of major scientific contributions in these subjects. In addition, this chapter provides guidelines for model specification and modelling process for structural equation models. Special emphasis is placed on determining the reflective or formative nature of measurement models.

4.1 Introduction

A variable is a characteristic measured or observed when an experiment is carried out or an observation is made (Upton, 2006). Observed characteristics are useful in describing the behavior of an interesting phenomena. What factors influence the behavior of people? How do societies change? What determines economic growth? How does public spending affect innovation? To study countries' economy, it is possible to use different measures such as income per capita, gross internal product or inflation index. These variables may be described and altogether give a sign of how good or bad the economy of a region is. Technological performance of a country can be studied by observing the evolution of its R&D investment, its number of scientific publications or its stock of patents. Analysis of these variables may provide insight into the reasons for a given change enabling better decision making.

Understanding how variables are related and how the strength of the relationships are among them, it is the foci of this thesis. Usually, real phenomena deal with complex variables. They represent a concept, an idea or a construct, and can only be indirectly estimated through other observed variables or measures. Nowadays, with the advent of this new era of technological development, there are a variety of "technologies" to describe a set of variables. Contributions come from different disciplines, ranging from the classical Bayesian theory –largely developed by statisticians– to the development of algorithms, where computer science community have contributed significantly.

From a social science standpoint, structural equation modelling (SEM) techniques are used to represent and describe relationships between constructs, and between constructs and observed variables. Structural equation models are a family of methods that deal with dependency relationships among variables studying their variance and covariance. They are called a second generation of multivariate analysis (Fornell, 1985). The first generation is mainly composed by Multiple Regression, Principal Components Analysis, Factor Analysis and Discriminant Analysis (Chin & Newsted, 1999). The advantage of the second generation tools is the flexibility to work with theory and data, providing the following key benefits (Bagozzi, 1980; Batista & Coenders, 2000; Wheaton *et al.*, 1977). From a theory and data perspectives:

- To approach the whole phenomenon and a more complete representation of a complex theory.
- To elucidate the theory due to the need for a clear definition of the constructs and their relationships.
- Assumption, constructs and hypothetical relationships are explicit.
- Causal relationships among constructs can be interpreted directly.
- SEMs provide a formal framework for building and testing theories and measures.

SEMs allow the definition of a large number of endogenous and exogenous variables, to simplify the multivariate matrixes and work with constructs estimated by indicators and then to assess the quality of this measurement. From a researcher's perspective, these techniques allow researchers to use their own discretion and knowledge specifying a model.

There are two main streams of research regarding SEM. The covariance- and the component-based approaches. Factorial Analysis and Linear Structural Relation Model (LISREL) are the most well-known covariance-based techniques. These and other models of covariance analysis involve procedures of parameters' estimation that seek to estimate the value of the variables and the strength of the dependency relationships among them, as well as to reproduce as nearly as possible the observed covariance matrix. Recently, soft modelling and component-based techniques, as Partial Least Squares (PLS) Path Modelling, have been increasingly used by researchers and practitioners. PLS aims the minimization of error in all endogenous constructs or equivalently the maximization of explained variance. Generally, researchers are familiar with Linear Structural Relation Model and apply it successfully. It is not so with PLS Path Modelling. But, in recent years, PLS Path Modelling has spread explosively and has blossomed a variety of approaches and applications coming from different disciplines.

Emphasizing a component-based approach, this chapter presents an overview of the main methods and algorithms for structural equation modelling from both component- and covariance-based approaches. A brief historical perspective, main contributions and advances are introduced.

4.2 Background

From a statistical theory and application perspectives[1], two main knowledge areas have generated synergies and made outstanding contributions to the development of structural models: psychometrics/psychology and econometrics.

4.2.1 Psychometrics and Factorial Analysis

The British psychologist Charles Spearman studied how to measure and objectively determine the –underlying and unobserved– human intelligence. Spearman (1904, 1927) found a general factor "g" of intelligence accounting for observed correlations between individual differences (observed variables). The one-factor model was presented in 1904. Since then, this method has become a widely used technique and the baseline of the covariance-based procedures. Factor analysis (FA) was disseminated in the U.S. by the mathematician Karl Holzinger and the psychologist Louis Leon Thurstone, but Truman Kelly and Thurstone (1931)[2] transformed Spearman's one Factor Analysis into Multi-Factor Analysis in the thirties and forties. Multiple common factors jointly account for intercorrelations between variables (or test scores).

At that time, estimation of factors models was frequently made by means of "a modified version of Principal Components [also called Principal Factor Method] and the Centroid Method" (Cudeck *et al.*, 2001, p. 34). Both procedures involve the computation of communalities, that is, that portion of the variance of the *ith*

[1] According to Bagozzi (1980), from a philosophical approach firsts in studying causal relations were David Hume, Immanuel Kant and John Stuart Mill. They defined the main characteristics of these relations: (a) their contiguity in time and place, (b) their temporal priority of cause and effect and (c) their constant conjunction. This paper does not seek to discuss the philosophical aspects of causation. Those researchers interested may consult the works of Bagozzi (1980), Pearl (2000), Woodward (2003) and references therein.

[2] Thurstone also contributed with the generalization of Spearman's tetrad analysis to the rank of the correlation matrix as the basis for determining the number of common factors. Another important contribution was made by Ledyard Tucker who proposed the Three-Mode Factor Analysis. The interested reader may review Cudeck & MacCallum (2007) for historical details.

variable contributed by the common factors. Common factors and communalities are obtained after model estimation, however. Jöreskog solved this problem in his PhD dissertation proposing a new method for extracting factors based on maximum likelihood estimation (see section 4.2.4).

4.2.2 Econometrics and Principal Components

Principal Components were introduced by the British statistician Karl Pearson in 1901 in order to factor a matrix of observable variables from a data reduction perspective[3]. Later on, it was the American economist Harold Hotelling (1933) who fully completed the formulation of Principal Component Analysis (PCA) as a method for the Factor Analysis of a correlation matrix. The main difference between PCA and FA is that the former does not include unobserved phenomena (or measurement errors). Instead, linear combinations of observed variables are formed. Hotelling (1936) also introduced the Canonical Correlation Analysis (CCA).

Econometricians have taken correlation and regression analysis from Pearson's work and the t-test and maximum likelihood estimation from Ronald Fisher's contributions[4]. Economical phenomena are determined by the relationship between a large number of economic variables. These relationships may be described by single equation models or a set of equations that must be estimated at the same time. Simultaneous equation models cannot be estimated using ordinary least squares method, because the assumption of zero covariance between disturbance terms and independent variables is not satisfied. Econometricians have intensively investigated to solve this issue. Among other contributions, the Norwegian economist Trygve Haavelmo clearly distinguished identification problems from estimation problems by the presence of simultaneous equations (Morgan, 1990), proposed a probabilistic approach to econometrics, and recommended maximum likelihood as a method of estimation (Haavelmo, 1947; Markus & Converse, 1979). In a parallel type of investigation –and in opposition to Haavelmo's ideas– the Swedish statistician and econometrician Herman Wold determined the conditions under which OLS estimates of simultaneous equation systems are consistent, and proposed an alternative approach for recursive models with observed and unobserved variables (see section 4.2.5). PLS Path Modelling has its origins in econometrics. As Areskoug (1982) remarked:

> The concept of latent variables in the PLS framework emerged from the estimation of reformulated interdependent systems [Mosback and Wold

[3]Both Pearson and Spearman were pupils of Francis Galton in his Anthropometric Laboratory in London; for more about Pearson, Spearman and the Anthropometric Laboratory, see Cudeck *et al.* (2001). There are a number of articles published by Galton in www.galton.com.

[4]See Morgan (1990) for a review of the origins and the evolution of econometrics.

Table 4.1: Comparison of Principal Component Analysis and Factorial Analysis

Principal Component Analysis	**Factor Analysis**
Karl Pearson (1901)	Charles Spearman (1904)
Natural scientific school	Social science school
No measurement errors	Measurement errors
No unobservable variables	Unobservable variables
To explain both common and unique variance of variables	To explain common variance of variables
To maximize the total variance of manifest variables	To explain off-diagonal correlations among manifest variables
Loadings tend to be larger	Loadings tend to be closer to population parameters (Cudeck & MacCallum, 2007)
Smaller standard errors for loadings	Larger standard errors for loadings (Ogasawara, 2003)
Component scores are unique	Component scores are not unique
No indeterminacy problems	Indeterminacy problems (Steiger, 1979, 1996)
Less computer intensive	More computer intensive (Velicer and Jackson, 1990)

> (1980)], where endogenous variables as regressors were substituted by their systematic parts obtained from other relations. Specifying estimated latent variables as systematic parts of OLS regressions, allows estimation to stay in the structural form of the model and avoids some of the identification problems connected with the reduced form. The reason for this is that less assumptions about the residuals need be made. Leaving the residual structure within blocks unspecified actually yields an unidentified model (p. 100).

Both Factor Analysis and Principal Component Analysis share striking similarities, since both techniques try to approximate the covariance between variables, although conceptually they are based on different models (see Table 4.1). The former looks for explain the common variance among manifest variables, whereas the latter seeks to maximize the total variance in a reduced-dimensional space.

Another important contribution for the development of structural equation models was made by the American geneticist and biometrician Sewall Wright (1918). The researcher defined three basic components of SEMs: (1) the path diagram, (2) the equations relating correlations or covariances to parameters and (2) the decomposition effects. These elements are the baseline in the modelling process allowing the expression of the covariances or correlations among variables

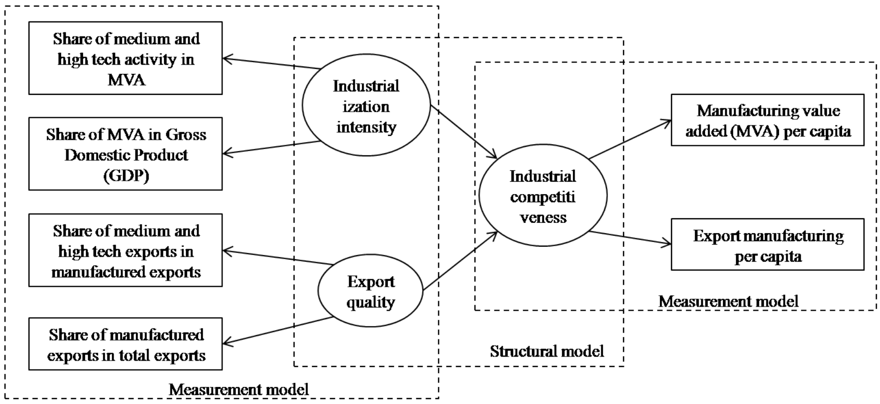


Figure 4.1: Path model with structural and measurement models. Industrialization intensity and export quality are modeled as determinant of industrial competitiveness. Their relationships are described by the structural model. Each latent variable is linked with measurement variables, emphasizing the factorial perspective.

as function of model parameters (Bollen, 1989). Path analysis and paths models are terms coined by psychometricians and econometricians, respectively, to refer structural equations models. SEMs formalize relations among latent and manifest variables, combining an econometric approach focused on prediction and a psychometric perspective focused on knowing the relationships among latent variables. Figure 4.1 shows an example of a path model with latent and manifest variables. Latent variables and their relationships are described by a structural model, while relationships between latent and manifest variables by measurement models.

4.2.3 Exploratory and Confirmatory Approaches

Both, Principal Components and Factorial Analysis are closely related to exploratory (theory building) and confirmatory (theory testing) perspectives. In this section, we tackle these two approaches from the point of view of the French, Anglo-Saxon and Dutch schools. French school is based on data analysis, Anglo-Saxon school is based on modelling, and Dutch school is placed on a middle point between the first two approaches. PLS procedures –with nordic roots– have brought the French school to the Anglo-Saxon school, integrating modelling into Exploratory Data Analysis. This is one of its great merits.

The English term Data Analysis includes both Exploratory Data Analysis and Confirmatory Data Analysis. This convention and the term "data analysis" were formalized by John Tukey in 1977: "Exploratory Data Analysis" to analyze data in order to formulate hypothesis to test, and "Confirmatory Data Analysis" focused on statistical hypothesis testing[5].

[5]Tukey's data analysis was mainly concerned with small samples and few variables (Gower, 2008).

From a exploratory perspective, a body of methods have been developed to explore and describe data in geometric/multidimensional spaces (orthogonal projections). Visual representations make it easier to understand data and their relationships. These exploratory multivariate data analysis techniques are based on singular value decomposition. They include Principal Component Analysis, Factor Analysis, Simple and Multiple Correspondence Analysis, Multiway Arrays and Generalized Canonical Correlation Analysis[6]. Some of these techniques are rooted in an exploratory approach, but contributions have been also made from a confirmatory data analysis framework.

Although Tukey coined the term, the French's school of "Analyse des Données" pioneered in the exploratory approach. The school was born in the 60s and 70s, mostly around developments of Jean Paul Benzécri, who introduced Correspondence Analysis and Taxonomy/Cluster Analysis (1969). French researchers exploited computational advances of those years and focused on the analysis of large data matrices, "letting data speak for themselves." Falguerolles (2008, p. 2) mentioned four elements that characterize the data analysis "á la francaise":

1. A firm belief in multidimensional descriptions.
2. A search for latent variables giving sense to the observed data and allowing dimension reduction.
3. A conviction that proper graphical representations best convey the structure of either the original data or the results of their analyzes.
4. A manifest (and possibly overplayed) claim from their authors to avoid any modelling driven by probabilistic considerations.

Formalizing a definition, exploratory analysis aims "to establish relationships between variables without giving more importance to any particular variable. It is to this family of methods that this issue is dedicated. Traditionally, in this phase of the study, the conclusions only concern the data that is analyzed and they are not inferred to a larger population. The exploratory analysis essentially reposes on graphical representations and on multidimensional description techniques" (Lebart, 2008, p. 3). PLS Path Modelling procedure tap into an exploratory approach.

French approach to data analysis contrast with the Anglo-Saxon school[7] which emphasizes confirmatory analysis and inferential. In a confirmatory approach, hypotheses are first formulated and they are validated/invalidated on the basis

[6]See the paper of Lebart (2008) for historical details. Only Single and Multiple Correspondence Analysis are rooted in the classical French school of the 70s.

[7]Anglo-Saxon school is mainly composed by scientist from the U.S. and England.

of statistical tests or probabilistic models. Some researchers from this school are Charles Spearman, Karl Pearson and Harold Hotelling.

A middle approach[8] is followed by the Dutch school of exploratory multidimensional data analysis, where scientists as Gerard Heymans, John van der Geer[9] and Jan de Leeuw –the latter led the Albert Gifi research group on data theory at the University of Leiden– have made important contributions to psychometrics, non linear multivariate analysis and correspondence analysis. They remark that there is not clear cut distinction between exploratory and confirmatory approaches. So, they focused on "building bridges between methods of l'analyse des données and statistical modelling methods" (Falguerolles, 2008). These researchers extended the French school, enriching the analysis including statistical modelling.

4.2.4 Covariance-based Approaches

As Bollen (1989) remind us, the works of the Royal Society of Edinburgh[10] (1940), Anderson & Rubin (1956) and Jöreskog (1969) helped lay the foundations for hypothesis testing in modern Factor Analysis. These three researchers incorporated almost simultaneously path diagram and path analysis in a structural modelling framework. However, the Swedish mathematician and statistician Karl Jöreskog has done one of the most important contributions to the development to structural modelling. Jöreskog first introduced the Confirmatory Factor Analysis in 1969 doing a psychometric application, and in 1970, a general method for Analysis of Covariance Structures (ACOVS) based on maximum likelihood estimation (Jöreskog, 1969, 1970). Jöreskog (1973), Keesling (1972, in Bollen 1989) and Wiley (1973) rose what is now known as structural equation modelling (SEM) or JKW models (Bollen, 1989). This technique rests on the assumption that the covariance matrix may be expressed in terms of a set of known parameters that are estimated using maximum likelihood. The goal is to minimize the difference between the sample covariance and the estimated covariance matrix making use of certain parameters (i.e., $\sum - \sum(\Theta)$). Thus, the process attempts to fit factor models to data. Among other reasons, this procedure has been widely used due to Jöreskog, together with Dag Sörbom, developed the LISREL computer program to implement it[11]. Nowadays, the name LISREL is sometimes used interchangeably to refer to the software or the statistical method.

Other researchers such as the sociologist Otis D. Duncan (1957), the psychol-

[8]Falguerolles (2008, p. 25) talked about "the between introduction of probabilistic models" in l'analyse des données.

[9]John van der Geer and Gerard Heymans –the latter from Groningen University– initiated the multidimensional approach to multivariate data in the Netherlands. See Heiser (2008).

[10]The Royal Society of Edinburgh is Scotland's National Academy of Science & Letters.

[11]Availability of a number of software such as LISREL, EQS, AMOS, SEPath, CALIS, RAMONA and LISCOMP, have facilitated the use of covariance-based techniques.

ogist Peter Bentler (1980; 1990; 1996), and the economists Arthur S. Goldberger (1973) and Dennis J. Aigner (1977) have extended these techniques in their own disciplines. In 1972, together with Goldberger, Jöreskog formalized the use of Generalized Least Squares (GLS) as an alternative to maximum likelihood (Jöreskog & Goldberger, 1972; Browne, 1977). Jöreskog and Goldberger also proposed procedures for Multiple Indicators and Multiple Causes of a single latent variable, the so-called MIMIC models (Jöreskog & Goldberger, 1975). Other important contributions are made by Browne (1984) who proposed estimations that assume arbitrary distributions, Bentler (1983) that suggested estimators that treat higher-order product moments of latent variables, and Muthén (1984) that generalized these models to ordinal or limited observed variables[12].

4.2.5 Component-based Approaches

In the early 60s and 70s, the Swedish statistician and econometrician Herman Wold wrote extensively on operative aspects of econometrics, sociological models and causal flows with latent variables. Among others, Wold proposed in 1973 the algorithm Nonlinear Estimation by Iterative Partial Least Squares (NIPALS) (Wold, 1973a,b, 1974, 1975). NIPALS was originally presented by Wold in 1966 for Principal Component Analysis with the name Nonlinear Estimation by Iterative Least Squares (NILES). The algorithm shows how principal components are extracted or how a matrix is factorized from a series of simple regressions by least squares. Hence, that using this technique, consistent estimators of parameters of a set of equations are found. This algorithm is the precursor of the PLS Path Modelling algorithm. Though the PLS design was completed in 1977 (Wold, 1982, p. 35), it was presented in 1979 in the article Partial Least Square Path Modelling with Latent Variables (Gerlach *et al.*, 1979). Herman Wold (1980) preferred the "soft modelling" for econometric modelling because this approach considers few cases and assumptions about data distribution, in contrast to LISREL, which assumes that data are multivariate normal distributed and where large sample sizes are required for their application ("hard modelling"). PLS Path Modelling has a partial nature because only a part of the model is involved at each iteration step of the algorithm.

Several methods and approaches have been proposed to the eaves of the PLS approach. Svante Wold, Harald Martens and Herman Wold proposed a particular case of PLS Path Modelling, the PLS Regression (Wold *et al.*, 1983; Martens *et al.*, 1983; Wold *et al.*, 2001). Extensively used in Chemometrics, PLS Regres-

[12] A summary of developments to the 80s can be found in Jöreskog & Sörbom (1982) and Bollen (1989). Both the covariance- and component-based perspectives are presented in the book System under Indirect Observation published jointly by Jöreskog & Wold (1982b). Recent approaches and further historical details can be found in Cudeck *et al.* (2001).

sion offers an alternative to Multiple Regression when independent variables are multicollinear. Recently, there has been a rapid progress in the development of PLS-based algorithms. Chemometrics and computer science scientific communities have made contributions to modelling nonlinearities between response and predictors; linear inner relationships have been replaced with quadratic polynomials (Wold *et al.*, 1989; Baffi *et al.*, 1999b), smooth bivariate spline functions (Frank, 1990; Wold, 1992), sigmoidal neural network functions (Qin & McAvoy, 1992; Baffi *et al.*, 1999a), kernel functions (Lindgren *et al.*, 1993; Rosipal & Trejo, 2001), radial basis functions (Wilson *et al.*, 1997), and feedforward neural networks (Malthouse *et al.*, 1997). After Rosipal & Trejo (2001) introduced the use of a kernel function, other researchers have proposed algorithms with classification purposes, such as PLS Logistic Regression and PLS Generalized Linear Regression (Bastien *et al.*, 2005), PLS for Discrimination (Pérez-Enciso & Tenenhaus, 2003; Barker & Rayens, 2003), Kernel PLS for Discrimination (Rosipal, 2003) and Kernel Logistic PLS (Tenenhaus *et al.*, 2007).

On the other hand, Jan-Bernd Lohmöller was one of the first researchers to work in the computational aspects (LVPLS 1.8), and theoretical developments of PLS Path Modelling (Lohmöller, 1989). Lohmöller proposed a matrix-based version of the algorithm[13] making easier its computational implementation (see section 5.3 for details on this). More recently, Wynne W. Chin has introduced a software with graphical interfaces (PLS Graph 3.0), improved the validation techniques, and extended the method in the information systems field (Chin, 1995; Chin & Marcolin, 1995; Chin, 1998a; Chin & Newsted, 1999; Chin *et al.*, 2003). Michel Tenenhaus has related PLS Path Modelling to Multi-Block Analysis and has stressed that the procedure is an alternative to handle missing values (Tenenhaus, 1998; Pagés & Tenenhaus, 2001; Tenenhaus *et al.*, 2005, 2007; Tenenhaus & Hanafi, 2010). Esposito-Vinzi has shown that both, PLS Regression and PLS Path Modelling, can be combined at technical levels (Esposito Vinzi & Lauro, 2005; Esposito Vinzi, 2007). Additionally, since 1999, Tenenhaus and Esposito-Vinzi have led the International Conference on Partial Least Square and Related Methods, the main international forum of PLS research, innovations and practical applications.

In the recent past, PLS Path Modelling scientific community has mainly focused on developing algorithms for multi-group analysis. Different approaches have been addressed, as permutation procedures (Chin, 2003; Henseler & Fassott, 2007; Chin & Dibbern, 2010), the segmentation tree algorithm PATHMOX (Sánchez & Aluja, 2006; Aluja & Sánchez, 2007; Sánchez & Aluja, 2007), the PLS typological path modelling routine PLS-TPM (Esposito Vinzi *et al.*, 2005, 2007; Squillacciotti,

[13]This procedure is described in Tenenhaus *et al.* (2005).

2005, 2007, 2010), the response-based units segmentation routine REBUS-PLS (Trinchera *et al.*, 2006; Trinchera, 2007; Esposito Vinzi *et al.*, 2008), the fuzzy PLS Path Modelling for latent class detection FPLS-LCD (Palumbo *et al.*, 2008), the PLS genetic algorithm segmentation PLS-GAS (Ringle & Schlittgen, 2007) and the finite mixture-PLS (FIMIX-PLS) procedure (Hahn *et al.*, 2002; Ringle *et al.*, 2005a, 2010). See Sarstedt (2008) for a non-technical comparison and a methodological taxonomy of these approaches.

Other lines of research have also been undertaken. Investigations have been concerned with convergence and consistency of the PLS Path Modelling algorithm (Hui & Wold, 1982; Dijkstra, 1983; Mathes, 1993; Schneeweiss, 1993; Hanafi & Qannari, 2005; Krämer, 2006; Hanafi, 2007; Henseler, 2009; Dijkstra, 2010), PLS Regression and PLS Path Modelling relationship (Esposito Vinzi *et al.*, 2005; Tenenhaus & Esposito Vinzi, 2005; Esposito Vinzi, 2007), the algorithm performance against multicollinearity, skewed distributions for observed variables and misspecification of structural models (Westlund *et al.*, 2001, 2008; Marcoulides & Saunders, 2006; Marcoulides *et al.*, 2009), the reflective versus formative specification of the measurement models (Chin & Gopal, 1995; Diamantopoulos & Winklhofer, 2001; Rossiter, 2002; Jarvis *et al.*, 2003; Diamantopoulos, 2006; Petter *et al.*, 2007; Diamantopoulos *et al.*, 2008), the algorithm performance with formative outer models and the tetrad analysis approach (Cassel *et al.*, 1999, 2000; Bucic & Gudergan, 2004; Ringle *et al.*, 2007; Westlund *et al.*, 2008; Gudergan *et al.*, 2008; Ringle *et al.*, 2009; Vilares *et al.*, 2010), the analysis of interaction and nonlinear effects among constructs with PLS Path Modelling Mode A (Chin *et al.*, 2003; Henseler & Fassott, 2005; Goodhue *et al.*, 2006, 2007; Henseler *et al.*, 2007; Qureshi & Compeau, 2009; Henseler & Fassott, 2010; Henseler & Chin, 2010), the robustness of the algorithm compared with covariance-based models (Fornell & Bookstein, 1982; Schneeweiss, 1991; Hsu *et al.*, 2006; Ringle *et al.*, 2007; Almeida *et al.*, 2007; Ringle *et al.*, 2009), the modelling of hierarchical constructs (Wetzels *et al.*, 2009), and the non-supervised model building with PLS Path Modelling (Marcoulides, 2003; Jakobowicz & Derquenne, 2007).

On the other hand, though PLS Path Modelling has its origins in econometrics, since the algorithm was proposed, has been extensively used in the social sciences. Investigations by Bagozzi (1980), Fornell & Larcker (1981), Bookstein (1982), Fornell & Bookstein (1982), Fornell (1992), and Fornell & Cha (1994) helped to spread the structural models in marketing, where PLS Path Modelling has been widely used (see the recent article of Henseler *et al.* (2009) on PLS Path Modelling applications in international marketing and references therein). Additionally, the procedure has been applied in a number of disciplines such as strategic management (Cool *et al.*, 1989; Hulland, 1999), information systems (Barclay *et al.*, 1995; Goodhue *et al.*, 2006), e-business and finances (Serrano-Cinca *et al.*, 2007; Sohn *et al.*, 2007), industrial management (Hadaya & Cassivi, 2007), and group and organization management (Bucic & Gudergan, 2004; Sosik *et al.*, 2009). The availability of software, such as XLSTAT-PLSPM, LVPLS developed by Lohmöller, PLS-Graph 3.0 developed by Wynne W. Chin, PLSPM in R-Project by Gastón Sanchez and SmartPLS (Ringle *et al.*, 2005b) have facilitated the use of this approach[14]. It is worth mention, that LVPLS, PLS-Graph and SmartPLS have implemented the Lohmöller procedure of the PLS Path Modelling algorithm.

This updated review of the literature shows that today PLS techniques have taken root in various knowledge fields. However, there is little research reporting (1) the robustness and performance of PLS Path Modelling with Mode C and few indicators per construct, and (2) the robustness and performance of a Two-Step PLS Path Modelling approach for estimating nonlinearities among formative constructs. It is in these aspects where this work is aimed to make contributions.

4.3 Conventional Rules for SEMs

Models describe relationships between variables; these may represent general concepts that cannot be directly observed. Thus, they are named constructs, factors, latent variables, unobserved or unmeasured variables. Examples of latent variables are intelligence, human development or the value of something. An structural model describes relationships between latent variables. Unmeasured variables can be estimated indirectly through other variables known as measurement, manifest or observed variables, and also as indicators. Measurement models describe relationships between latent and measurement variables. A path diagram consists of a visual representation of latent and manifest variables and their relationships, which facilitates the understanding of a phenomenon and model parameters specification (Jöreskog & Wold, 1982a; Batista & Coenders, 2000). A set of circles, squares and arrows are used to display concepts, indicators and relationships, respectively (Figure 4.2). Hence, in these diagrams it is also possible to observe the dependency or causality relationships that govern the modeled phenomenon. From a PLS-based approach, conventions are as follows (see Figure 4.3):

- Latent variables (LVs), constructs, unobserved or unmeasured variables are represented by circles; they may contain a disturbance term.
- Manifest variables (MVs), observed or measured variables are represented by squares; they may contain random or systematic measurement error.
- Causal relationships among variables are indicated by unidirectional arrows

[14] See Temme & Kreis (2005) for a comparison of PLS software to that date (LVPLS 1.8, PLS-GUI 2.0.0, PLS-Graph 3.00, SPAD-PLS, SmartPLS 1.0).

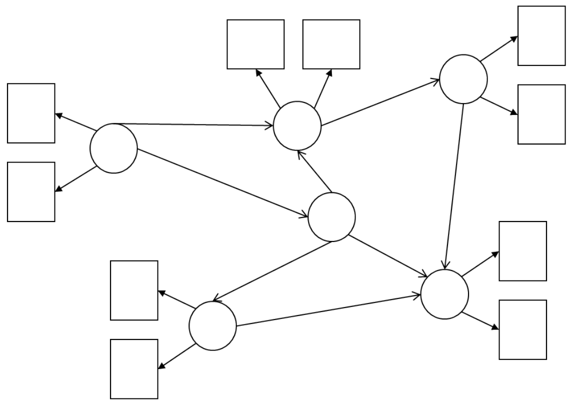


Figure 4.2: Path diagram, visual representation of variables and their relationships.

(recursive relationship) from latent to manifest variables in the case of reflective models, and from manifest to latent variables in the case of formative models.

- Endogenous latent variables are affected by a random disturbance term that is included in the diagram as an additional arrow pointing to the endogenous variable.

4.4 Modelling Process

Modelling process comprises conceptual and methodological considerations (Wheaton *et al.*, 1977; Wold, 1980; Bagozzi, 1984; Bollen, 1989; Hulland, 1999). Researchers begin the process of defining a theoretical model, identifying the concepts that best represent the phenomenon under study, and specifying whether these concepts can be represented by latent variables to be estimated by means of manifest variables. Indicators are related to latent variables forming or reflecting the constructs; variables and relationships are represented graphically by a path diagram.

Implicitly, conceptual modelling stage leads to the definition of the dimensionality problem by taking into account the differences between constructs and measures and the causality relationships among variables. Wold (1985, p. 582) emphasizes that "using prior knowledge and intuition the investigator is free to specify the LVs, to design the inner relations, and to compile a selection of indicators for each LV." The path model "is usually tentative since the model construction is an evolutionary process. The empirical content of the model is extracted from the data, and the model is improved by interactions through the estimation between the model and the data and the reactions of the researcher" (Wold, 1980, p. 70). Wheaton *et al.* (1977) highlights the importance of doing a good model specifica-

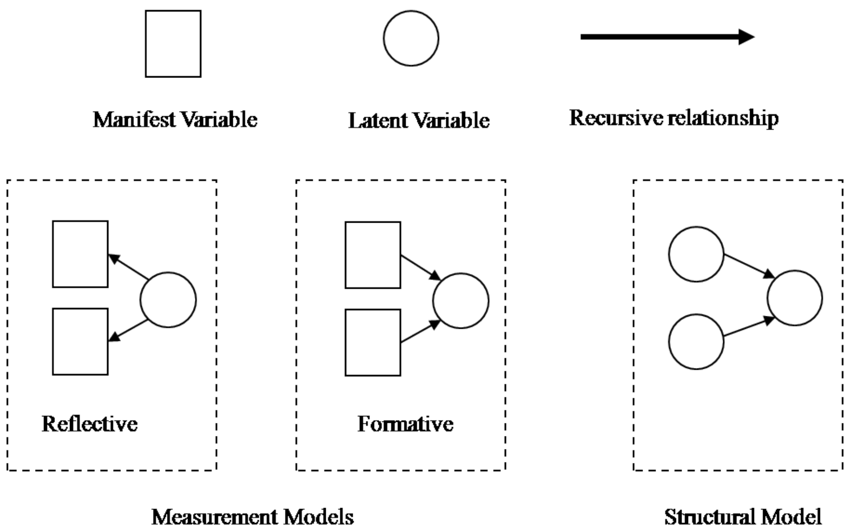


Figure 4.3: Graphical conventions of path models

tion, considering all latent and measurement variables that may explain a event because this influences the model's estimation results. From a methodological standpoint, the modelling process aims:

1. To determine the model and relationships between observed and unobserved variables.

2. To interpret the path coefficients.

3. To assess the reliability and validity of measures.

4. To select a plausible final model.

On the other hand, some strategies are given by Chin & Newsted (1999, p. 337) and Wold (1980, 1982) to decide if PLS Path Modelling is a suitable procedure for estimating a structural equation model. They are summarized as follows:

- To determine the modelling orientation: prediction versus parameter estimation[15].

- To determine the prior knowledge about a phenomenon, if it is relatively new or changing and if the theoretical model or measures are not well formed.

- To determine the complexity of the model, the numbers of indicators and/or latent variables.

[15] A covariance-based approach should be preferred if the objective is to estimate the population parameters.

- To determine data conditions relating to normal distribution, independence, and/or sample size.

To specify a model, a first attempt to define latent and manifest variables may be done through an extensively review of the literature on the phenomenon being studied. In addition, relationships in each measurement model must be determined. Traditionally, reflective relationships have been preferred by researchers. Especially important for the development of this thesis is the distinction between reflective and formative measurement models. Hence, a literature review on these topics is presented in what follows.

4.5 Reflective versus Formative Constructs

The distinction between reflective and formative measurement models for structural equation models (SEM) is an issue that has been addressed by several scientific communities. Major contributions have been made by researchers from statistics (Cohen *et al.*, 1990), psychology and sociology (Bollen & Ting, 1993, 2000), information science (Chin & Newsted, 1999; Petter *et al.*, 2007), and business and marketing research (Fornell & Bookstein, 1982; Diamantopoulos & Winklhofer, 2001; Jarvis *et al.*, 2003; Bucic & Gudergan, 2004; Gudergan, 2005). In general, the literature is diverse and contributions come from researchers using covariance- and component-based approaches. We are interested in both perspectives, but the emphasis is placed on Partial Least Squares (PLS) Path Modelling. Although the literature review does not seek to be exhaustive, we attempt to present the main contributions that have been made and which are related to formative outer models.

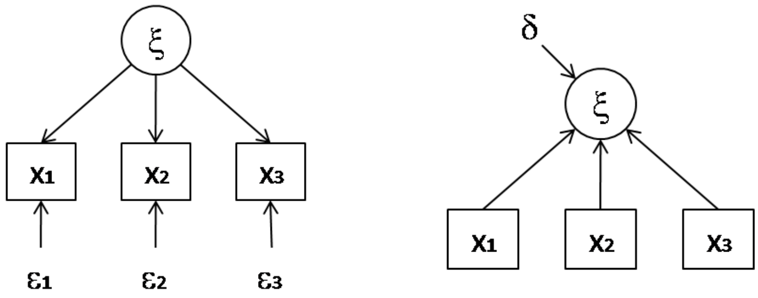
Depending on its nature, manifest variables have been referred to as effects or reflective indicators, or causes or formative indicators. We prefer to refer to reflective and formative measurement models or reflective and formative relationships or constructs, thus emphasizing the link between observed and unobserved variables. Wold's basic design usually refers to Mode A and Mode B, or simple and multiple regressions (Wold, 1985). Although formative measurement models were first discussed by Curtis & Jackson (1962) and Blalock (1964), and a number of variables can be modeled in a better way through formative relationships (Hulland, 1999), measurement variables have been traditionally modeled in a reflective mode. Several authors have reviewed the scientific literature investigating the advisability of this practice. They have found that a number of articles indeed misspecified formative constructs and this has had an impact on the quality of results and conclusions that can be inferred from the estimated models. So, roadmaps based on decision rules or statistical procedures have been provided for some authors for proper definition of formative constructs.

4.5.1 Some Definitions

Reflective relationships seek to represent variances and covariances between the manifest variables that are generated or caused by a latent variable. So, observed variables are treated as an effect of unobserved variables (Cohen *et al.*, 1990; Bollen & Lennox, 1991). In a reflective measurement model, the manifest variables are measured with error (Figure 4.4(a)). Alternatively, formative relationships are used to minimize residuals in the structural relationship (Fornell & Bookstein, 1982), and here, manifest variables are treated as forming the unobserved variables (see Table 4.2 for a comparison of the differences between the reflective and formative outer models). MacCallun & Browne (1993) said that observed variables in a formative model are exogenous measured variables. In a formative outer model the manifest variables are presumed to be error-free and the unobserved variable is estimated as a linear combination of the manifest variables plus a disturbance term, so they are not true latent variables (Figure 4.4(b)). As in this case all variables forming the construct should be considered, the disturbance term represents all those non-modeled causes (Diamantopoulos, 2006).

Traditionally, the relationships between manifest and latent variables have been assumed to be reflective. There may be some reasons for this. (1) A theoretical definition of the model may impose the reflective relationships between the variables. That is, past empirical evidence drives the researchers to define the manifest variables as a reflection of the latent variables and not vice-versa. Additionally, (2) the classical test theory, which includes factorial analysis and maximum likelihood covariance structure analysis, estimates the models assuming that the variances and covariances between the manifest variables are caused by an underlying construct. Fornell & Bookstein (1982) and Chin (1998b) pointed out that modelling formative modes using a covariance-based approach may lead to identification problems and Heywood cases. So, researchers may tend to define outer models as reflective.

There are many relationships that may be modeled as formative instead of reflective. This is for instance, the relationship between education, occupational prestige and income as indicators of socioeconomic status (SES); or job, divorce, recent accident and death in the family for the latent variable life stress (Cohen *et al.*, 1990; Chin & Newsted, 1999). In these examples, the observed variables trigger the socioeconomic status or the life stress of a person and not vice-versa. This may have something to do with the temporary situation or the order in which events occur. In a formative relationship, one might think that the observed variables are generated first and that from these variables the construct is generated at a later stage. In the SES example, education, prestige and income are variables



(a) Reflective model (b) Formative model

Figure 4.4: Reflective and formative measurement models (focus on component-based approach)

that are developed from a certain moment in time and over time, and then they could generate the socioeconomic status of a person[16]. This is similar to what Bollen & Ting (2000) refer to when they talk about "causal priority" between an indicator and a latent variable or the "surplus meaning" of Jarvis *et al.* (2003).

To model a relationship as reflective when indeed it is formative has several consequences. An incorrect definition of the relationships between variables leads to a misspecification of the model, and thus to an inaccurate interpretation of the results[17]. There are studies that have provided criteria to determine whether a measurement model is reflective or formative, both in a covariance- and component-based approach. Several authors have focused on decision rules (Cohen *et al.*, 1990; Chin & Newsted, 1999; Diamantopoulos & Winklhofer, 2001; Jarvis *et al.*, 2003; MacKenzie *et al.*, 2005; Petter *et al.*, 2007) and others have proposed statistical procedures (Bollen & Ting, 1993, 1998, 2000; Gudergan *et al.*, 2008). Some of the presented procedures can be targeted to meet the needs of a particular area, since many of them have been conceived within psychological, sociological, and information systems, as well as within the marketing and business scientific communities.

4.5.2 Decision Rules Criteria

Interestingly, the fundamental contributions to distinguish between reflective and formative measurement models were made several decades ago. Fornell & Bookstein (1982), Cohen *et al.* (1990), Bollen & Lennox (1991), and McDonald (1996) have provided the basic guidelines to ensure a proper definition of the nature of a measurement model, and their results have encouraged the development of new

[16]See Cohen *et al.* (1990) for a discussion of the reflective or formative nature of SES.
[17]See the literature review for marketing research at Jarvis *et al.* (2003), for information system research at Petter *et al.* (2007), and a summary at Diamantopoulos *et al.* (2008).

Table 4.2: Comparison of reflective and formative measurement models

Reflective outer model	**Formative outer model**
The observed variables are treated as an effect of the unobserved variables	Observed variables are treated as forming the latent variable
Latent variables cause a change in the manifest variables	Observed variables cause a change in the latent variable
Latent variable is first revealed in time	Observed variables are first revealed in time
It seeks to represent variances and covariances between the observed variables that are generated or caused by a latent variable	Observed variables are used to minimize residuals in the structural relationships
It is possible to say something about the validity of the block of variables	It is not possible to say something about the validity of the block of variables (Bollen & Lennox, 1991)

contributions[18]. Moreover, these authors discuss how several common practices may influence the parameters estimation. The theoretical and empirical guidelines to distinguish between a reflective and a formative outer model, and thus make an adequate model specification, can be summarized in five points, as follows (see also Table 4.3).

(1) Strong theory and the goals that are pursued

All researchers agree with the importance of a strong theory in the model specification (Fornell & Bookstein, 1982; Cohen *et al.*, 1990; Bollen & Lennox, 1991). The theory and the previous knowledge of a phenomenon under study should help to clarify the generative nature of the construct (Cohen *et al.*, 1990). Considering an approach to index construction from a marketing perspective, Diamantopoulos & Winklhofer (2001) recommend specifying the content domain that the construct attempts to capture, and to examine the causal priority between observed and unobserved variables. When a formative relationship is considered, manifest variables must cover the entire scope of construct. Jarvis *et al.* (2003) suggest analyzing the nomological net of the manifest variables that is to determine whether the observed variable shares the same backward and forward variables. Moreover, Jarvis

[18]See the studies of Diamantopoulos & Winklhofer (2001), Jarvis *et al.* (2003), Bucic & Gudergan (2004), Petter *et al.* (2007), Gudergan *et al.* (2008), and Marcoulides *et al.* (2009).

et al. (2003) and Diamantopoulos (2006) pointed out that constructs have a surplus meaning that also have to be considered. That is because a latent variable in a reflective mode exists previously and independently of its manifest variables. In contrast, a construct in a formative mode does not exist without its observed variables and in this case, the surplus meaning is associated with the disturbance term.

(2) Correlations among manifest variables

The most important condition for reflective outer models is that all manifest variables that are considered for measuring a construct have to explain it (internal consistency). Correlation between manifest variables depends on the magnitude of correlations between manifest variables and the latent variable that they measure (reliability, Bollen & Lennox (1991)). So, this means that observed variables have to be positively and highly correlated with one another. On the other hand, in a formative outer model, manifest variables do not have to be especially correlated (Fornell & Bookstein, 1982). If a construct changes, the observed variables that are related do not have to change simultaneously and the model does not need to explain variance and covariances between observed variables. In a formative mode, there is not a procedure to assess the consistency and reliability of the block of manifest variables.

(3) Within-construct correlations versus between-construct correlations

When the model is being specified, a common practice among researchers is to test the within-constructs and between-constructs correlations. The applied condition is that the former should be greater than the latter. This is usually tested by means of a cross-validation technique (Cudeck & Browne, 1983). However, Bollen & Lennox (1991) show that this may lead to an incorrect indicator selection for reflective and formative outer models, because this rule may have exceptions. The researchers clearly demonstrated this showing that in some cases when the correlation between two latent variables is greater than zero, the between-construct exceeds the within-construct correlations.

(4) Sample size and multicollinearity

It is well known that sample size and indicator multicollinearity affects the stability of indicator coefficients (Fornell & Bookstein, 1982). Recall that multicollinearity is a frequent problem in multiple regressions. If X is the data matrix, parameter inversion of the regression matrix requires the inversion of X'X. If one of the explanatory variables is exactly a linear combination (collinear with the rest) or is highly correlated with the other, the matrix will have a range smaller

Table 4.3: Criteria for correlations and interchangeability of manifest variables in reflective and formative outer models

Criteria	**Reflective outer model**	**Formative outer model**
Correlations	Positively and highly correlated (unidimensionality)	High, moderate or low correlations
	Difficult to separate the impact of manifest variables on latent variables	Positive, negative or no correlations
		Multicollinearity affects the stability of indicator coefficients
Interchangeability	They are interchangeable	Not to be interchangeable
	Low effect if omitting an indicator	Serious effects if omitting an indicator

than k+1 (number of parameters), X'X will be singular and the equation system that determines the parameters will not have a unique solution. The consequences of this situation are widely known; the regression coefficients are unstable and may not be significant. So, the regression equation is difficult to interpret due to the erratic signs of regression coefficients (Tenenhaus, 1998). Fornell & Bookstein (1982) and Diamantopoulos & Winklhofer (2001) recommend that when observed variables are collinear, "one might estimate mode B but use loadings, rather than regression weights, for interpretation" (Fornell & Bookstein, 1982, p. 442). However, PLS Regression may be applied in case of multicollinearity.

(5) Interchangeability

Interchangeability refers to whether or not the manifest variables share the same concept (Diamantopoulos & Winklhofer, 2001; Jarvis *et al.*, 2003). All manifest variables in a reflective model explain the same construct. So, removing an indicator from the block of variables should not have a significant effect on the construct. The situation is completely different when considering formative outer models. The indicators do not have to be interchangeable or share the same concept. That is what Bollen & Lennox (1991) called "sampling facets of a construct"; in other words manifest variables of a formative block of variables should represent all the aspects that form the concept. These authors also point out that "omitting an indicator is omitting a part of the construct ... With causal indicators we need a census of indicators, not a sample. That is, all indicators that form the latent variable should be included" (p. 308).

4.5.3 Statistical Procedures

The tetrad analysis approach

The term "tetrad" was introduced by Spearman (1904, 1927). He noted that when it is possible to find a common factor to explain a set of manifest variables, certain pairs of product covariances between manifest variables are zero. He called the analysis of this phenomenon "exploratory tetrad analysis." Much later, Bollen & Ting (1993, 1998, 2000) introduced the "confirmatory tetrad analysis" (CTA) in structural equation models to differentiate reflective from formative measurement models. If a, b, c and d are a set of four manifest variables, three tetrads can be computed as $\tau_{abcd} = \sigma_{ab}\sigma_{cd} - \sigma_{ac}\sigma_{bd}$, $\tau_{acdb} = \sigma_{ac}\sigma_{db} - \sigma_{ad}\sigma_{cb}$, and $\tau_{adcb} = \sigma_{ad}\sigma_{cb} - \sigma_{ab}\sigma_{db}$, where σ denotes population covariance. Bollen & Ting (1993, p. 148) said that "the structure of each model often implies population tetrads that should be zero [vanishing tetrad]... Significant nonzero tetrads for the model implied vanishing tetrads cast doubt on the appropriateness of the model." The steps suggested by Bollen & Ting (2000, p. 5) to perform a tetrad test are: (1) specify the most plausible models of the relations between indicators and latent variables, (2) identify the model-implied vanishing tetrads for each model, (3) eliminate redundant vanishing tetrads, and (4) perform a simultaneous vanishing tetrad test.

From a PLS perspective, the first person to turn his attention to tetrad analysis was Chin (1998b) who recommended examining the TETRAD II methodology developed by Scheiness *et al.* (1994). These researchers used tetrad analysis to develop an algorithm that, given data, automatically identify causal relationship discovering causal patterns. The approach is completely exploratory and in those years "the IS field was still in the formative stages" (Chin, 1998b, p. xii). However, Bucic & Gudergan (2004), Gudergan (2005) and Gudergan *et al.* (2008) were the first in using the confirmatory tetrad analysis (CTA) as a test to distinguish between a reflective and formative measurement model in a component-based approach. CTA-PLS procedure is similar to that followed by Bollen & Ting (1993) and tetrads should not vanish in formative outer models. Additionally, two important issues are pointed out by Gudergan *et al.* (2008); first "neither CTA-SEM nor CTA-PLS are applicable for correlations or covariances close to zero in the measurement model" (p. 1243), and second when an outer model has less than four observed variables, this procedure requires adding manifest variables from other measurement models.

Chapter 5

Partial Least Squares Approaches

Abstract. This chapter presents the partial least squares (PLS) approach for modelling latent variables. The Wold's basic design and Lohmöller's implementation are examined in detail. Recent advances about the sensitivity of the algorithm to starting values, weighting schemes, and consistency are discussed. Convergence and validation techniques are examined as well. We begin by introducing the nonlinear estimation by iterative PLS and the PLS regression algorithms.

5.1 Nonlinear Estimation by Iterative Partial Least Squares

PLS Path Modelling has its origins in the Nonlinear Estimation by Iterative Partial Least Squares (NIPALS) algorithm (Wold, 1973a). In an iterative process, this algorithm takes principal components out computing simple regressions by least squares. Let $X = x_{ji}$ be a matrix of observations on p predictor variables of rank a. Variables, $X_1, X_2, ..., X_p$, are centered. NIPALS algorithm may be formalized as follows (Tenenhaus, 1998):

Step 1: $X_0 = X$
Step 2: For $h = 1, 2, ..., a$:
  Step 2.1: t_h = first column of X_{h-1}
  Step 2.2: To repeat until convergence of p_h
    Step 2.2.1: $p_h = X'_{h-1} t_h / t'_h t_h$
    Step 2.2.2: To standardize p_h to 1
    Step 2.2.3: $t_h = X_{h-1} p_h / p'_h p_h$
  Step 2.3: $X_h = X_{h-1} - t_h p'_h$

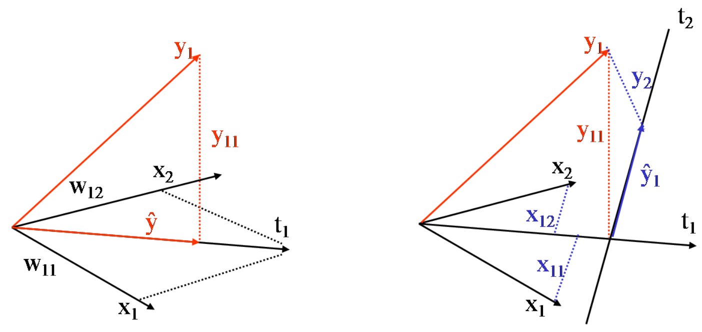
NIPALS leads to principal component analysis when examining a matrix with no missing data. Otherwise, t_h component's and p_h vector's estimates describe the data matrix X, estimating missing values.

5.2 Partial Least Squares Regression

Multicollinearity is a frequent problem in multiple regressions. If X is a matrix of observations on p explanatory variables, inversion of regression parameters requires the inversion of the matrix $X'X$. If one of the regressors is a linear combination of the others, that is collinear with the rest, the matrix rank will be smaller than the number of parameters $(k+1)$, $X'X$ will be singular and the equations determining the parameters will not have a unique solution (Peña, 1998). But, short of this extreme case, multicollinearity occurs when some or all independent variables are highly correlated to each other. Consequences are known, unstable and non-significant regression coefficients, and difficulty in interpreting the regression equation due to erratic signs of regression coefficients (Tenenhaus, 1998). In addition, multiple regression can use many factors. But if the number is too large –for example larger than the number of observations– the model perfectly will fit the sample, but it will fail in predicting new data, overfitting them (Valencia & Díaz-Llanos, 2003).

PLS regression is an alternative to multiple regression when predictors are collinear. It is an iterative algorithm that computes regressions by least squares to extract a series of orthogonal components. These components aims to explain the independent variables at the same time being related to the response variables. In each successive step, residuals are minimized until the algorithm reaches convergence. Let $X = x_1, x_2, ..., x_p$ be a matrix of observations on p explanatory variables of rank a and $Y = y_1, y_2, ..., y_q$ a matrix of observations on q response variables. The classic PLS regression algorithm can be expressed as follows (Tenenhaus, 1998):

Step 1: $X_0 = X; Y_0 = Y$
Step 2: For $h = 1, 2, ..., a$:
  Step 2.1: $u_h =$ first column of Y_{h-1}
  Step 2.2: To repeat until convergence of w_h
    Step 2.2.1: $w_h = X'_{h-1}u_h/u'_h u_h$
    Step 2.2.2: To standardize w_h to 1
    Step 2.2.3: $t_h = X_{h-1}w_h/w'_h w_h$
    Step 2.2.4: $c_h = Y'_{h-1}t_h/t'_h t_h$
    Step 2.2.5: $u_h = Y_{h-1}c_h/c'_h c_h$



(a) First component (b) Second component

Figure 5.1: Geometrical representation of components extraction by PLS regression

Step 2.3: $p_h = X'_{h-1}t_h/t'_h t_h$
Step 2.4: $X_h = X_{h-1} - t_h p'_h$
Step 2.5: $Y_h = Y_{h-1} - t_h c'_h$

This algorithm delivers five data matrices, (1) the matrix W_{hn} with the regression coefficients of X on components u, (2) the matrix C_h with the regression coefficients of Y on components t, (3) the matrix P_h with the regression coefficients of X on components t, (4) the matrix T_h with the PLS components t_h, and (5) the matrix U_h with the PLS components u_h.

Figures 5.1(a) and 5.1(b) show a geometrical representation of the algorithm when two explanatory variables are taken into account. The first component t_1 is computed as a linear combination of the independent variables, $w_{11}x_1 + w_{12}x_2$, or what is the same thing, as the orthogonal projection of the response variable y_1 in the plane formed by the predictors, x_1 and x_2. So, the dependent variable may be estimated as a function of this component plus a residual, $y = c_1w_{11}x_1 + c_1w_{12}x_2 + y_{11}$. Then, the residual, y_{11}, is orthogonally projected into the space formed by the residuals (x_{11} and x_{12}) from the regressions of each independent variable, x_1 and x_2, on the first component t_1. Thus, the second component t_2 is found. This procedure is repeated until convergence; hence, the residuals are minimized. If the first two components are taken into account ($H = 2$), the regression equation of y_k on the components t_1, t_2 is written as:

$$y_k \approx c_{1k}t_1 + c_{2k}t_2 \tag{5.1}$$

Components t_h may be expressed in terms of variables x_k:

$$t_h = w^*_{h1}x_1 + w^*_{h2}x_2 \tag{5.2}$$

Finally, PLS regression computed from two components is formalized as follow:

$$y_k \approx c_{1k}(w_{11}^* x_1 + w_{12}^* x_2) + c_{2k}(w_{21}^* x_1 + w_{22}^* x_2) \tag{5.3}$$

$$y_k \approx (c_{1k} w_{11}^* + c_{2k} w_{21}^*) x_1 + (c_{1k} w_{12}^* + c_{2k} w_{22}^*) x_2 \tag{5.4}$$

$$y_k \approx b_{k1} x_1 + b_{k2} x_2 \tag{5.5}$$

5.3 PLS Path Modelling Design

PLS Path Modelling –in contrast to PLS regression– works with unobserved variables. It is a soft modelling technique and a data analytic tool for estimating structural equations models and building a sequence of latent variables. There are two main implementations of the algorithm as proposed by Wold (1982, 1985) and Lohmöller (1989). This section mainly refers to the basic design of the algorithm as introduced by Herman Wold in 1985, and put forward by Lohmöller (1989) and Tenenhaus *et al.* (2005). Recall that Wold's procedure was presented in 1985 for a particular case of six blocks of variables and only for the centroid weighting scheme. Hence, Lohmöller's procedure and some extensions, as proposed by Hanafi (2007), are also reviewed.

5.3.1 Structural or Inner Model

The structural model or inner model, also called inner relations and substantive theory, describes relationships among latent variables by means of multiple regressions, that is through linear functions. Usually, structural model specification is supported by theory. PLS Path Modelling explores if variables and relationships are hold up or whether other theory-based specifications, that may be proposed, help better explain a particular phenomenon. Equation 5.6 describes relationships among latent variables. ξ_i and ξ_j are the exogenous and endogenous latent variables, respectively, and β_{ji} are the parameters that measure the relationship among constructs. They are called path coefficients. The condition imposed by Herman Wold is predictor specification, $E(\xi_j/\xi_i) = \sum_i \beta_{ji}\xi_i$, that is, there is no linear relationship between predictor and residual. This condition implies that $E(\nu_j/\forall\xi_i) = 0$, and $cov(\nu_j, \xi_i) = 0$.

$$\xi_j = \beta_{j0} + \sum_i \beta_{ji}\xi_i + \nu_j \tag{5.6}$$

5.3.2 Measurement or Outer Model

Measurement or outer models, also called outer relations, may be modeled in two different ways. Relationships between manifest and latent variables may be reflective or formative depending on whether the indicators are reflecting or forming the construct, respectively[1]. Normally, PLS Path Modelling considers the relationships between manifest and latent variables as follows.

Reflective outer models. Manifest variables revealing or reflecting the effect of a construct are modeled as indicators of it in a reflective measurement model. Each manifest variable x_{jh} is related by simple ordinary least squares regression with the underlying construct ξ_j (Equation 5.7).

$$x_{jh} = \lambda_{jh0} + \lambda_{jh}\xi_j + \epsilon_{jh} \tag{5.7}$$

The parameters λ_h are called loadings and determine the extent to which each indicator reflects a construct; ξ_j is a common factor with mean m, standard deviation one and indirectly observable by the manifest variables. The condition imposed by Herman Wold is predictor specification, $E(x_h/\xi) = \lambda_{h0} + \lambda_h\xi$. This condition implies that ϵ_h has zero mean, and it is uncorrelated with ξ_j. Moreover, the basic design of Herman Wold assumes that the covariance matrices of all ϵ_j are diagonal (Dijkstra, 1983).

As in a reflective model, where all the indicators of the block of variables reflect the same construct, there should be high collinearity among these variables. That is, the blocks of variables must be one-dimensional. Three methods are commonly used to check this constraint: principal component analysis, the classical Cronbach's alpha and the Dillon-Goldstein's ρ.

Formative outer models. Here, the latent variable is formed by a set of manifest variables as a linear function of them plus a residual (Equation 5.8).

$$\xi_j = \sum_h \pi_{jh}x_{jh} + \delta_j \tag{5.8}$$

The parameters π_h are called weights and determine the extent to which each indicator contribute to the formation of the constructs. Each block of manifest variables may be multidimensional, and multicollinearity among indicators is not a necessary constraint. The condition imposed by Herman Wold is predictor specification $E(\xi/X_1, ..., X_{pj}) = \sum_h \pi_h x_h$. This condition implies that the residual δ has zero mean, and it is uncorrelated with the manifest variables x_h. Since each latent variable is formed by a linear combination of the manifest variables, the sign of each weight π_h should be the same sign as the correlation between x_h and ξ (Tenenhaus *et al.*, 2005).

[1] See Chapter 4 for a detailed discussion about reflective and formative nature of measurement models

5.3.3 PLS Path Modelling Algorithm

The PLS Path Modelling algorithm is structured in three stages (Wold, 1980, 1982, 1985). The first stage computes the case values of the latent variables, the second stage focuses on the inner and outer relationships, and in the third stage, location parameters of the latent variables, λ_{jh0} and β_{j0}, are estimated. Only the first stage is iterative. The Wold's procedure algorithm is as follows.

The first stage. The algorithm starts choosing an arbitrary weight vector –outer weights– to first relate each latent variable with their own manifest variables. Usually this vector is a vector of ones. Each standardized latent variable Y_j –zero mean, unit variance– is computed as an exact linear combination of their own centered manifest variables:

$$Y_j = \sum w_{jh} x_{jh} \tag{5.9}$$

where w_{jh} are called the outer weights.

An auxiliary latent variable[2] Z_j is introduced as a counterpart to the variable Y_j. Each Z_j is computed as a weighting sum of the latent variables which is related to:

$$Z_j \propto \sum e_{ji} Y_i \tag{5.10}$$

where e_{ji} are called the inner weights. There are three different weighting schemes that may be used to compute e_{ji}: the centroid, the factorial and the path weighting schemes. The first one was introduced by Wold, and the last two by Lohmöller (1989).

The simplest scheme is the centroid scheme where the e_{ji} are equal to the signs of the correlations between Y_j and the Y_i's. This scheme is suitable when "the latent variable correlation matrix is singular because the weights are based only on the bivariate correlations among component scores" (Chin & Newsted, 1999, p. 317), and it is not suitable when latent variable correlations are close to zero. The inner weights are equal to the correlation between Y_j and Y_i when the factorial scheme is considered. According to Lohmöller (1989), factor scheme "maximizes the variance of the principal component of the latent variables when the number of latent variables goes to infinity" (Chin & Newsted, 1999, p. 309). The inner weights in a path weighting scheme are (a) equal to the regression coefficients of Y_i in the multiple regression of Y_j on all the Y_i related to the predecessor of Y_j, or (b) are equal to the correlation between the successor of Y_i and Y_j. "This scheme attempts to produce a component that ideally can both be predicted and at the same time be a good predictor for subsequent dependent variables" (Chin & Newsted, 1999, p. 318). Taking into account the directionality of the model,

[2] Also called environmental variables for some researchers (Schneeweiss, 1993)

this scheme can be used when the underlying theory is important to the model[3]. Recent contributions have shown that there are no significant differences between the estimates obtained with PLS Path Modelling and the three weighting schemes (Tenenhaus *et al.*, 2005).

Once the auxiliary latent variables are estimated, the weights w_{jh} are recomputed. Recall that, in the iterative process, these weights are used to estimate all latent variable scores as a linear combination of their own indicators. The procedure considers two ways to recompute the outer weights depending on the reflective or formative nature of the outer models: **Mode A and Mode B**. Usually, Mode A is considered for recomputing the outer weights when outer models are reflective, and Mode B is considered for recomputing the outer weights when outer models are formative. In addition, "the algorithm is called PLS Mode C if each of Modes A and B is chosen at least once in the model" (Wold, 1982, p. 10). Only Mode B is scale invariant, that is, linear transformations of the indicators do not affect the latent variable scores. The attributes of the modes A and B are combined in Mode C. Considering two constructs with their respective blocks of variables, Mode C yields a type of the MIMIC model (Dijkstra, 2010).

Dijkstra (2010, p. 32-33) pointed out that: mode A and principal components share a lack of scale-invariance, they are both sensitive to linear scale transformation. According to McDonald (1996), Mode A corresponds to maximization of the sum of absolute values of the covariances of the proxies, where the sum excludes the terms corresponding to latent variables which are not directly related. For Mode A: $\hat{w}_i \propto \sum_{j \in C_i} sign_{ij} S_{ij} \hat{w}_j$ and $\hat{w}_i{}' S_{ii} \hat{w}_i = 1$. Mode B is scale-invariant, in the sense that linear scale transformations of the indicators leave $\hat{\eta}_i$ and $\bar{\eta}_i$ undisturbed. Mode B is a genuine generalization of canonical variables: it is equivalent to the maximization of the sum of absolute values of the correlations between the proxies. For Mode B: $\hat{w}_i \propto S_{ii}^{-1} \sum_{j \in C_i} sign_{ij} S_{ij} \hat{w}_j$ and $\hat{w}_i{}' S_{ii} \hat{w}_i = 1$.

In the algorithm, for Mode A, the w_{jh} is the regression coefficient of Z_j in the simple regression of x_{jh} on the inner estimation of Z_j:

$$w_{jh} = cov(x_{jh}, Z_j) \tag{5.11}$$

For Mode B, the vector w_j of weights w_{jh} is the vector of the regression coefficient in the multiple regression of Z_j on the manifest variables $(x_{jh} - \widetilde{x}_{jh})$ related to the same latent variable Z_j:

$$w_j = (X_j' X_j)^{-1} X_j' Z_j \tag{5.12}$$

[3]See three different inside approximation situations in the path weighting scheme in Chin & Newsted (1999, p. 318).

The first stage is iterated until convergence.

The second stage. Once the algorithm converges, the latent variable scores estimated in stage 1 are used to estimate the inner and outer relationships by ordinary least squares regression without location parameters. If reflective blocks of variables are modeled, simple regression is used to estimate loadings (Equation 5.7). If formative blocks of variables are modeled, weights are estimated by ordinary multiple regression (Equation 5.8).

The third stage. The third stage focus on estimation of the location parameters, and the values of π_{jh0} and β_{j0} (Equation 5.6).

5.3.4 Lohmöller's Implementation

There are two implementations for iteration step of the PLS Path Modelling algorithm, the Wold's and the Lohmöller's procedures. The main difference between them is how they use the information. While the Wold's procedure uses the latest available information at each iteration –of this and previous iteration– the Lohmöller's algorithm only uses the information of the previous iteration. Lohmöller's procedure is commonly implemented in popular software as it is based on matrix algebra (Lohmöller, 1989; Tenenhaus *et al.*, 2005; Hanafi, 2007).

Let J denotes a set of latent variables, each related to a block of manifest variables X_j. Let $C = [c_{jl}]$ be a binary matrix indicating which latent variables are linked by specifying that:

$$c_{jl} = \begin{cases} 1, & \text{if the latent variable j is related to l or viceversa;} \\ 0, & \text{otherwise.} \end{cases} \tag{5.13}$$

Considering the centroid weighting scheme, denoting $R = [r_{jl}]$ the correlation matrix between latent variables, and $\Theta = [\theta_{jl}]$ the matrix with the signs of the correlations between latent variables, the Lohmöller's algorithm can be written as follows:

Step 1: Choose J arbitrary initial vectors $w_j^{(0)}$, $j = 0, 1, 2, ..., J$
To repeat until convergence, $s = 0, 1, 2, ...$
Step 2: External estimation
Step 2.1: To normalize $w_j^{(s)}$ so that $V(Y_j^{(s)}) = 1$
Step 2.2: $Y_j^{(s)} = X_j w_j^{(s)}$
Step 3: Internal estimation
Step 3.1: $r_{jl}^{(s)} = r(Y_j^{(s)}, Y_l^{(s)})$
Step 3.2: $\theta_{jl}^{(s)} = sign(r_{jl}^{(s)})$

Step 3.3: $Z_j^{(s)} = \sum c_{jl}\theta_{jl}^{(s)} Y_l^{(s)}$
Step 4: Updating w_j
$w_j^{(s+1)} = X_j' Z_j^{(s)}$ (Mode A)
$w_j^{(s+1)} = (X_j' X_j)^{-1} X_j' Z_j^{(s)}$ (Mode B)

It is worth noting that, the environmental variable $Z_j^{(s)}$ –in step 3.3– is computed as a linear combination of latent variables computed in the iteration s. The way of calculating the auxiliary variable is the key difference in the Lohmöller's and Wold's implementations.

5.3.5 Wold's Implementation

Here, it is presented the Wold's algorithm as further extended by Hanafi (2007)[4]. That is, for (a) any number of blocks and (b) any conceptual design. Some modifications are made in order to clarify the presentation. If J denotes a set of latent variables, let $C = [c_{jl}]$ be a binary matrix that indicate which latent variables are linked by specifying that:

$$c_{jl} = \begin{cases} 1, & \text{if the latent variable j is related to l or viceversa;} \\ 0, & \text{otherwise.} \end{cases} \tag{5.14}$$

Considering the centroid weighting scheme, denoting $R = [r(Y_j, Y_l)]$ the correlation matrix between latent variables, and $\Theta = [\theta_{jl}]$ the matrix with the signs of the correlations between latent variables, the Wold's algorithm can be written as follows:

Step 1: Choose J arbitrary initial vectors $w_j^{(0)}$, $j = 0, 1, 2, ..., J$
To repeat until convergence, $s = 0, 1, 2, ...$
Step 2: External estimation
Step 2.1: To normalize $w_j^{(s)}$ so that $V(Y_j^{(s)}) = 1$
Step 2.2: $Y_j^{(s)} = X_j w_j^{(s)}$
Step 3: Internal estimation
Step 3.1: if $j < l$, $\quad r_{jl}^{(s)} = r(Y_j^{(s)}, Y_l^{(s)})$
Step 3.1: if $j > l$, $\quad r_{jl}^{(s)} = r(Y_j^{(s)}, Y_l^{(s+1)})$
Step 3.2: $\theta_{jl}^{(s)} = sign(r_{jl}^{(s)})$
Step 3.3: $Z_j^{(s)} = \sum_{j<l} c_{jl}\theta_{jl}^{(s)} Y_l^{(s)} + \sum_{j>l} c_{jl}\theta_{jl}^{(s)} Y_l^{(s+1)}$
Step 4: Updating w_j
$w_j^{(s+1)} = X_j' Z_j^{(s)}$ (Mode A)
$w_j^{(s+1)} = (X_j' X_j)^{-1} X_j' Z_j^{(s)}$ (Mode B)

[4] Recall that Wold (1985)'s implementation was presented for only 6 constructs.

5.4 Starting Values of Weight Vectors

The PLS Path Modelling algorithm requires an arbitrary vector of weights to initialize the iterative procedure. Some researchers have been concerned about whether different initial vectors generate different estimates for a model. Wold (1982, p. 13) and Wold (1985, p. 585) recommend to use the first indicator of each block of variables for initializing of outer estimates of latent variables. For each outer model, this is equivalent to consider a weights' vector where the first weight is equal to 1, and the others are equal to zero. Lohmöller (1989) suggests to take all the weights equal to 1, but the last one equals -1. The researcher argues that this should prevent against possible negative estimates of the outer weights. Tenenhaus *et al.* (2005) argue that Lohmöller's choice entails sign problems in cross validation, so it should not be the standard choice to initialize the algorithm. On the contrary, Tenenhaus *et al.* (2005) recommend to choose initial weights as the sign of the correlations between manifest and latent variable, $w_{jh} = sign(cor(x_{jh}, \xi_h))$. Wold (1980, p. 58) also suggests to take an initial vector of ones. This is a common practice among researchers. For instance, Henseler (2009) thus initializes the algorithm, computing the outer estimates of the latent variables as the sum of the indicators of each block, $\xi_j = \sum_h x_{jh}$. According to Temme & Kreis (2005) the initial values of weights have an impact in the sign of the final estimates of loadings and path coefficients[5]. Henseler (2009) also agree that the selection of the initial weight vector is an important aspect in the calculation of the latent variable scores, especially for the algorithm convergence. The author recommends that it "should be choose the average of the indicators as the standard initialization of the latent variable score" (p. 11). These views are not shared by Dijkstra (2010, p. 13), who reminds us that "PLS algorithms will converge for every choice of starting values to unique fixed points" (see section 5.7 for details).

5.5 Weighting Schemes

As seen above, centroid weighting scheme was introduced by Wold in his original design of the PLS Path Modelling algorithm, whereas factorial and path weighting schemes were later proposed by Lohmöller (1989). According to Tenenhaus *et al.* (2005), factorial and path weighting schemes do not significantly influence the estimates. But, these researchers argue that these schemes are useful to relate

[5]These researchers also pointed out that the sign of weights/factor loadings and path coefficients can vary considerably across the different computational implementations of the algorithm, LVPLS, PLS-GUI, PLS-Graph, SmartPLS.

PLS Path Modelling to multi tables analysis methods. Ringle *et al.* (2009) comparing the performance of the PLS Path Modelling algorithm with a covariance-based approach, checked the differences between the estimates given by the algorithm if different weighting schemes are used. They conclude that "on average, the alternative weighting schemes provide the same parameter estimates for the model under investigation" (p. 18). This result was previously found by Noonan & Wold (1982). Nevertheless, results should be interpreted with caution because some studies have shown that the weighting scheme is an important factor for the algorithm convergence. Henseler (2009) studied this latter topic assessing the PLS performance under the different weighting schemes. Problems on the algorithm convergence are reported when estimating a model with 2 reflective exogenous and 1 reflective endogenous latent variables, with 6 specific data set. Henseler found that the PLS Mode A algorithm may not converge if the factorial or path weighting schemes are used. On the contrary, it has not been reported convergence problems when using the centroid weighting scheme, which is recommended for all researchers.

5.6 Consistency and Consistency at Large

Consistency refers to sample size and consistency at large refers to the number of observable variables per construct. According to Anderson & Gerbing (1988) "PLS estimates will be asymptotically correct only under the joint conditions of consistency (sample size become large) and consistency at large (the number of indicators per latent variable becomes large)" (p. 412). In addition, and as stated by Schneeweiss (1993, p. 301), "consistency at large is a property of the model and not a property of an estimation method"[6]. This claim is very important to understand the results of a number of simulations studies about PLS Path Modelling robustness and performance. Studies from Hui & Wold (1982), Chin & Newsted (1999), Cassel *et al.* (1999), Chin *et al.* (2003), and Westlund *et al.* (2008) have confirmed that increasing the number of manifest variables per latent variable increases the accuracy and precision of the PLS Mode A estimates[7]. However, practical models or "real-world models" have to be able to involve an appropriate number of manifest variables per latent variable. Some constructs are not reflected in or not formed by a large number of variables, by the way. It is worth noting that Anderson & Gerbing (1988, p. 416) pointed out that when two indicators per latent variables are available, larger samples "may be needed to obtain a converged and proper solutions" (p. 146).

[6] The same researcher pointed out that this concept should be distinguished "from the common concept of consistency in estimation theory" (Schneeweiss, 1993, p. 301).

[7] Most of these simulations have been carried out with reflective measurement models.

5.7 Optimization Problem and Convergence

Even though PLS Path Modelling belongs to the family of fixed-point methods[8] (Wold, 1966; Lyttkens, 1973; Dijkstra, 1983, 2010), it is known that the algorithm does not optimize any global function. Herman Wold himself clarified this issue. Wold (1980, p. 66) notes that:

> PLS procedure minimizes in its first stage each residual variance in the weight relations that for the various blocks of the model are given by [Mode A or Mode B]... Mode A uses simple OLS regressions to minimize with respect to the parameters one by one, whereas Mode B uses multiple OLS regression to minimize with regard to the parameters jointly... PLS procedure remains partial in the sense that no total residual variance or other overall criterion is set up for optimization.

In Bookstein (1982, p. 56)'s words "instead of a global optimization, the LVs are jointly characterized using a complicated non-linear operator for which the vector of all estimated item weights (outer relations) serves as fixed-point."

On the other hand, it is a common belief among researchers that the PLS algorithm always converge in practice. However, some convergence problems have been recently reported (Henseler, 2009). It is somewhat surprising that the literature on PLSPM convergence does not seem to be widespread, especially considering the fact that besides the Wold's works, other researchers as Bookstein (1982), Areskoug (1982), Dijkstra (1983), Schneeweiss (1993), Mathes (1993), Hanafi & Qannari (2005), Krämer (2006), Hanafi (2007), and Dijkstra (2010) have studied this issue. Early, Wold (1982, p. 24) stressed that "for one- and two-block soft models the iterative stage of the PLS estimation is almost always convergent. For multi-block soft models the convergence has not been proved." Wold (1980, p. 66) asserted that "for PLS models with three or more blocks, convergence of the estimation procedure has never been a problem in applications to real-world models and data." Wold (1982, p. 24) goes on saying that "it seems that the PLS algorithm will fail to converge only in the exceptional case when the largest eigenvalues are equal or nearly equal."

Krämer (2006) have studied the mathematical properties of PLS Path Modelling. Extending the canonical correlation analysis to more than two blocks of variables, the researcher posed the optimization problem as follow. Let $Y_j = X_j w_j$ be a set of latent variables such that Y_j and Y_l are maximally correlated if the block of variables are linked. For J blocks of variables with p_j indicators each, the

[8] That is because PLS Path Modelling algorithm found fixed-points ("the weight vector") by means an iterative sequence of regressions starting with an arbitrary choice of weights (Dijkstra, 1983, p. 78).

optimization problem is:

$$\begin{array}{ll} \arg\max_w & \sum_{j,l:c_{jl}\neq 0} g(cov(X_j w_j, X_l w_l)) \\ \\ \text{subject to} & \frac{1}{n} \parallel X_j w_j \parallel = 1 \end{array} \tag{5.15}$$

where $g(\cdot)$ is the centroid, factorial or Horst schemes. Raising the Lagrangian function and differentiating with respect to w_j and the Lagrangian multiplier λ_i, Krämer obtained the Lagrangian equations of the optimization problem 5.15:

$$\begin{array}{rcl} S_g(w)w & = & \Lambda S_D w \\ w_j' S_{jj} w_j & = & 1 \end{array} \tag{5.16}$$

where S is the covariance matrix between block of manifest variables, S_D is the diagonal matrix of S, and Λ is a diagonal matrix of the form $\Lambda = diag[\lambda_1 I_{p_1}, \ldots, \lambda_k I_{p_J}] \in \mathbf{R}^{p\times p}$. Krämer poited out that "any solution of the equations 5.16 are called stationary points of the optimization problem 5.15. In general, there is more than one stationary point. The stationary point w that is solution of 5.15 is the one such that the corresponding multivariate eigenvalues fulfill, $\sum_{j=1}^{J} \lambda_j = \max$" (p. 7).

If the variance among latent variables instead of correlations is maximized, the optimization problem is:

$$\begin{array}{ll} \arg\max_w & \sum_{j,l:c_{jl}\neq 0} g(cov(X_j w_j, X_l w_l)) \\ \\ \text{subject to} & \frac{1}{n} \parallel w_j \parallel = 1 \end{array} \tag{5.17}$$

and the Lagrangian equations of this optimization problem are:

$$\begin{array}{rcl} S_g(w)w & = & \Lambda w \\ \frac{1}{n} w_j' w_j & = & 1 \end{array} \tag{5.18}$$

Krämer (2006) proved that if a structural equation model have all block of variables modeled in Mode B, and if the PLS algorithm converges, the PLS estimates are stationary points of the optimization problem 5.15. This is valid for the Wold's and Lohmöller's procedure, and for the three analyzed weighting schemes. If in a structural equation model, at least one block of variables is modeled as Mode A, Krämer's findings are quite the opposite. The researcher found that the PLS estimates are not stationary points of a optimization problem, or what is the same, that "equations that determine the solution of Mode A cannot be the Lagrangian equations of any twice differentiable optimization problem." Regarding to this latter result for Mode A, Wold (1980) early mentioned that this is precisely one of the advantages of PLS Path Modelling. The researcher pointed out that "the weight relations avoid the nonlinear side relations that would arise if the aggregate relations were treated as side conditions and had to be taken into account through Lagrange multipliers" (p. 66). In addition, Krämer (2006) showed that PLS Path Modelling algorithm with Mode B not necessarily converge to the solution of 5.15, that is an optimum solution, although Hanafi (2007) has proved that the PLS Path Modelling algorithm with Mode B converges monotonically. Krämer (2006) concluded that "PLS path algorithms in Mode A produce algebraic equations that are not linked to any sufficiently smooth optimization problem. This marks a severe setback in the search of a justification of Mode A in terms of optimality criteria" (p. 17).

For PLS Path Modelling with Mode B, Mathes (1993) optimized the correlation structure of PLS-mode-B latent variables for the following global criteria:

Theorem 1 The PLS estimation of Mode B with centroid weighting scheme is a critical point of the function "sum of absolute correlations of the adjacent latent variables in the structural system."

Theorem 2 The PLS estimate of Mode B with correlation weighting scheme is a critical point of the function "sum of squared correlations of adjacent latent variables in the structural system."

Table 5.1 shows a summary of the results of different researchers related to the PLS Path Modelling convergence. Finally, as for PLS Path Modelling with Mode A convergence is not assured, a new Mode A has been proposed by Tenenhaus (2009) introducing a normalization on the weights. The optimization problem is:

$$\begin{array}{rl} \arg\max_w & \sum_{j<l:c_{jl}\neq 0} g(cov(X_j w_j, X_l w_l)) \\ \text{subject to} & \parallel w_j \parallel = 1 \end{array} \tag{5.19}$$

However, further research is needed to ensure the monotone convergence of the new procedure.

5.8 PLS Path Modelling and Related Approaches

As seen above, PLS Path Modelling with Mode B is a generalization of Canonical Correlation Analysis (Horst, 1961) to more than two blocks of variables. Tenenhaus *et al.* (2005) present a discussion about this and how the PLS Path Modelling finds other methods, such as multiple factor analysis (Escofier & Pagés, 1994), among others. Under certain conditions, the procedure also leads to the following results.

Table 5.1: Summary of results about the convergence of the PLS algorithm

Author	**Constraint**	**Results**
Chu & Watterson (1993)	Path weighting scheme (or Horst scheme)	Proving the convergence of Lohmöller's procedure with Mode B
Krämer (2006)	All block of variables modeled as Mode B Wold's and Lohmöller's procedure Three weighting schemes	PLS estimates are stationary points of a optimization problem
Krämer (2006)	If at least one block of variables modeled as Mode A	PLS estimates are not stationary points of a optimization problem
Hanafi (2007)	All block of variables modeled as Mode B	Proving the monotone convergence of Wold's procedure with Mode B

The first principal component. One-block PLS Path Modelling with Mode A is numerically equivalent to the first principal component (Wold, 1980, 1985).

The first canonical correlation. Two-block PLS Path Modelling with Mode B gives the first canonical coefficient as the estimated correlation between the two unobserved variables (Wold, 1980; Areskoug, 1982; Wold, 1985).

The first component of the inter-battery factor analysis. Two-block PLS Path Modelling with Mode A finds the first component of the inter-battery factor analysis between two sets of manifest variables (Tenenhaus *et al.*, 2005).

The first component of the redundancy analysis. PLS Path Modelling with Mode B for an exogenous construct and Mode A for an endogenous latent variable gives the first component of the redundancy analysis of the path model.

5.9 Validation Techniques

5.9.1 Assessment of Reflective Outer Models

Reliability and validity of reflective outer models must be assessed. According to Tenenhaus *et al.* (2005), both criteria must consider three aspects, (1) convergent validity of the measurement variables, (2) discriminant validity, and (3) reliability of individual items.

Convergent validity. The internal consistency, unidimensionality or convergent validity of reflective blocks of variables is related to the coherence between constructs and their measurement variables. All indicators of a reflective block of variables must reflect the same construct. Three indexes are often tested to assess unidimensionality. (1) Principal component analysis (PCA) of the block of variables; the rule of thumb is that the first eigenvalue of the correlation matrix of the reflective outer model should be greater than 1, and the second one smaller than 1. (2) Cronbach's alpha index should be greater than 0.7. The alpha index is defined as:

$$\alpha = \frac{\sum_{i \neq i'} cor(x_i, x_{i'})}{p + \sum_{i \neq i'} cor(x_i, x_{i'})} \times \frac{p}{p-1} \tag{5.20}$$

where x_i and $x_{i'}$ are indicators of a p-standardized block of variables (X). (3) Dillon-Goldstein's ρ (also called composite reliability) should be greater than 0.7. The ρ index is defined as:

$$\rho = \frac{(\sum_{i=1}^{p} cor(x_i, t))^2}{(\sum_{i=1}^{p} cor(x_i, t))^2 + \sum_{i=1}^{p}(1 - cor^2(x_i, t))} \tag{5.21}$$

where t is the first principal component of X. According to Chin (1998a, p. 320) "alpha tends to be a lower bound estimate of reliability whereas composite reliability is a closer approximation under the assumption that the parameter estimates are accurate." Low internal consistency suggests a poorly defined model or multidimensional constructs.

Reliability of individual items. Loadings indicate how well the indicators reflect the latent variable with which they are related. These parameters represent the correlation between indicators and component scores. So, reliability is evaluated examining loadings. A rule of thumb generally accepted is 0.7 or more. "This implies that there are more shared variance between construct and variable than error variance" (Hulland, 1999, p. 198). A low value of a loading factor suggests that the indicator has little relation to the associated construct.

Discriminant validity. Discriminant validity represents "the extent to which measures of a given construct differ from measures of other constructs in the same model" (Hulland, 1999, p. 199). For this, a construct should share more variance with its indicators than with other constructs in a given model. This is measured computing communality index for each reflective construct j:

$$Communality_j = \frac{1}{p}\sum_{i=1}^{p} cor^2(x_i, Y_j) \tag{5.22}$$

where Y_j is the construct estimated by p-manifest variables x_i. The average communality index is measured as:

$$\overline{Communality} = \frac{1}{\sum_{j=1}^{J} p_j}\sum_{i=1}^{J} p_j \times communality_j \tag{5.23}$$

Average Variance Extracted (AVE) suggested by Fornell & Larcker (1981) is also used to measure the average variance shared between a construct and its measures. That is, the percentage of variance that is captured by a construct in relation to the variance due to random measurement error. This value should be greater than the variance shared between construct and other constructs in the model. A rule of thumb is to accept an AVE greater than 0.5. For a construct j, the index is defined as:

$$AVE_j = \frac{\sum_{i=1}^{p} cor^2(x_i, Y_j)}{\sum_{i=1}^{p} var(x_i)} \tag{5.24}$$

In addition, cross loadings may be calculated when two or more reflective constructs are in the model. Cross loadings may be obtained by calculating the correlations between latent variable component scores and indicators associated with other reflective constructs. If a parameter has higher correlation with other latent variable instead of the associated, it its position in the model should be reconsidered.

5.9.2 Assessment of Formative Outer Models

In regard to formative blocks of variables, weights allow us to determine the extent to which each indicator contributes to the formation of a construct. Recall that in this case, unidimensionality is not a necessary condition. Additionally, measures to asses validity are not necessary valid for formative outer models (Bollen & Lennox, 1991; Cohen *et al.*, 1990; MacCallun & Browne, 1993). See Chapter 4 for a discussion about formative relationships and details about Confirmatory Tetrad Analysis. Reliability of formative outer models may also be assessed by examining the correlations between the constructs and their corresponding manifest variables. As seen above, weight estimates and correlations should have the same sign for the sake of interpretation.

5.9.3 Assessment of Structural Models

The inner model is assessed by examining path coefficients among constructs, the R-square coefficient, and redundancy indexes for each endogenous construct j. The value of path coefficients provide evidence about the strength of the association among latent variables. R-square or multiple determination coefficient of each endogenous constructs gives the overall fit of a model or the percentage of variance explained by the exogenous constructs. This coefficient is sensible to model set up and the election of the dependent variable. The effectiveness of R-square depends on the quotient between the number of variables and the sample size. Assessing changes in R-squares allows us to determine the effect size f^2 of a exogenous construct on an endogenous constructs (Chin, 1998a):

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \tag{5.25}$$

where $R^2_{included}$ and $R^2_{excluded}$ are the R-squares obtained for the endogenous construct when exogenous construct is included and excluded of the estimated structural model. Size effects of 0.02, 0.15, and 0.35 indicate a small, medium, or large effect at the structural level.

Redundancy index for an endogenous constructs is defined as:

$$Redundancy_j = Communality_j \times R^2_j \tag{5.26}$$

Here, the index measures the portion of variance explained by the exogenous constructs but also considering the correlations between these constructs and their corresponding manifest variables. The average redundancy index is measured as:

$$\overline{Redundancy} = \frac{1}{J} \sum_{j=1}^{J} Redundancy_j \tag{5.27}$$

An overall measure of goodness-of-fit of the models is given by the GoF index (Tenenhaus *et al.*, 2004). This is defined as the geometric mean of the average communality and the average R^2:

$$GoF = \sqrt{\overline{communality} \times \overline{R^2}} \tag{5.28}$$

5.9.4 Resampling Techniques

Resampling techniques are used to estimate the predictive power of the models and significance of relationships. While cross-validation is used for prediction, bootstrap or jackknife allows the assessment of the stability and significance of model parameters.

Rooted in data mining procedures, cross-validation technique provides a way to analyze how generalizable the results are to a new data set. Using different partitions of the sample into a training data set and a test set, the former is used to perform a statistical analysis, and the latter for validating it. Average results are taken from replications to validate a model. Wold (1982, p. 30) proposed to adapt the Stone-Geisser's test –"a kind of cross-validated R-Square" (Tenenhaus *et al.*, 2005)– for proving the "predictive significance of principal components of consecutive orders." This approach follows a blindfolding procedure. That is, part of the data and indicators are "blinded", the model is estimated, and the obtained parameters are used to predict the blinded "folds." The procedure is repeated until all folds have been blinded once. For each re-estimated prediction, the sum of squares of prediction errors (SS) and the sum of square errors using the mean prediction (SSM) are computed. Wold (1982) sum up by a test criterion Q^2 as follows:

$$Q^2 = 1 - \frac{\sum SS}{\sum SSM} \tag{5.29}$$

If $Q^2 > 0$ the model has predictive relevance. Furthermore, discriminant validity and redundancy indexes may also be computed by cross-validation (see Tenenhaus *et al.* (2005)).

On the other hand, Bootstrap is a method proposed by Bradley Efron in 1979 to assess measures of accuracy to statistical estimator (Efron & Tibshirani, 1993). Given an independent data set denoted by the vector $x = (x_1, x_2, ..., x_n)$, to estimate an statistic $s(x)$, n random samples (bootstrap samples) are taken with replacement from the original sample, $x^* = (x_1^*, x_2^*, ..., x_n^*)$. A sufficiently large number (for instance 1000) of independent bootstrap samples should be taken, each of size n. The statistic for each bootstrap sample is evaluated, $s(x^{*b})$. The bootstrap estimate of standard error is the standard deviation of the bootstrap replications. So,

$$\widehat{se}_{boot} = \{\sum_{b=1}^{B} [s(x^{*b}) - s(\cdot)]^2/(B-1)\}^{1/2} \tag{5.30}$$

where $s(\cdot) = \sum_{b=1}^{B} s(x^{*b})/B$. Standard errors are usually estimated by bootstrapping. But, they may also be computed by Jackknife (Lebart, 1985). Unlike bootstrap, jackknife leaves out one observation in each n random sample. Let the ith jackknife sample a data set with the ith observation removed. So, if $\widehat{\theta}_{(i)} = s(x_{(i)})$ is the ith jackknife replication of $\widehat{\theta}$, the jackknife estimate of standard error is defined by,

$$\widehat{se}_{jack} = [\frac{(n-1)}{n} \sum (\widehat{\theta}_{(i)} - \widehat{\theta}_{(\cdot)})^2]^{1/2} \tag{5.31}$$

where $\widehat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \widehat{\theta}_{(i)}/n$. The jackknife sample tends to be more similar to the original data than the bootstrap sample. Jackknife is easier to compute than bootstrap if n is less than 100 or 200 replicates. However, the former uses limited information about the statistic, so it is less efficient than bootstrap. Jackknife is a procedure that may be seen as an approximation to the bootstrap procedure.

Chapter 6

Monte Carlo Simulations and Computational Experiments for PLS Path Modelling with Mode C

Abstract. Monte Carlo simulations and computational experiments were carried out to study the performance of Partial Least Squares Path Modelling algorithm with Mode C and few indicators per construct. The empirical results are in line with the theoretical PLS framework. Outer relationships are overestimated and inner relationships underestimated.

6.1 Introduction

Depending on its nature, manifest variables have been referred to either as effects (reflective) indicators or causes (formative) indicators. Wold's basic design of Partial Least Squares (PLS) Path Modelling usually refers to Mode A and Mode B, or simple and multiple regressions (Wold, 1985), depending on the features of the outer models. Reflective relationships seek to represent variances and covariances between the manifest variables that are generated or caused by a latent variable. So, observed variables are treated as an effect of unobserved variables (Cohen *et al.*, 1990; Bollen & Lennox, 1991). In a reflective measurement model, the manifest variables are measured with error. Alternatively, formative relationships are used to minimize residuals in the structural relationship (Fornell & Bookstein, 1982). Here, manifest variables are treated as forming the unobserved variables, they are presumed to be error–free, and the unobserved variable is estimated as a linear combination of the manifest variables plus a disturbance term. As in this case all variables forming the construct should be considered, the disturbance term represents all those non-modeled causes (Diamantopoulos, 2006). Although formative measurement models were first discussed by Curtis & Jackson (1962) and Blalock (1964), and a number of variables can be modeled in a better way through formative relationships (Hulland, 1999), measurement variables have been traditionally modeled in a reflective mode. Fornell & Bookstein (1982) and Chin (1998b) pointed out that modelling formative modes using a covariance-based approach may lead to identification problems and Heywood cases[1]. So, researchers may tend to define outer models as reflective. However, a number of researchers have pointed out that PLS Path Modelling overcomes the identification problems that arise when implementing a covariance-based approach (Wold, 1980, 1985; Tenenhaus *et al.*, 2005). That is because a PLS Path Modelling algorithm consists of a series of ordinary least squares (OLS) analysis. From a component-based approach, and "because the off-diagonal elements are not among the unknown parameters of the model and because the unobservables are explicitly estimated, there are no identification problems for recursive PLS models" (Fornell & Bookstein, 1982, p. 443).

In Wold's PLS approach (1985), a construct is completely determined by a linear combination of its indicators. The procedure usually uses a Mode A or Mode B to model a structural equation model (SEM). Mode A or simple regression if the SEM includes reflective outer models. Mode B or multiple regression if formative outer models are included. However, "the algorithm is called PLS Mode C if each of Modes A and B is chosen at least once in the model" (Wold, 1982, p. 10). To the best of our knowledge, there are only a small number of published articles that examine the performance of PLS Path Modelling algorithm in the presence of formative outer models, and they are not conclusive. Findings by Cassel *et al.* (1999) and Ringle *et al.* (2009) are quite different. For instance, Cassel et al. found that measurement relationships in formative outer models are overestimated, while Ringle et al. found that these relationships are underestimated. Thus, this paper aims to provide evidence regarding how well PLS Path Modelling performs if formative exogenous outer models are modeled using PLS Mode B and reflective endogenous latent variables are modeled using PLS Mode A. That is, PLS Path Modelling with Mode C. This is also the set-up under which patent value models are proposed.

On the other hand, the issue of few indicators per construct is a topic less often addressed by the scientific community. Using simulations, some authors have studied the performance of the partial least square procedure, including in

[1]A Heywood case in common factor analysis occurs when the minimum of the discrepancy function is obtained with one or more negative values as estimates for the variables of the unique variables. Heywood cases occur when too many factors are extracted, or the sample size is too small.

their analysis the case of two indicators per latent variable, but with reflective outer models. The accepted general recommendation is to have at least three or four manifest variables per latent variable, and the belief among researchers is that with few indicators, the estimates are neither acceptable nor significant. Recall the consistency at large property of the PLS Path Modelling algorithm, "the larger the number of indicators in a block, the more the 'essence' of the LV is confirmed by the data" (Chin & Newsted, 1999, p. 329). However, in reality, there are proposed models that have used only one or two indicators in some latent variables. See for instance the ECSI (European Customer Satisfaction Index) model where the perceived value is related to two manifest variables and the customer complaints are related to one indicator (Tenenhaus *et al.*, 2005). In addition, in the case of a formative measurement model, the manifest variables should be a census of all variables that generate the latent variable (Bollen & Lennox, 1991). So, what happens if the census consists of only two manifest variables? Would the algorithm then be able to capture the relationship between the variables? In a number of cases regarding technology change and technology watch, for instance, two indicators per construct could be considered. So, we are interested in knowing how reliable and robust the parameter estimates are when few indicators per unobservable variable are considered.

6.2 PLS Path Modelling

The PLS Path Modelling procedure –presented by Gerlach, Kowalski, and Wold in 1979– is a soft modelling technique and a data analytic tool for estimating structural equation models (SEM) and building a sequence of latent variables. PLS Path Modelling first estimates the unobservable variables and then the parameters with an aim toward maximizing the total variance and minimizing residuals of endogenous models regardless of the covariances among manifest variables. The structural model or inner model describes relationships among constructs ξ_i by means of multiple regressions (Equation 6.1). ξ_j and ξ_i are the endogenous and exogenous latent variables, respectively, and β_{ji} are the path coefficients that measure the relationship among constructs. The condition imposed by Herman Wold is predictor specification, $E(\xi_j/\xi_i) = \sum_i \beta_{ji}\xi_i$, that is, there is no linear relationship between predictor and residual. This condition implies that $E(\nu_j/\forall\xi_i) = 0$, and $cov(\nu_j, \xi_i) = 0$.

$$\xi_j = \beta_{j0} + \sum_i \beta_{ji}\xi_i + \nu_j \tag{6.1}$$

Manifest variables revealing or reflecting the effect of a construct are modeled as indicators of it in a reflective measurement model. Each manifest variable x_{jh} is related by simple ordinary least squares regression with the underlying construct ξ_j (Equation 6.2). The loadings λ_h determine the extent to which each indicator reflects a construct; ξ_j is a common factor with mean m, standard deviation one and it is indirectly observable by the manifest variables. The condition imposed by Herman Wold is predictor specification, $E(x_h/\xi) = \lambda_{h0} + \lambda_h \xi$. This condition implies that ϵ_h has zero mean, and it is uncorrelated with ξ_j. Moreover, the basic design of Herman Wold assumes that the covariance matrices of all ϵ_j are diagonal. As in a reflective model, where all the indicators of the block of variables reflect the same construct, there should be high collinearity among these variables. That is, the blocks of variables must be one-dimensional.

$$x_{jh} = \lambda_{jh0} + \lambda_{jh}\xi_j + \epsilon_{jh} \tag{6.2}$$

The latent variable is formed by a set of manifest variables as a linear function of them plus a residual in formative outer models (Equation 6.3). The weights π_h determine the extent to which each indicator contributes to the formation of the constructs. Each block of manifest variables may be multidimensional, and multicollinearity among indicators is not a necessary constraint. The condition imposed by Herman Wold is predictor specification $E(\xi/X_1, ..., X_{pj}) = \sum_h \pi_h x_h$. This condition implies that the residual δ has a zero mean, and it is uncorrelated with the manifest variables x_h. Since each construct is formed by a linear combination of the manifest variables, to facilitate interpretation of the estimates, it is desirable that the sign of each weight π_h is the same sign of the correlation between x_h and ξ (Tenenhaus *et al.*, 2005).

$$\xi_j = \sum_h \pi_{jh} x_{jh} + \delta_j \tag{6.3}$$

The PLS Path Modelling algorithm is structured in three stages (Wold, 1980, 1982, 1985). The first stage computes the case values of the latent variables; the second stage focuses on the inner and outer relationships; and in the third stage, location parameters of the latent variables, λ_{jh0} and β_{j0}, are estimated. Only the first stage is iterative.

6.3 Monte Carlo Simulations in a PLS Path Modelling Context

To the best of our knowledge, there are only a small number of published articles that examine the performance of PLS Path Modelling algorithm in the presence of formative measurement models with the aim of studying (1) how well the method recovers true population parameters, and (b) how the algorithm performs against multicollinearity, skewed data and model misspecification. Cassel *et al.* (1999), Cassel *et al.* (2000), Ringle *et al.* (2007), and Westlund *et al.* (2008) performed simulation studies based on Monte Carlo data generation for a hypothetical model. Recently, a working paper and a book chapter have been published by Ringle *et al.* (2009) and by Vilares *et al.* (2010), respectively. Albeit some of these studies have different objectives, we present here the main results.

Cassel *et al.* (1999) examined the behavior of a simplified version of the Swedish Customer Satisfaction Index (SCSI) structural model in terms of the bias and standard deviations of the estimates. They studied the presence of three difficulties: (1) skew distribution for observed variables, (2) multicollinearity within a block of manifest variables and between latent variables, and (3) misspecification of the structural model. This last condition refers to when the structural model is not completely defined; that is, an important latent variable in terms of its regression coefficients is not included in the model. The authors studied a model composed of three formative exogenous constructs and one reflective endogenous latent variable. The data for the observed variables were generated from a symmetric beta distribution (6,6) and the model coefficients were assumed to be known. The error terms in the measurement models were generated from a continuous uniform distribution, and the error in the structural model from a normal distribution (all errors with mean zero and accounting for 30% of the variance of the variable). They compared the estimates of the PLS algorithm for a basic model, with the following cases: (a) a model with skewed manifest variables generated from beta distributions; (b) a model with multicollinearity within a block of manifest variables and between latent variables; and (c) models where one of the constructs (regressor) is omitted. When Cassel *et al.* (1999) report the results of their estimates, they mention that "the basic model is specified without specification error and correlation between the latent or manifest variables" (p. 442). The researchers found that a PLS algorithm is quite robust in the presence of these conditions for all sample sizes. They also reported that in the absence of specification error and correlation, the determination coefficient for the structural model is below 50 percent. So, in an ideal (and unrealistic) case, we would expect that the model fit the data 50 percent at most (structural model). Additionally, there was a "substantial" increase in the bias in the estimates only when the data analyzed is extremely skewed or when an important latent variable is omitted. The estimates of loadings (underestimated), weights (overestimated) and path coefficients (underestimated) are close to the true values in all scenarios studied (reported bias are between 20% and 50% depending on the estimated parameter). They confirmed that the PLS algorithm does not depend on data distributional assumptions. The method showed a better performance when a greater number of observed variables was considered. The researchers concluded that "PLS estimators of the inner structure coefficients are inconsistent; they are only consistent at large, which means that they are consistent with the increasing size of the blocks of manifest variables. Hence, for finite sample sizes, biased PLS estimation must be expected, and the question arises of in what may and to what extent this bias is affected by distributional properties" (p. 445). Nevertheless, although Cassel *et al.* (1999) said that "biases can be expected to be reduced only when the number of explanatory variables in the corresponding sub-model is increased" (p. 443), they do not analyze the effects of changing the number of manifest variables in each block. We address this aspect in the Monte Carlo simulation performed below.

Table 6.1 shows a comparison between the studies by Cassel *et al.* (1999) and Chin & Newsted (1999). Even though Chin & Newsted (1999) did not work with formative relationships, their results are summarized in Table 6.1 because it provides a reference for how other researchers with similar objectives have performed their simulations[2].

Another study related to formative measurement models has been conducted by Ringle *et al.* (2007, 2009)[3]. They analyzed the robustness properties of formative indicators using an experimental setup that allowed the estimation of the model through component- and covariance-based approaches. The set-up consisted of a model with three formative exogenous constructs, each with five, three and five indicators, and two reflective endogenous latent variables with three indicators each. Unlike Cassel *et al.* (1999), Ringle *et al.* (2009) generated 1000 sets of multivariate normal data from a correlation matrix (300 cases). Among others, findings reported by Ringle *et al.* (2009) are: (1) centroid, factor and path schemes provide the same results from PLS Path Modelling; (2) PLS Path Modelling has a tendency to underestimate weights in the formative measurement models (for both normal and non-normal scenarios); (3) PLS has a tendency to overestimate loadings in reflective outer models; and (4) PLS underestimates inner relationships

[2]Chin & Newsted (1999) investigated the behavior of the PLS algorithm using reflective measurement models, and compared the results of a partial least square simulation to the simple path-analytic regression using a summation of the indicators. This research also examined how well the PLS algorithm performs in recovering the true population parameters. Their research found that partial least square always performed better than the simple summed regression approach.

[3]The first work was presented as a poster in the PLS'07 conference (Ås, Norway). The working paper of 2009 seems to be a revised and extended version of this study.

(path coefficients).

Study results by Cassel *et al.* (1999) and Ringle *et al.* (2009) are quite different. Cassel et al. found that measurement relationships in formative blocks of variables are overestimated, while Ringle et al. found that these relationships are underestimated. The opposite happens in the case of reflective outer models. Cassel et al. found that reflective relationships are underestimated, contradicting the findings of other researchers. Cassel et al. and Ringle et al. agree that estimates of structural relationships are underestimated. Although the models tested by Cassel et al. and Ringle et al. are different, they are not so very different. Cassel et al. considered a model with three formative exogenous constructs and one reflective endogenous latent variable, while Ringle et al. considered a model with three formative exogenous constructs and two reflective endogenous latent variables. The main difference between the study of Ringle *et al.* (2009) and others lies in how the data are generated. So, in this research we are interested in confirming part of these results.

Along similar lines of research, Vilares *et al.* (2010) also performed a simulation study to assess the effects of two assumptions, the symmetry of the distributions and the reflective modelling of the indicators. They used the ECSI (European Customer Satisfaction Index) model and compared, among others, (a) a model "where all blocks are reflective and the measurement variables show a symmetric distribution" (base model); and (b) the base model but with the exogenous latent variable (image) in a formative outer model. One of the main results reported by the authors is that the PLS approach was very robust under formative outer models and skewed data, but unfortunately the authors did not report additional results for formative estimates. It is worth noting, that Vilares *et al.* (2010) always found an adequate PLS Path Modelling performance, even for loadings in the perceived value measurement model. In this outer model, the block of variables included only two manifest variables related to the construct in a reflective mode.

Westlund *et al.* (2001) also studied the robustness of the PLS Path Modelling algorithm against the skewness of the observed variables, misspecification of the model, and measurement errors. They studied the basic European Performance Satisfaction Index (EPSI) model, and the data generation follows the procedure implemented by Cassel *et al.* (1999). They found that "the introduction of measurement errors in data does not add much to the estimated bias, aside from the consequences already noticed due to misspecification problems, or problems due to skew response distributions" (p. 879). The study of Westlund et al. was extended in 2008 also through Monte Carlo simulations. The authors analyzed the robustness of PLS estimates against multicollinearity in the data. So correlations are introduced between manifest variables, and between exogenous latent variables. Again the ECSI model is tested and authors conclude that PLS Path Modelling is

Table 6.1: Comparison of results of previous studies

Model	**Simulation conditions**	**Sample size (n)**	**Latent variables**	**Manifest variables**	**Recovering True population parameters**
Cassel, Hackl and Westlund (1999). 3 formative exogenous and 1 reflective endogenous latent variables.	Sample size: 50, 200, 1000. Indicator per latent variable: 3 in formative block of variables, 4 in reflective block of variables. True path coefficients: 0.8, 0.1, 0.1. True loadings: 1.1, 1, 0.9, 0.8. True weights: 1/3. 500 replications.	The biases of the estimates seem to be unaffected by the increasing sample size. Increasing n lowered standard errors.	If a latent variable with a large regression coefficient is omitted, the remaining coefficient have a large upward bias.	The biases can be expected to be reduced only when the number of explanatory variables in the corresponding submodel is increased. In the presence of skew data, the medium-sized weights showed "an upward bias."	Basic Model: PLS performs very well, even in small samples. Bias in weights is < 10%. Bias in loadings is about 30%, PLS tends to underestimate the true loadings. The estimates of small coefficients are close to the true values; large coefficients show a downwards bias (20%).
Chin and Newated (1999). 1 reflective exogenous and M reflective endogenous latent variables.	Sample size: 20, 50, 100, 150 and 200 cases. Number of latent variables: 2, 4, 8, 12 and 16. Indicators per latent variable: 4, 8, 12, 16 or 32. True path coefficients: 0.4. True loadings: 0.2, 0.6, 0.8. 100 replications. Data were generated using PRELIS 2.14.	Increasing n alone does not provide better approximation to the population values. Increasing n lowered standard errors.	The number of latent variables did not seem to help the loading estimates.	The number of observed variables has to increase together with n to improve the approximation to the population values.	PLS always performed better than the summed regression approach. With greater number of indicators (16 or more) the PLS and regression estimates were essentially same. Loadings of 0.2 was not detected until n=150 and 200. Loadings of 0.6 and 0.8 were detected with n=20. Minimum n for a medium effect size 0.4 is 53.

robust under the presence of multicollinearity. The case analyzed for these authors is similar to that examined by Vilares *et al.* (2010).

6.4 Designing the Monte Carlo Simulation Study

A Monte Carlo simulation study was performed to address several issues (Paxton *et al.*, 2001; Gentle, 2003). The aims were:

1. To analyze the performance of PLS Path Modelling when considering formative measurement models.

2. To analyze the performance of PLS Path Modelling when few indicators are

considered per unobservable variable.

3. To analyze the performance of PLS Path Modelling when considering different sample sizes.

Based on a patent value model (Martínez-Ruiz & Aluja-Banet, 2009), the underlying population model considered a simple structure with three formative exogenous constructs and one reflective endogenous latent variable (Figure 6.1). The experimental design considered models with two, four, six and eight indicators per construct, and four different sample sizes (50, 100, 250, 500) were studied. Five hundred random data sets were generated for each of the 4×4 cells of the two-factor design. PLS Path Modelling with centroid scheme –as described in Tenenhaus *et al.* (2005)– was performed in R-project (R Development Core Team, 2007)[4]. Five hundred replications (t) were made for each cell in the design. Results are provided in terms of the:

- Mean value of the estimates.
- Mean standard deviation.
- Mean confidence intervals.
- Bias (accuracy, $\frac{1}{t}\sum_{i=1}^{t} E[\theta_i] - \theta$).
- Variance.
- Mean square error (precision, MSE= $Bias^2 + Variance$).
- Mean relative bias (MRB= $100 * \frac{1}{t}\sum_{i=1}^{t} \frac{\theta - E[\theta_i]}{\theta}$, Chin *et al.* (2003)).

6.4.1 Generating data

To generate data, we considered three different strategies with the aim of having a baseline model and two other cases. The data were generated from a component-based model (Schneeweiss, 1991; Chin & Newsted, 1999). Although it is also possible to work with centered variables, we began generating standardized manifest variables x_{jh} for each formative outer model as independent normal data (case A). We assumed that the covariances –in this case equal to the correlations– among manifest variables are zero. This is quite consistent with the literature review above, where manifest variables in a formative measurement model do not have to have a special type of relationship and should rather represent different

[4]Results of the implemented algorithm were obtained with several data sets and contrasted with those obtained using SmartPLS (Ringle *et al.*, 2005b); we also replicated part of Ringle et al.'s and Cassel et al.'s studies to ensure reliable outcomes.

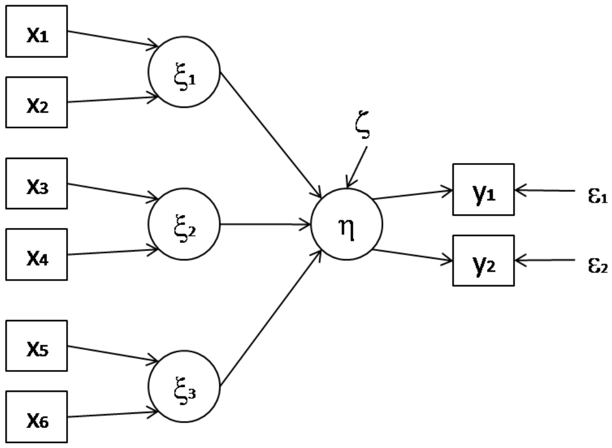


Figure 6.1: Structural and measurement models of the simulated setups; measurement models consider two, four, six, and eight indicators per construct.

facets of a construct. Once the manifest variables were generated, we computed the exogenous constructs ξ_j and the endogenous latent variable η, so that the variance of the unobservable variables is one.

Case A considered that disturbance terms δ in the inner relationships and errors ϵ_i in the outer relationships are zero. So, this baseline model allowed us to observe how close the estimates of the true values in an ideal situation are. Finally, standardized observed variables y_i of reflective measurement models were generated as independent normal data.

Case B followed the same procedure as in case A, except that the endogenous latent variable was calculated as a linear combination of the exogenous constructs plus a disturbance term. Disturbance terms were computed as random normal data with a zero mean and the corresponding standard deviation. They were distributed independently of unobservable variables. Errors of the reflective relationships were computed as random normal data with a zero mean and the corresponding standard deviation; they were also uncorrelated with the latent variable.

The more general case C took into account the correlations between manifest variables of the formative blocks of variables as well as the correlations between exogenous constructs. It is worth noting that in all cases, the generated exogenous constructs are not collinear.

Schneeweiss (1991) pointed out that "in a PLS model we only postulate which block of variables should depend on which other block(s) of variables, and the latent variables and their relations to each other are constructed by definition to reflect these dependencies. The covariance structure of the manifest variables is unrestricted" (p. 146). Nevertheless –and as Dijkstra (1983), Dijkstra (1992) and Dijkstra (2010) have stressed– covariance structure of the variables "restricts the feasible models" and this should also be considered in PLS path models.

6.4.2 Setting the true population parameters

To set the true population parameters for the models, we took into account different combinations of permissible values so as to see whether they are recovered by the PLS Path Modelling algorithm. Moreover, the behavior of the algorithm when there are few indicators per construct is of special interest. In social sciences, particularly in the literature regarding technological change and technology watch, the relationships between variables are often small or moderate[5]. So, we wish to establish whether or not this method is able to recover this type of relationship. Hence, the values of the coefficients were also set up in an attempt to get closer to the obtained coefficients when estimating patent value models. Table 6.2 shows the true population values of weights, path coefficients and loadings. We consider large values for all the true loadings, at least 0.7 in the case of two manifest variables per construct. This ensures the unidimensionality of the block of variables and it satisfies the condition imposed by the PLS Path Modelling algorithm.

Table 6.2: Vectors of true population values for weights, path coefficients and loadings; a model with three formative exogenous constructs and one reflective endogenous latent variable; cases for two, four, six and eight indicators in each outer model.

MVs	**Coefficient**	**Case A**	**Case B**	**Case C**
2	Weights	(0.8,0.6)	(0.8,0.6)	(0.8,0.5)
		(0.4,0.917)	(0.4,0.917)	(0.4,0.8)
		(0.1,0.995)	(0.1,0.995)	(0.1,0.9)
	Path Coefficients	(0.5,0.4,0.768)	(0.5,0.4,0.6)	(0.5,0.4,0.6)[a]
	Loadings	(0.7,0.8)	(0.7,0.8)	(0.7,0.8)
4	Weights	(0.2,0.3,0.5,0.782)	(0.2,0.3,0.5,0.782)	(0.2,0.3,0.5,0.7)
		(0.2,0.4,0.6,0.663)	(0.2,0.4,0.6,0.663)	(0.2,0.4,0.6,0.5)
		(0.3,0.5,0.7,0.412)	(0.3,0.5,0.7,0.412)	(0.3,0.5,0.7,0.2)
	Path Coefficients	(0.5,0.4,0.768)	(0.5,0.4,0.6)	(0.5,0.4,0.6)[a]
	Loadings	(0.6,0.7,0.8,0.9)	(0.6,0.7,0.8,0.9)	(0.6,0.7,0.8,0.9)
6	Weights	(0.5,0.3,0.4,0.3,0.6,0.224)	(0.5,0.3,0.4,0.3,0.6,0.224)	(0.5,0.3,0.4,0.3,0.5,0.1)
		(0.2,0.4,0.6,0.4,0.2,0.490)	(0.2,0.4,0.6,0.4,0.2,0.490)	(0.2,0.4,0.6,0.4,0.2,0.3)
		(0.3,0.7,0.2,0.3,0.4,0.361)	(0.3,0.7,0.2,0.3,0.4,0.361)	(0.3,0.6,0.2,0.3,0.4,0.2)
	Path Coefficients	(0.5,0.4,0.768)	(0.5,0.4,0.6)	(0.5,0.4,0.6)[a]
	Loadings	(0.6,0.7,0.8,0.9,0.6,0.7)	(0.6,0.7,0.8,0.9,0.6,0.7)	(0.6,0.7,0.8,0.9,0.6,0.7)
8	Weights	(0.2,0.3,0.4,0.5,0.4,0.3,0.2,0.412)	(0.2,0.3,0.4,0.5,0.4,0.3,0.2,0.412)	(0.3,0.3,0.4,0.3,0.4,0.3,0.2,0.3)[b]
		(0.3,0.3,0.4,0.4,0.2,0.3,0.5,0.346)	(0.3,0.3,0.4,0.4,0.2,0.3,0.5,0.346)	(0.3,0.3,0.4,0.4,0.2,0.3,0.4,0.2)[b]
		(0.4,0.5,0.5,0.3,0.2,0.1,0.3,0.332)	(0.4,0.5,0.5,0.3,0.2,0.1,0.3,0.332)	(0.4,0.5,0.4,0.3,0.2,0.1,0.3,0.2)[b]
	Path Coefficients	(0.5,0.4,0.768)	(0.5,0.4,0.6)	(0.5,0.4,0.6)
	Loadings	(0.6,0.7,0.8,0.9,0.6,0.7,0.8,0.9)	(0.6,0.7,0.8,0.9,0.6,0.7,0.8,0.9)	(0.6,0.7,0.8,0.9,0.6,0.7,0.8,0.9)

[a] For N=50 the true path coefficient vector was (0.5,0.4,0.5).
[b] For N=50 and N=100 the true weight vectors were (0.3,0.1,0.4,0.3,0.4,0.3,0.2,0.2), (0.3,0.1,0.4,0.4,0.2,0.3,0.4,0.1), and (0.2,0.4,0.4,0.3,0.2,0.1,0.3,0.2).

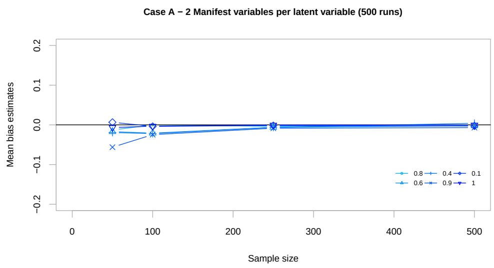
[5]Cohen (1988) suggest that correlations of 0.1, 0.3, and 0.5 express small, medium and large effect sizes, respectively.

6.5 Simulation Results

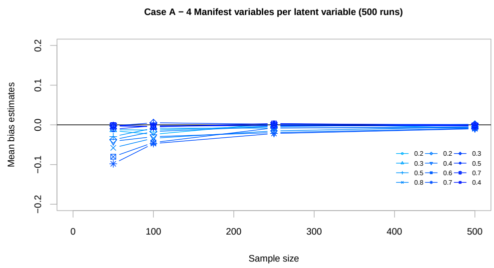
Figures 6.2 to 6.11 show the mean bias of the weights, path coefficients and loadings when sample sizes and the number of indicators increases for cases A, B and C. Figures 6.5 and 6.10 show the mean relative bias of a weight, loading and path coefficient depending on the sample size and the number of indicators. In the Appendix 6.7, Tables 6.3 to 6.44 show the mean estimates, standard deviation, confidence intervals, bias, variance, mean square error (MSE) and mean relative bias (MRB) for weights, path coefficients and loadings for the three analyzed cases.

6.5.1 Estimating Weights in Formative Outer Models

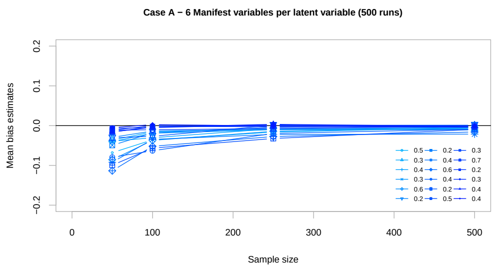
Figure 6.2 reports the mean bias of weight estimates for case-A models with two, four, six and eight indicators when the sample size varies from 50 to 500 observations. The statistical assumptions considered for this case –that is, uncorrelated variables and no disturbance terms nor errors– clearly show their influence on the weight estimates, and the algorithm underestimates the true weights for all sample sizes. This confirms that PLS estimates are biased. Increasing the sample size, the bias decreases, and for N=500, PLS almost exactly recovers all the population values (small, moderate and large values).



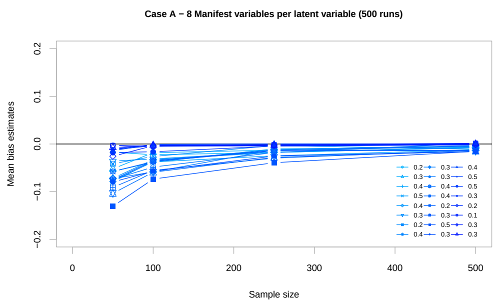
(a) Two indicators



(b) Four indicators



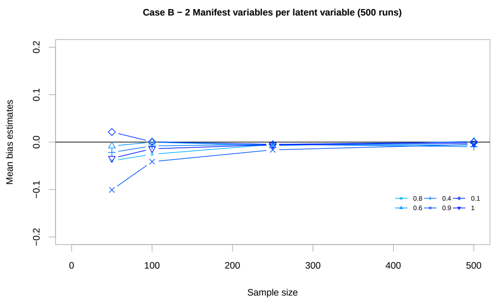
(c) Six indicators



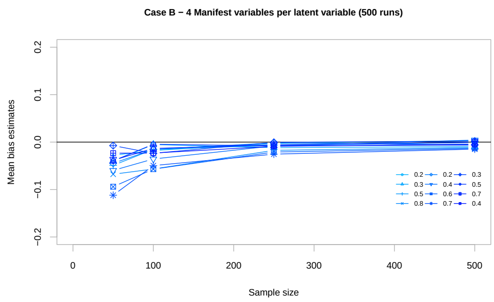
(d) Eight indicators

Figure 6.2: Mean bias of weight estimates for baseline case A. Highlighting the influence of the sample size and the number of indicators.

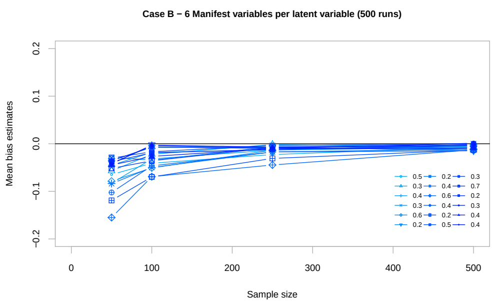
Unlike case A, case B considers disturbance terms and errors, and observations are generated as normal and independent data. This is a more realistic setting, but correlations between manifest variables, and between latent variables are not considered. Case B performs in a similar way to case A (see Figure 6.3). The true weights are underestimated, and as expected, the biases are larger compared to those obtained in case A. For four, six and eight indicators per construct, the largest MRBs are 22%, 26% and 33% when N=50, respectively (see Tables 6.17 to 6.22). The MRB decreases with increasing sample size. Besides case-B assumptions, the more general case C takes into account the correlations between manifest variables, and between latent variables. The empirical results are in line with the theoretical PLS framework (Dijkstra, 2010), and the true weights are overestimated by the PLS Path Modelling algorithm. Figure 6.4 clearly shows this and also how the biases decrease with increasing sample size and number of indicators for all analyzed cases. For case C, the largest MRBs are exhibited for models with the smallest sample sizes (N=50). Figure 6.5(a) and Tables 6.31 to 6.17 report the results.



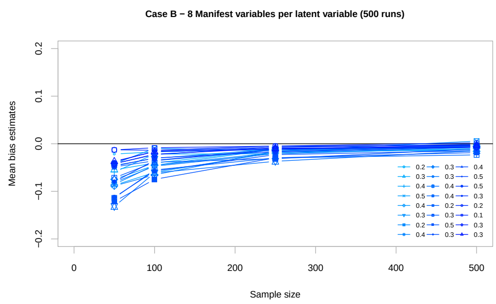
(a) Case B - Two indicators



(b) Case B - Four indicators



(c) Case B - Six indicators

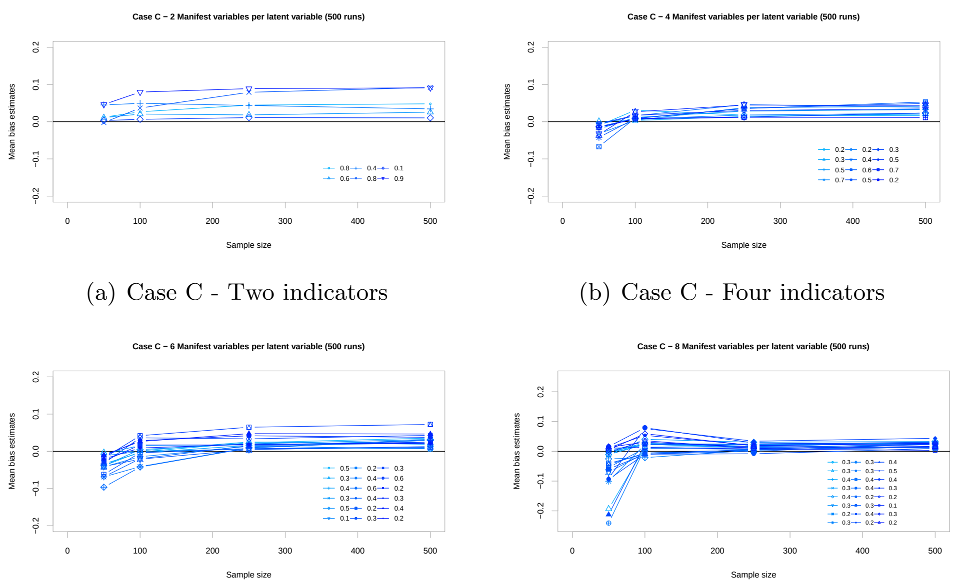


(d) Case B - Eight indicators

Figure 6.3: Mean bias of weight estimates for cases B. Highlighting the influence of the sample size and the number of indicators.

Interestingly, for cases A and B, the PLS algorithm yields similarly when considering two, four, six or eight indicators per construct. Higher values of bias in models with only two manifest variables are not observed when compared to models with four, six or eight indicators per construct. Quite the contrary. For small sample sizes (N=50), the bias proved to be slightly higher when four, six or eight indicators were included in the formative models. However, for the more realistic case C, the number of manifest variables helps to decrease the bias of the estimates. This can be seen in Figure 6.5(a). In addition, the variability and MSE decrease by increasing the sample size or increasing the number of manifest variables in all the simulated cases (see Tables 6.3 to 6.8, 6.17 to 6.22, 6.31 to 6.36).

Simulations performed by Chin & Newsted (1999) for PLS models with reflective relationships showed that, by themselves, neither the number of indicators nor the sample size substantively improve the quality of the estimates. Rather, it is necessary to increase both factors at the same time for an improvement in the quality of the estimates. Here, the simulations for PLS models with formative-reflective blocks of variables render the same aforementioned result. So, PLS Path Modelling is consistent and consistent at large. Nevertheless –and recalling that PLS algorithm computes the latent variables as an exact linear combination of the observed variables– the results suggest that in real-world applications with formative outer models, estimates will improve by increasing the sample size more than increasing the number of observable variables, depending on the correlations between manifest variables. So, the researcher may suspect the type of relationship that she expects to find.



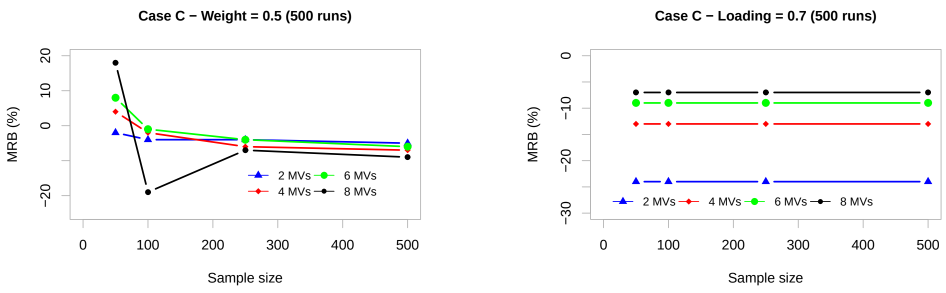
(a) Case C - Two indicators

(b) Case C - Four indicators

(c) Case C - Six indicators

(d) Case C - Eight indicators

Figure 6.4: Mean bias of weight estimates for cases C. Highlighting the influence of the sample size and the number of indicators.



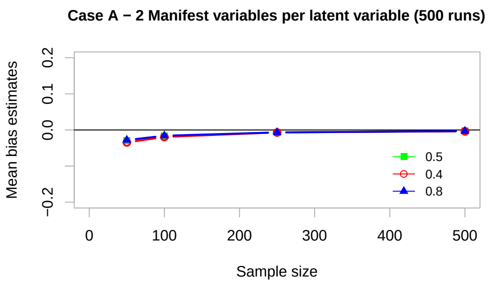
(a) MRB for a weight of 0.5 (b) MRB for a loading of 0.7

Figure 6.5: Mean relative bias of a weight and a loading, case C. Highlighting the influence of the sample size and the number of indicators per construct.

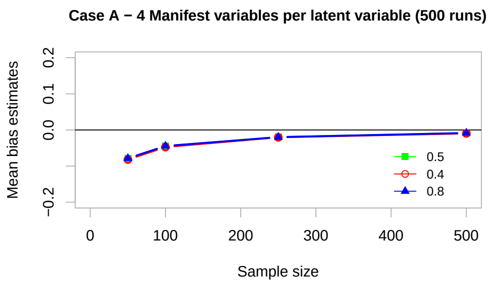
6.5.2 Estimating Path Coefficients in Structural Models

Results for estimates of inner relationships are quite conclusive. For baseline case A, the algorithm underestimates the true path coefficients in all the analyzed cases (see Figure 6.6). This means that PLS Path Modelling estimates are biased even though no disturbance terms and no errors are considered. As the sample size increases, the estimates increasingly approach to true values and the biases decrease. By introducing disturbance terms and errors, as in case B, as well as correlations, as in case C, PLS Path Modelling also underestimates the structural relationships.

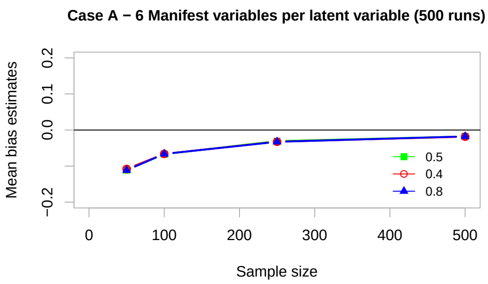
Figures 6.7, 6.8 and 6.9 allow us to see how an increase in both the number of manifest variables and the sample size reduces the bias of the estimates for both cases B and C, for all assumed true values. The variability and the mean square errors of the estimates also tend to decrease with increasing sample size. In addition, by way of example, Figure 6.10 shows the mean relative bias of the path coefficients whose true value are 0.5. Figure 6.10(a) reports the change in the MRB when the sample size increases for models with two, four, six and eight manifest variables in the measurement models. Case C is shown here. In accordance with expectations, when the models consider only two indicators, the MRB is the largest and proves to be quite the same when sample size increases. Its value ranges from 18% (N=50) to 15% (N=500). For models with four, six and eight indicators, the MRB decreases when sample size increases. Figure 6.10(a) also shows that by increasing the number of indicators per latent variable, it yields closer estimates to the true values; but this factor tends to have less influence on the quality of the estimates than the sample size does. Figure 6.10(b) compares the MRB of cases A, B and C for models with four indicators per construct when the sample size increases. As can be seen, the MRB does not exceed 20%. For case C, the mean relative biases are about 15% and 10% for N=50 and N=500, respectively. As



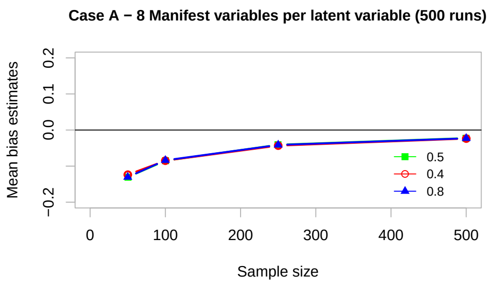
(a) Two indicators



(b) Four indicators



(c) Six indicators



(d) Eight indicators

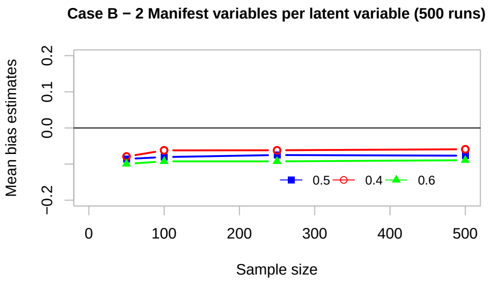
Figure 6.6: Mean bias of path coefficients for baseline case A. Highlighting the influence of the sample size and the number of indicators.

expected, case A shows the lowest MRB. See Tables 6.9 to 6.12, 6.23 to 6.26, 6.37 to 6.40 for the full list of results.

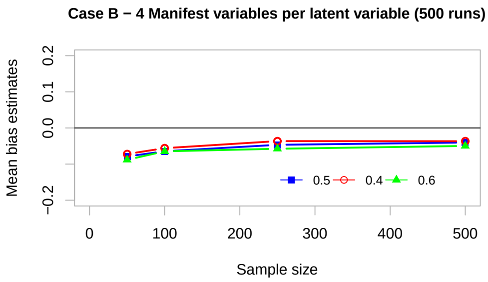
Summing up, in all analyzed cases, the results obtained are better when each outer model considers more manifest variables per construct. However, it is worth noting that the estimates are shown to be of the same level of accuracy and precision when the measurement models include only two indicators per construct. This suggests that PLS Path Modelling may be a robust alternative when estimating structural equation models with formative relationships and few indicators per construct.

6.5.3 Estimating Loadings in Reflective Outer Models

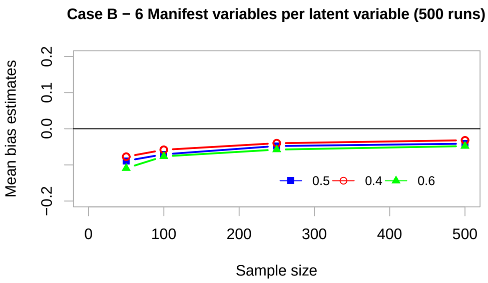
As can be seen in Figure 6.11, the estimates of loadings in all cases are very close to the true values, regardless of the sample sizes and number of manifest variables per construct. PLS Path Modelling overestimates the population values. Moreover, according to the results, a higher number of manifest variables seems to be more important than a higher sample size for decreasing the bias of the estimates in reflective outer models. This is in contrast with the formative relationships and coincides with the results found by other researchers (Chin & Newsted, 1999).



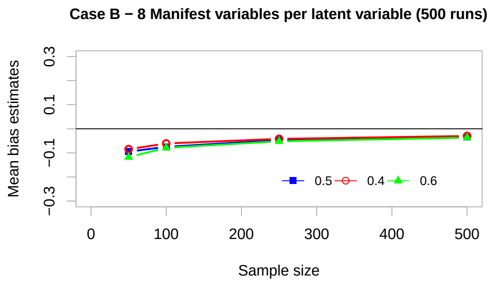
(a) Case B - Two indicators



(b) Case B - Four indicators



(c) Case B - Six indicators



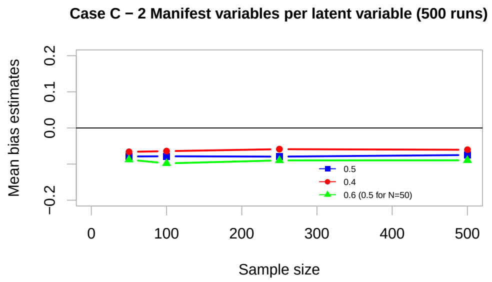
(d) Case B - Eight indicators

Figure 6.7: Mean bias of path coefficients for cases B. Highlighting the influence of the sample size and the number of indicators.

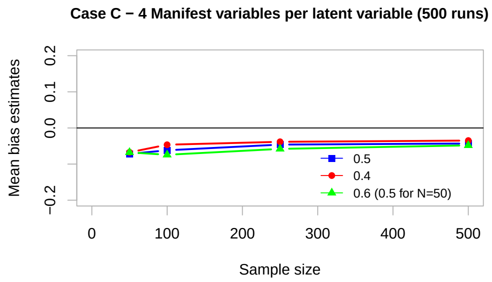
This is clearly seen in Figure 6.5(b) where the mean relative bias for a loading of 0.7 strongly decreases when the number of indicators increases. So, this confirms that PLS estimates are "inconsistent" in a reflective setting (Hui & Wold, 1982, p. 123); they are only consistent at large (Wold, 1982; Chin & Newsted, 1999).

6.6 Final Remarks

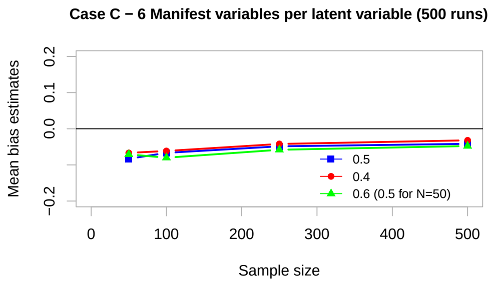
For the studied model, the findings suggest that PLS Path Modelling offers a way to build "proper indices" for unobservable variables and to estimate the relationships between them. The estimates are always biased. The procedure shows a tendency to overestimate outer relationships and underestimate inner relationships. It is worth noting, however, that the estimates are shown to be robust when the measurement models include only two indicators per construct. That is, the procedure is able to recover proper estimates of the true values. It is true that when the number of observed variables and sample size increase, the quality of the PLS Path Modelling estimates increases. But when few indicators and a small sample size are considered, we can obtain estimates of the parameters, at least for the simulated case. Vilares *et al.* (2010) have noted the same behavior



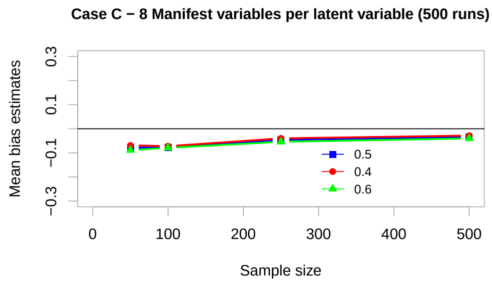
(a) Case C - Two indicators



(b) Case C - Four indicators



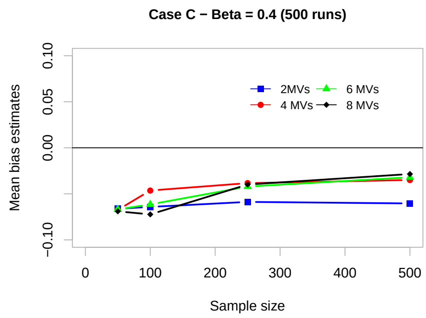
(c) Case C - Six indicators



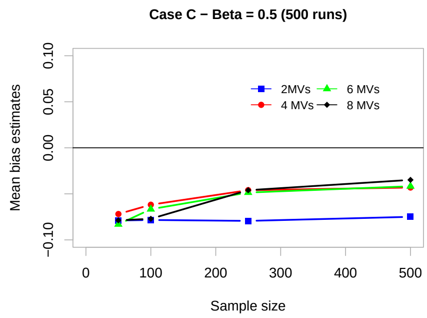
(d) Case C - Eight indicators

Figure 6.8: Mean bias of path coefficients for cases C. Highlighting the influence of the sample size and the number of indicators.

in a reflective blocks of variables with two indicators. Finally, we think that the model simulated here represents a number of models that can be studied in real-world applications: those in which formative exogenous outer models are modeled using PLS Mode B and reflective endogenous latent variables are modeled using PLS Mode A. That is, PLS Mode C, in terms of Wold's approach.

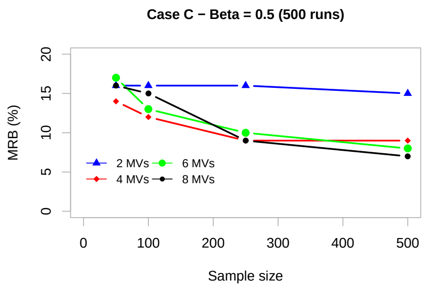


(a) True value = 0.4

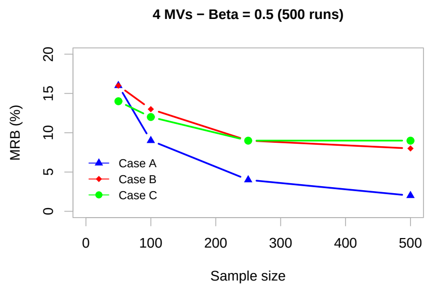


(b) True value = 0.5

Figure 6.9: Mean bias of path coefficients. Highlighting the influence of the sample size and the number of indicators.

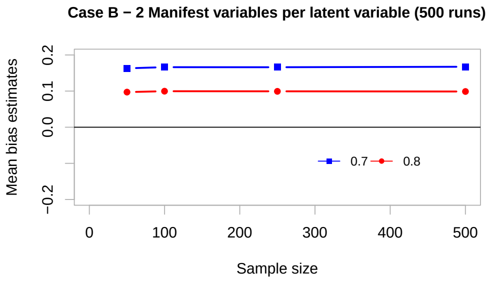


(a) MRB for case C

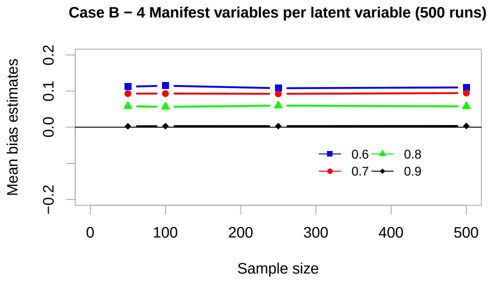


(b) MRB for models with 4 manifest variables in the outer models

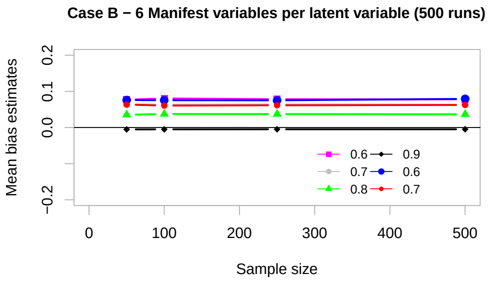
Figure 6.10: Mean relative bias of a path coefficient, true value equal to 0.5. Highlighting the influence of the sample size and the number of indicators per construct.



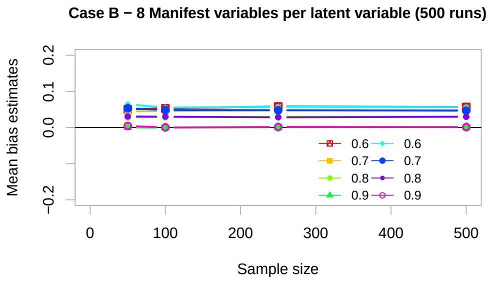
(a) Case B - Two indicators



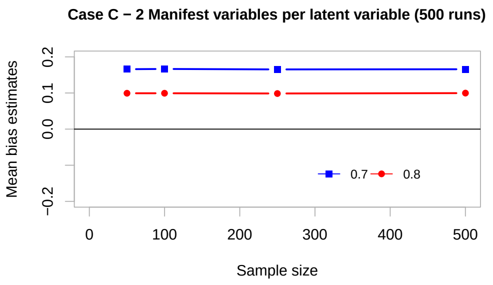
(b) Case B - Four indicators



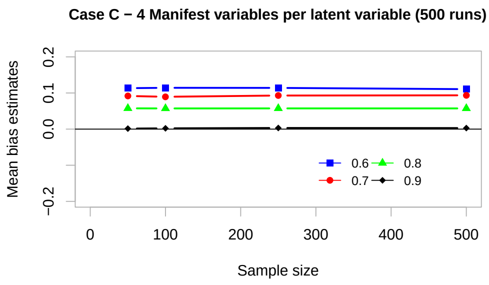
(c) Case B - Six indicators



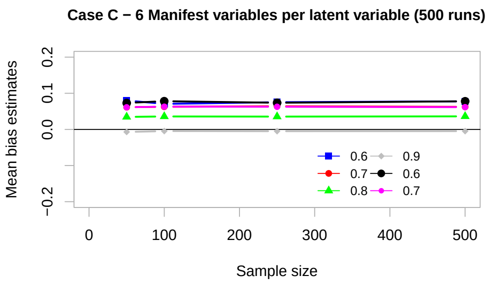
(d) Case B - Eight indicators



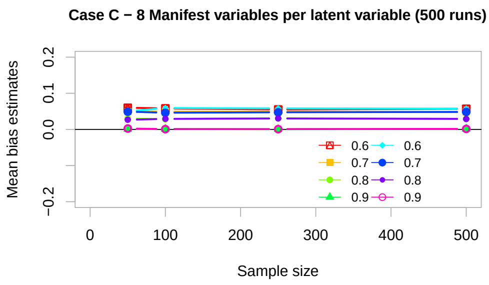
(e) Case C - Two indicators



(f) Case C - Four indicators



(g) Case C - Six indicators



(h) Case C - Eight indicators

Figure 6.11: Mean bias of loadings for cases B and C. Highlighting the influence of the sample size and the number of indicators.

6.7 Appendix: Tables

Table 6.3: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with two indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.800	0.787	0.158	0.772	0.803	-0.013	0.025	0.025	2
		0.600	0.583	0.206	0.570	0.597	-0.017	0.043	0.043	3
	2	0.400	0.381	0.303	0.350	0.409	-0.019	0.092	0.092	5
		0.917	0.860	0.192	0.843	0.872	-0.056	0.037	0.040	6
	3	0.100	0.107	0.119	0.094	0.116	0.007	0.014	0.014	-7
		0.995	0.988	0.028	0.986	0.991	-0.007	0.001	0.001	1
100	1	0.800	0.800	0.116	0.788	0.810	0.000	0.013	0.013	0
		0.600	0.579	0.148	0.567	0.591	-0.021	0.022	0.022	3
	2	0.400	0.378	0.231	0.360	0.406	-0.022	0.053	0.054	5
		0.917	0.892	0.107	0.881	0.900	-0.025	0.011	0.012	3
	3	0.100	0.096	0.084	0.088	0.104	-0.004	0.007	0.007	4
		0.995	0.992	0.016	0.991	0.993	-0.003	0.000	0.000	0
250	1	0.800	0.796	0.072	0.791	0.804	-0.004	0.005	0.005	0
		0.600	0.594	0.094	0.587	0.603	-0.006	0.009	0.009	1
	2	0.400	0.394	0.134	0.382	0.406	-0.006	0.018	0.018	2
		0.917	0.908	0.064	0.903	0.914	-0.008	0.004	0.004	1
	3	0.100	0.098	0.051	0.093	0.102	-0.002	0.003	0.003	2
		0.995	0.994	0.009	0.994	0.995	-0.001	0.000	0.000	0
500	1	0.800	0.798	0.053	0.792	0.803	-0.002	0.003	0.003	0
		0.600	0.598	0.066	0.592	0.604	-0.002	0.004	0.004	0
	2	0.400	0.404	0.099	0.396	0.413	0.004	0.010	0.010	-1
		0.917	0.910	0.046	0.906	0.913	-0.007	0.002	0.002	1
	3	0.100	0.097	0.036	0.094	0.101	-0.003	0.001	0.001	3
		0.995	0.995	0.006	0.994	0.995	0.000	0.000	0.000	0

Table 6.4: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with four indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.200	0.185	0.235	0.166	0.208	-0.015	0.055	0.056	7
		0.300	0.288	0.215	0.267	0.312	-0.012	0.046	0.046	4
		0.500	0.470	0.209	0.451	0.493	-0.030	0.044	0.045	6
		0.787	0.729	0.177	0.712	0.744	-0.058	0.031	0.035	7
	2	0.200	0.162	0.301	0.136	0.193	-0.038	0.091	0.092	19
		0.400	0.359	0.275	0.335	0.382	-0.041	0.075	0.077	10
		0.600	0.520	0.260	0.490	0.549	-0.080	0.068	0.074	13
		0.663	0.565	0.261	0.545	0.589	-0.098	0.068	0.078	15
	3	0.300	0.297	0.113	0.284	0.312	-0.003	0.013	0.013	1
		0.500	0.489	0.112	0.479	0.502	-0.011	0.013	0.013	2
		0.700	0.697	0.105	0.689	0.705	-0.003	0.011	0.011	0
		0.412	0.412	0.114	0.401	0.424	-0.001	0.013	0.013	0
100	1	0.200	0.176	0.152	0.166	0.188	-0.024	0.023	0.024	12
		0.300	0.287	0.166	0.271	0.304	-0.013	0.028	0.028	4
		0.500	0.492	0.144	0.478	0.505	-0.008	0.021	0.021	2
		0.787	0.754	0.118	0.744	0.764	-0.034	0.014	0.015	4
	2	0.200	0.183	0.222	0.167	0.205	-0.017	0.049	0.050	8
		0.400	0.370	0.213	0.348	0.392	-0.030	0.046	0.046	7
		0.600	0.555	0.189	0.529	0.579	-0.045	0.036	0.038	7
		0.663	0.616	0.168	0.602	0.631	-0.047	0.028	0.030	7
	3	0.300	0.305	0.079	0.299	0.311	0.005	0.006	0.006	-2
		0.500	0.498	0.080	0.489	0.507	-0.002	0.006	0.006	0
		0.700	0.696	0.076	0.691	0.703	-0.004	0.006	0.006	1
		0.412	0.407	0.081	0.402	0.414	-0.005	0.006	0.007	1
250	1	0.200	0.204	0.107	0.196	0.214	0.004	0.011	0.011	-2
		0.300	0.297	0.107	0.286	0.307	-0.003	0.011	0.011	1
		0.500	0.492	0.093	0.486	0.497	-0.008	0.009	0.009	2
		0.787	0.772	0.072	0.766	0.781	-0.015	0.005	0.005	2
	2	0.200	0.196	0.140	0.184	0.207	-0.004	0.020	0.020	2
		0.400	0.381	0.131	0.368	0.395	-0.019	0.017	0.017	5
		0.600	0.592	0.119	0.579	0.604	-0.008	0.014	0.014	1
		0.663	0.642	0.118	0.630	0.654	-0.022	0.014	0.014	3
	3	0.300	0.302	0.051	0.295	0.308	0.002	0.003	0.003	-1
		0.500	0.498	0.050	0.494	0.502	-0.002	0.003	0.003	0
		0.700	0.699	0.047	0.695	0.703	-0.001	0.002	0.002	0
		0.412	0.416	0.051	0.411	0.420	0.003	0.003	0.003	-1
500	1	0.200	0.192	0.079	0.184	0.199	-0.008	0.006	0.006	4
		0.300	0.297	0.075	0.291	0.304	-0.003	0.006	0.006	1
		0.500	0.495	0.067	0.486	0.504	-0.005	0.005	0.005	1
		0.787	0.782	0.049	0.778	0.786	-0.005	0.002	0.002	1
	2	0.200	0.201	0.094	0.196	0.209	0.001	0.009	0.009	-1
		0.400	0.392	0.093	0.383	0.400	-0.008	0.009	0.009	2
		0.600	0.592	0.085	0.582	0.601	-0.008	0.007	0.007	1
		0.663	0.653	0.080	0.647	0.661	-0.010	0.006	0.006	1
	3	0.300	0.301	0.037	0.297	0.305	0.001	0.001	0.001	0
		0.500	0.498	0.033	0.494	0.501	-0.002	0.001	0.001	0
		0.700	0.697	0.032	0.694	0.701	-0.003	0.001	0.001	0
		0.412	0.411	0.036	0.408	0.414	-0.001	0.001	0.001	0

Table 6.5: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with six indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.500	0.432	0.218	0.415	0.447	-0.068	0.048	0.052	14
		0.300	0.264	0.231	0.249	0.286	-0.036	0.053	0.055	12
		0.400	0.361	0.217	0.342	0.383	-0.039	0.047	0.048	10
		0.300	0.260	0.226	0.244	0.279	-0.040	0.051	0.053	13
		0.600	0.514	0.205	0.496	0.530	-0.086	0.042	0.049	14
		0.224	0.194	0.231	0.170	0.217	-0.029	0.053	0.054	13
	2	0.200	0.150	0.278	0.127	0.176	-0.050	0.077	0.080	25
		0.400	0.308	0.264	0.285	0.333	-0.092	0.070	0.078	23
		0.600	0.486	0.259	0.470	0.511	-0.114	0.067	0.080	19
		0.400	0.321	0.290	0.292	0.352	-0.079	0.084	0.090	20
		0.200	0.169	0.273	0.149	0.190	-0.031	0.075	0.076	15
		0.490	0.390	0.258	0.366	0.422	-0.100	0.066	0.076	20
	3	0.300	0.284	0.114	0.272	0.295	-0.016	0.013	0.013	5
		0.700	0.689	0.110	0.679	0.700	-0.011	0.012	0.012	2
		0.200	0.193	0.124	0.177	0.209	-0.007	0.015	0.015	3
		0.300	0.289	0.123	0.281	0.297	-0.011	0.015	0.015	4
		0.400	0.384	0.120	0.373	0.397	-0.016	0.014	0.015	4
		0.361	0.355	0.123	0.347	0.362	-0.005	0.015	0.015	1
100	1	0.500	0.464	0.146	0.454	0.478	-0.036	0.021	0.022	7
		0.300	0.284	0.156	0.270	0.301	-0.016	0.024	0.025	5
		0.400	0.369	0.158	0.362	0.377	-0.031	0.025	0.026	8
		0.300	0.281	0.166	0.268	0.298	-0.019	0.028	0.028	6
		0.600	0.563	0.141	0.549	0.583	-0.037	0.020	0.021	6
		0.224	0.210	0.168	0.192	0.231	-0.014	0.028	0.028	6
	2	0.200	0.182	0.214	0.161	0.204	-0.018	0.046	0.046	9
		0.400	0.365	0.204	0.343	0.387	-0.035	0.042	0.043	9
		0.600	0.548	0.181	0.537	0.560	-0.052	0.033	0.035	9
		0.400	0.337	0.200	0.316	0.353	-0.063	0.040	0.044	16
		0.200	0.173	0.201	0.155	0.191	-0.027	0.041	0.041	13
		0.490	0.433	0.197	0.412	0.450	-0.057	0.039	0.042	12
	3	0.300	0.298	0.076	0.292	0.305	-0.002	0.006	0.006	1
		0.700	0.690	0.070	0.683	0.697	-0.010	0.005	0.005	1
		0.200	0.195	0.082	0.187	0.203	-0.005	0.007	0.007	3
		0.300	0.300	0.081	0.293	0.306	0.000	0.007	0.007	0
		0.400	0.396	0.081	0.389	0.403	-0.004	0.007	0.007	1
		0.361	0.363	0.082	0.357	0.369	0.003	0.007	0.007	-1

Table 6.6: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with six indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.500	0.484	0.096	0.475	0.493	-0.016	0.009	0.009	3
		0.300	0.291	0.103	0.279	0.302	-0.009	0.011	0.011	3
		0.400	0.392	0.099	0.382	0.401	-0.008	0.010	0.010	2
		0.300	0.290	0.107	0.280	0.302	-0.010	0.011	0.011	3
		0.600	0.586	0.093	0.577	0.594	-0.014	0.009	0.009	2
		0.224	0.214	0.106	0.205	0.223	-0.009	0.011	0.011	4
	2	0.200	0.196	0.142	0.181	0.215	-0.004	0.020	0.020	2
		0.400	0.378	0.127	0.369	0.387	-0.022	0.016	0.016	6
		0.600	0.573	0.114	0.562	0.585	-0.027	0.013	0.014	5
		0.400	0.380	0.136	0.370	0.392	-0.020	0.019	0.019	5
		0.200	0.201	0.140	0.187	0.216	0.001	0.020	0.020	0
		0.490	0.458	0.122	0.445	0.470	-0.032	0.015	0.016	7
	3	0.300	0.298	0.053	0.292	0.303	-0.002	0.003	0.003	1
		0.700	0.692	0.046	0.688	0.696	-0.008	0.002	0.002	1
		0.200	0.202	0.051	0.197	0.208	0.002	0.003	0.003	-1
		0.300	0.297	0.052	0.293	0.302	-0.003	0.003	0.003	1
		0.400	0.403	0.047	0.399	0.407	0.003	0.002	0.002	-1
		0.361	0.360	0.051	0.355	0.364	-0.001	0.003	0.003	0
500	1	0.500	0.495	0.070	0.490	0.500	-0.005	0.005	0.005	1
		0.300	0.293	0.076	0.285	0.300	-0.007	0.006	0.006	2
		0.400	0.389	0.074	0.383	0.397	-0.011	0.005	0.006	3
		0.300	0.291	0.082	0.285	0.298	-0.009	0.007	0.007	3
		0.600	0.595	0.069	0.589	0.600	-0.005	0.005	0.005	1
		0.224	0.220	0.075	0.213	0.227	-0.004	0.006	0.006	2
	2	0.200	0.201	0.100	0.193	0.208	0.001	0.010	0.010	0
		0.400	0.379	0.098	0.367	0.391	-0.021	0.010	0.010	5
		0.600	0.584	0.082	0.578	0.589	-0.016	0.007	0.007	3
		0.400	0.391	0.095	0.382	0.403	-0.009	0.009	0.009	2
		0.200	0.201	0.101	0.192	0.209	0.001	0.010	0.010	-1
		0.490	0.480	0.093	0.472	0.488	-0.010	0.009	0.009	2
	3	0.300	0.296	0.036	0.292	0.300	-0.004	0.001	0.001	1
		0.700	0.698	0.033	0.696	0.701	-0.002	0.001	0.001	0
		0.200	0.196	0.036	0.193	0.199	-0.004	0.001	0.001	2
		0.300	0.298	0.038	0.294	0.302	-0.002	0.001	0.001	1
		0.400	0.398	0.037	0.395	0.402	-0.002	0.001	0.001	0
		0.361	0.359	0.038	0.356	0.363	-0.001	0.001	0.001	0

Table 6.7: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with eight indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.200	0.183	0.223	0.165	0.201	-0.017	0.050	0.050	9
		0.300	0.248	0.226	0.225	0.268	-0.052	0.051	0.054	17
		0.400	0.343	0.227	0.325	0.359	-0.057	0.052	0.055	14
		0.500	0.424	0.207	0.410	0.441	-0.076	0.043	0.049	15
		0.400	0.328	0.215	0.302	0.346	-0.072	0.046	0.052	18
		0.300	0.264	0.217	0.248	0.279	-0.036	0.047	0.048	12
		0.200	0.159	0.225	0.136	0.182	-0.041	0.050	0.052	21
		0.412	0.333	0.217	0.313	0.354	-0.079	0.047	0.053	19
	2	0.300	0.227	0.248	0.197	0.259	-0.073	0.061	0.067	24
		0.300	0.226	0.266	0.200	0.247	-0.074	0.071	0.076	25
		0.400	0.296	0.263	0.274	0.322	-0.104	0.069	0.080	26
		0.400	0.309	0.256	0.285	0.331	-0.091	0.066	0.074	23
		0.200	0.142	0.267	0.111	0.173	-0.058	0.071	0.075	29
		0.300	0.225	0.245	0.202	0.253	-0.075	0.060	0.066	25
		0.500	0.369	0.274	0.334	0.402	-0.131	0.075	0.092	26
		0.346	0.267	0.258	0.242	0.297	-0.080	0.067	0.073	23
	3	0.400	0.386	0.120	0.376	0.397	-0.014	0.014	0.015	3
		0.500	0.488	0.117	0.479	0.499	-0.012	0.014	0.014	2
		0.500	0.482	0.112	0.472	0.491	-0.018	0.013	0.013	4
		0.300	0.289	0.123	0.276	0.301	-0.011	0.015	0.015	4
		0.200	0.197	0.116	0.187	0.209	-0.003	0.014	0.014	2
		0.100	0.097	0.119	0.089	0.106	-0.003	0.014	0.014	3
		0.300	0.274	0.119	0.260	0.288	-0.026	0.014	0.015	9
		0.332	0.324	0.120	0.314	0.333	-0.008	0.014	0.014	2
100	1	0.200	0.177	0.162	0.165	0.192	-0.023	0.026	0.027	11
		0.300	0.282	0.156	0.268	0.294	-0.018	0.024	0.025	6
		0.400	0.362	0.156	0.348	0.376	-0.038	0.024	0.026	10
		0.500	0.462	0.150	0.450	0.476	-0.038	0.023	0.024	8
		0.400	0.367	0.159	0.353	0.377	-0.033	0.025	0.026	8
		0.300	0.269	0.160	0.251	0.284	-0.031	0.025	0.026	10
		0.200	0.174	0.167	0.161	0.185	-0.026	0.028	0.029	13
		0.412	0.376	0.150	0.361	0.389	-0.036	0.023	0.024	9
	2	0.300	0.241	0.205	0.226	0.255	-0.059	0.042	0.046	20
		0.300	0.252	0.203	0.229	0.278	-0.048	0.041	0.044	16
		0.400	0.341	0.184	0.322	0.364	-0.059	0.034	0.037	15
		0.400	0.345	0.191	0.329	0.360	-0.055	0.037	0.040	14
		0.200	0.163	0.206	0.145	0.183	-0.037	0.042	0.044	18
		0.300	0.265	0.207	0.243	0.289	-0.035	0.043	0.044	12
		0.500	0.426	0.187	0.409	0.446	-0.074	0.035	0.040	15
		0.346	0.289	0.199	0.273	0.306	-0.057	0.040	0.043	16
	3	0.400	0.399	0.080	0.392	0.408	-0.001	0.006	0.006	0
		0.500	0.496	0.080	0.488	0.505	-0.004	0.006	0.006	1
		0.500	0.484	0.082	0.474	0.492	-0.016	0.007	0.007	3
		0.300	0.299	0.083	0.290	0.305	-0.001	0.007	0.007	0
		0.200	0.195	0.087	0.187	0.202	-0.005	0.008	0.008	2
		0.100	0.095	0.082	0.087	0.104	-0.005	0.007	0.007	5
		0.300	0.296	0.085	0.289	0.303	-0.004	0.007	0.007	1
		0.332	0.329	0.086	0.320	0.337	-0.003	0.007	0.007	1

Table 6.8: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with eight indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.200	0.182	0.109	0.175	0.191	-0.018	0.012	0.012	9
		0.300	0.287	0.110	0.276	0.299	-0.013	0.012	0.012	4
		0.400	0.390	0.101	0.378	0.401	-0.010	0.010	0.010	3
		0.500	0.475	0.097	0.466	0.484	-0.025	0.009	0.010	5
		0.400	0.382	0.099	0.369	0.396	-0.018	0.010	0.010	4
		0.300	0.283	0.103	0.275	0.293	-0.017	0.011	0.011	6
		0.200	0.197	0.101	0.184	0.207	-0.003	0.010	0.010	2
		0.412	0.401	0.101	0.393	0.410	-0.012	0.010	0.010	3
	2	0.300	0.288	0.135	0.276	0.299	-0.012	0.018	0.018	4
		0.300	0.282	0.135	0.268	0.295	-0.018	0.018	0.019	6
		0.400	0.372	0.130	0.356	0.391	-0.028	0.017	0.018	7
		0.400	0.370	0.125	0.358	0.384	-0.030	0.016	0.016	7
		0.200	0.186	0.132	0.175	0.195	-0.014	0.017	0.018	7
		0.300	0.287	0.131	0.278	0.295	-0.013	0.017	0.017	4
		0.500	0.461	0.127	0.449	0.474	-0.039	0.016	0.018	8
		0.346	0.321	0.127	0.312	0.329	-0.025	0.016	0.017	7
	3	0.400	0.400	0.054	0.395	0.406	0.000	0.003	0.003	0
		0.500	0.496	0.052	0.492	0.500	-0.004	0.003	0.003	1
		0.500	0.496	0.053	0.491	0.501	-0.004	0.003	0.003	1
		0.300	0.299	0.052	0.295	0.303	-0.001	0.003	0.003	0
		0.200	0.196	0.054	0.193	0.200	-0.004	0.003	0.003	2
		0.100	0.098	0.053	0.095	0.102	-0.002	0.003	0.003	2
		0.300	0.298	0.052	0.294	0.301	-0.002	0.003	0.003	1
		0.332	0.329	0.053	0.322	0.334	-0.003	0.003	0.003	1
500	1	0.200	0.193	0.078	0.186	0.200	-0.007	0.006	0.006	3
		0.300	0.294	0.072	0.289	0.299	-0.006	0.005	0.005	2
		0.400	0.394	0.075	0.387	0.401	-0.006	0.006	0.006	2
		0.500	0.488	0.066	0.483	0.494	-0.012	0.004	0.004	2
		0.400	0.393	0.074	0.387	0.399	-0.007	0.006	0.006	2
		0.300	0.292	0.076	0.287	0.298	-0.008	0.006	0.006	3
		0.200	0.189	0.074	0.182	0.197	-0.011	0.005	0.006	6
		0.412	0.409	0.068	0.404	0.414	-0.003	0.005	0.005	1
	2	0.300	0.287	0.094	0.278	0.296	-0.013	0.009	0.009	4
		0.300	0.301	0.100	0.293	0.312	0.001	0.010	0.010	0
		0.400	0.385	0.095	0.375	0.393	-0.015	0.009	0.009	4
		0.400	0.387	0.097	0.380	0.396	-0.013	0.009	0.009	3
		0.200	0.196	0.098	0.187	0.204	-0.004	0.010	0.010	2
		0.300	0.286	0.097	0.275	0.296	-0.014	0.009	0.010	5
		0.500	0.485	0.085	0.475	0.493	-0.015	0.007	0.007	3
		0.346	0.330	0.094	0.323	0.338	-0.016	0.009	0.009	5
	3	0.400	0.400	0.034	0.398	0.403	0.000	0.001	0.001	0
		0.500	0.501	0.037	0.497	0.506	0.001	0.001	0.001	0
		0.500	0.498	0.034	0.495	0.501	-0.002	0.001	0.001	0
		0.300	0.300	0.039	0.297	0.303	0.000	0.002	0.002	0
		0.200	0.199	0.037	0.194	0.202	-0.001	0.001	0.001	1
		0.100	0.102	0.035	0.098	0.105	0.002	0.001	0.001	-2
		0.300	0.299	0.037	0.296	0.302	-0.001	0.001	0.001	0
		0.332	0.332	0.035	0.329	0.335	0.000	0.001	0.001	0

Table 6.9: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with two indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.471	0.040	0.467	0.474	-0.029	0.002	0.002	6
	0.400	0.365	0.053	0.361	0.370	-0.035	0.003	0.004	9
	0.768	0.740	0.052	0.736	0.745	-0.028	0.003	0.003	4
100	0.500	0.483	0.027	0.480	0.486	-0.017	0.001	0.001	3
	0.400	0.381	0.027	0.378	0.383	-0.019	0.001	0.001	5
	0.768	0.752	0.038	0.749	0.755	-0.016	0.001	0.002	2
250	0.500	0.493	0.016	0.492	0.495	-0.007	0.000	0.000	1
	0.400	0.393	0.014	0.392	0.394	-0.007	0.000	0.000	2
	0.768	0.761	0.024	0.759	0.764	-0.007	0.001	0.001	1
500	0.500	0.497	0.012	0.496	0.497	-0.003	0.000	0.000	1
	0.400	0.396	0.010	0.395	0.397	-0.004	0.000	0.000	1
	0.768	0.765	0.018	0.763	0.766	-0.003	0.000	0.000	0

Table 6.10: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with four indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.420	0.051	0.415	0.425	-0.080	0.003	0.009	16
	0.400	0.318	0.050	0.314	0.323	-0.082	0.002	0.009	20
	0.768	0.690	0.044	0.686	0.693	-0.079	0.002	0.008	10
100	0.500	0.455	0.028	0.452	0.458	-0.045	0.001	0.003	9
	0.400	0.353	0.030	0.350	0.355	-0.047	0.001	0.003	12
	0.768	0.724	0.034	0.721	0.728	-0.044	0.001	0.003	6
250	0.500	0.480	0.018	0.479	0.482	-0.020	0.000	0.001	4
	0.400	0.380	0.016	0.378	0.381	-0.020	0.000	0.001	5
	0.768	0.748	0.025	0.746	0.750	-0.020	0.001	0.001	3
500	0.500	0.491	0.012	0.491	0.493	-0.009	0.000	0.000	2
	0.400	0.390	0.010	0.390	0.391	-0.010	0.000	0.000	2
	0.768	0.760	0.017	0.758	0.762	-0.008	0.000	0.000	1

Table 6.11: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with six indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.389	0.049	0.384	0.393	-0.111	0.002	0.015	22
	0.400	0.292	0.059	0.287	0.299	-0.108	0.004	0.015	27
	0.768	0.657	0.045	0.653	0.660	-0.111	0.002	0.014	14
100	0.500	0.433	0.028	0.431	0.435	-0.067	0.001	0.005	13
	0.400	0.334	0.032	0.331	0.337	-0.066	0.001	0.005	17
	0.768	0.702	0.032	0.699	0.705	-0.066	0.001	0.005	9
250	0.500	0.469	0.017	0.468	0.471	-0.031	0.000	0.001	6
	0.400	0.368	0.016	0.366	0.369	-0.032	0.000	0.001	8
	0.768	0.735	0.023	0.733	0.738	-0.033	0.001	0.002	4
500	0.500	0.482	0.011	0.481	0.483	-0.018	0.000	0.000	4
	0.400	0.382	0.011	0.381	0.382	-0.018	0.000	0.000	5
	0.768	0.750	0.017	0.749	0.752	-0.018	0.000	0.001	2

Table 6.12: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with eight indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.369	0.048	0.365	0.373	-0.131	0.002	0.019	26
	0.400	0.276	0.058	0.270	0.283	-0.124	0.003	0.019	31
	0.768	0.638	0.043	0.634	0.641	-0.130	0.002	0.019	17
100	0.500	0.415	0.031	0.412	0.418	-0.085	0.001	0.008	17
	0.400	0.315	0.033	0.312	0.318	-0.085	0.001	0.008	21
	0.768	0.684	0.031	0.681	0.687	-0.084	0.001	0.008	11
250	0.500	0.459	0.017	0.458	0.461	-0.041	0.000	0.002	8
	0.400	0.357	0.019	0.355	0.359	-0.043	0.000	0.002	11
	0.768	0.727	0.022	0.725	0.730	-0.041	0.000	0.002	5
500	0.500	0.478	0.012	0.477	0.479	-0.022	0.000	0.001	4
	0.400	0.376	0.011	0.375	0.377	-0.024	0.000	0.001	6
	0.768	0.745	0.017	0.743	0.747	-0.023	0.000	0.001	3

Table 6.13: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with two indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
100	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
250	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
500	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25

Table 6.14: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with four indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.7	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.8	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.9	1.000	0.000	-	-	0.100	0.000	0.010	-11
100	0.6	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.7	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.8	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.9	1.000	0.000	-	-	0.100	0.000	0.010	-11
250	0.6	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.7	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.8	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.9	1.000	0.000	-	-	0.100	0.000	0.010	-11
500	0.6	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.7	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.8	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.9	1.000	0.000	-	-	0.100	0.000	0.010	-11

Table 6.15: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with six indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
100	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
250	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
500	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43

Table 6.16: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case A, models with eight indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
100	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
250	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
500	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11
	0.600	1.000	0.000	-	-	0.400	0.000	0.160	-67
	0.700	1.000	0.000	-	-	0.300	0.000	0.090	-43
	0.800	1.000	0.000	-	-	0.200	0.000	0.040	-25
	0.900	1.000	0.000	-	-	0.100	0.000	0.010	-11

Table 6.17: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with two indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.800	0.760	0.203	0.744	0.773	-0.040	0.041	0.043	5
		0.600	0.592	0.243	0.572	0.613	-0.008	0.059	0.059	1
	2	0.400	0.378	0.356	0.348	0.405	-0.022	0.127	0.127	5
		0.917	0.816	0.280	0.793	0.836	-0.101	0.078	0.088	11
	3	0.100	0.121	0.249	0.099	0.143	0.021	0.062	0.063	-21
		0.995	0.961	0.069	0.953	0.967	-0.034	0.005	0.006	3
100	1	0.800	0.774	0.144	0.759	0.790	-0.026	0.021	0.021	3
		0.600	0.600	0.174	0.586	0.616	0.000	0.030	0.030	0
	2	0.400	0.392	0.251	0.372	0.413	-0.008	0.063	0.063	2
		0.917	0.875	0.132	0.866	0.882	-0.041	0.017	0.019	5
	3	0.100	0.101	0.166	0.086	0.117	0.001	0.027	0.027	-1
		0.995	0.981	0.033	0.978	0.983	-0.014	0.001	0.001	1
250	1	0.800	0.795	0.088	0.786	0.802	-0.005	0.008	0.008	1
		0.600	0.592	0.114	0.584	0.603	-0.008	0.013	0.013	1
	2	0.400	0.395	0.161	0.381	0.410	-0.005	0.026	0.026	1
		0.917	0.900	0.078	0.893	0.908	-0.016	0.006	0.006	2
	3	0.100	0.095	0.105	0.086	0.104	-0.005	0.011	0.011	5
		0.995	0.989	0.018	0.987	0.991	-0.006	0.000	0.000	1
500	1	0.800	0.794	0.061	0.788	0.801	-0.006	0.004	0.004	1
		0.600	0.601	0.078	0.597	0.608	0.001	0.006	0.006	0
	2	0.400	0.390	0.116	0.380	0.401	-0.010	0.014	0.014	2
		0.917	0.912	0.054	0.907	0.917	-0.005	0.003	0.003	1
	3	0.100	0.100	0.075	0.095	0.106	0.000	0.006	0.006	0
		0.995	0.992	0.010	0.991	0.993	-0.003	0.000	0.000	0

Table 6.18: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with four indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.200	0.172	0.242	0.151	0.194	-0.028	0.059	0.059	14
		0.300	0.260	0.266	0.231	0.294	-0.040	0.071	0.072	13
		0.500	0.451	0.239	0.435	0.467	-0.049	0.057	0.059	10
		0.787	0.720	0.192	0.707	0.733	-0.068	0.037	0.041	9
	2	0.200	0.155	0.322	0.123	0.189	-0.045	0.104	0.106	22
		0.400	0.340	0.302	0.306	0.370	-0.060	0.091	0.095	15
		0.600	0.506	0.287	0.476	0.539	-0.094	0.082	0.091	16
		0.663	0.551	0.267	0.524	0.573	-0.112	0.072	0.084	17
	3	0.300	0.293	0.196	0.276	0.310	-0.007	0.039	0.039	2
		0.500	0.460	0.202	0.443	0.477	-0.040	0.041	0.042	8
		0.700	0.662	0.172	0.651	0.677	-0.038	0.030	0.031	5
		0.412	0.389	0.197	0.378	0.401	-0.023	0.039	0.039	6
100	1	0.200	0.182	0.183	0.169	0.197	-0.018	0.033	0.034	9
		0.300	0.295	0.185	0.277	0.312	-0.005	0.034	0.034	2
		0.500	0.485	0.170	0.472	0.498	-0.015	0.029	0.029	3
		0.787	0.731	0.137	0.716	0.747	-0.057	0.019	0.022	7
	2	0.200	0.185	0.241	0.169	0.200	-0.015	0.058	0.058	8
		0.400	0.364	0.226	0.338	0.392	-0.036	0.051	0.052	9
		0.600	0.543	0.204	0.523	0.566	-0.057	0.042	0.045	10
		0.663	0.614	0.186	0.595	0.634	-0.050	0.035	0.037	7
	3	0.300	0.276	0.140	0.259	0.291	-0.024	0.020	0.020	8
		0.500	0.495	0.133	0.485	0.506	-0.005	0.018	0.018	1
		0.700	0.687	0.115	0.677	0.700	-0.013	0.013	0.014	2
		0.412	0.389	0.140	0.377	0.399	-0.023	0.020	0.020	6
250	1	0.200	0.197	0.117	0.185	0.213	-0.003	0.014	0.014	1
		0.300	0.288	0.115	0.276	0.299	-0.012	0.013	0.013	4
		0.500	0.493	0.104	0.483	0.504	-0.007	0.011	0.011	1
		0.787	0.770	0.081	0.763	0.777	-0.017	0.007	0.007	2
	2	0.200	0.198	0.147	0.182	0.213	-0.002	0.022	0.022	1
		0.400	0.392	0.141	0.378	0.404	-0.008	0.020	0.020	2
		0.600	0.579	0.125	0.567	0.591	-0.021	0.016	0.016	4
		0.663	0.638	0.124	0.629	0.648	-0.025	0.015	0.016	4
	3	0.300	0.299	0.087	0.294	0.305	-0.001	0.008	0.008	0
		0.500	0.492	0.082	0.488	0.499	-0.008	0.007	0.007	2
		0.700	0.694	0.071	0.686	0.702	-0.006	0.005	0.005	1
		0.412	0.403	0.090	0.397	0.409	-0.009	0.008	0.008	2
500	1	0.200	0.203	0.085	0.196	0.213	0.003	0.007	0.007	-1
		0.300	0.291	0.083	0.284	0.297	-0.009	0.007	0.007	3
		0.500	0.499	0.073	0.493	0.506	-0.001	0.005	0.005	0
		0.787	0.776	0.055	0.771	0.780	-0.012	0.003	0.003	1
	2	0.200	0.194	0.110	0.185	0.202	-0.006	0.012	0.012	3
		0.400	0.404	0.102	0.397	0.414	0.004	0.010	0.010	-1
		0.600	0.587	0.093	0.580	0.594	-0.013	0.009	0.009	2
		0.663	0.648	0.086	0.641	0.655	-0.015	0.007	0.008	2
	3	0.300	0.302	0.064	0.296	0.308	0.002	0.004	0.004	-1
		0.500	0.494	0.056	0.489	0.499	-0.006	0.003	0.003	1
		0.700	0.700	0.050	0.694	0.705	0.000	0.003	0.003	0
		0.412	0.408	0.062	0.402	0.414	-0.004	0.004	0.004	1

Table 6.19: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with six indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.500	0.435	0.213	0.416	0.450	-0.065	0.045	0.050	13
		0.300	0.243	0.240	0.220	0.269	-0.057	0.057	0.061	19
		0.400	0.317	0.250	0.294	0.337	-0.083	0.062	0.069	21
		0.300	0.255	0.252	0.231	0.283	-0.045	0.064	0.066	15
		0.600	0.521	0.224	0.504	0.538	-0.079	0.050	0.056	13
		0.224	0.188	0.245	0.161	0.211	-0.035	0.060	0.061	16
	2	0.200	0.171	0.285	0.151	0.198	-0.029	0.081	0.082	14
		0.400	0.316	0.287	0.291	0.339	-0.084	0.083	0.090	21
		0.600	0.445	0.280	0.411	0.479	-0.155	0.079	0.103	26
		0.400	0.298	0.286	0.276	0.321	-0.102	0.082	0.092	26
		0.200	0.158	0.301	0.140	0.176	-0.042	0.091	0.092	21
		0.490	0.371	0.292	0.341	0.397	-0.119	0.085	0.100	24
	3	0.300	0.270	0.191	0.246	0.291	-0.030	0.037	0.037	10
		0.700	0.648	0.165	0.631	0.668	-0.052	0.027	0.030	7
		0.200	0.162	0.201	0.146	0.183	-0.038	0.040	0.042	19
		0.300	0.264	0.200	0.244	0.281	-0.036	0.040	0.041	12
		0.400	0.355	0.189	0.337	0.380	-0.045	0.036	0.038	11
		0.361	0.319	0.206	0.296	0.339	-0.042	0.043	0.044	12
100	1	0.500	0.459	0.160	0.447	0.473	-0.041	0.026	0.027	8
		0.300	0.278	0.184	0.264	0.289	-0.022	0.034	0.034	7
		0.400	0.368	0.176	0.351	0.388	-0.032	0.031	0.032	8
		0.300	0.282	0.174	0.265	0.297	-0.018	0.030	0.030	6
		0.600	0.549	0.155	0.531	0.566	-0.051	0.024	0.027	8
		0.224	0.216	0.175	0.201	0.231	-0.007	0.030	0.031	3
	2	0.200	0.166	0.224	0.149	0.185	-0.034	0.050	0.051	17
		0.400	0.350	0.218	0.333	0.365	-0.050	0.047	0.050	12
		0.600	0.531	0.191	0.512	0.550	-0.069	0.037	0.041	12
		0.400	0.353	0.211	0.334	0.368	-0.047	0.044	0.047	12
		0.200	0.196	0.217	0.176	0.219	-0.004	0.047	0.047	2
		0.490	0.420	0.198	0.399	0.441	-0.070	0.039	0.044	14
	3	0.300	0.283	0.128	0.273	0.295	-0.017	0.016	0.017	6
		0.700	0.664	0.112	0.655	0.675	-0.036	0.012	0.014	5
		0.200	0.181	0.141	0.168	0.198	-0.019	0.020	0.020	10
		0.300	0.293	0.138	0.281	0.306	-0.007	0.019	0.019	2
		0.400	0.373	0.133	0.363	0.386	-0.027	0.018	0.018	7
		0.361	0.358	0.136	0.345	0.373	-0.003	0.018	0.018	1

Table 6.20: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with six indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.500	0.477	0.103	0.469	0.487	-0.023	0.011	0.011	5
		0.300	0.298	0.109	0.288	0.309	-0.002	0.012	0.012	1
		0.400	0.387	0.113	0.375	0.401	-0.013	0.013	0.013	3
		0.300	0.287	0.113	0.277	0.298	-0.013	0.013	0.013	4
		0.600	0.586	0.095	0.578	0.597	-0.014	0.009	0.009	2
		0.224	0.216	0.115	0.203	0.228	-0.007	0.013	0.013	3
	2	0.200	0.194	0.151	0.179	0.210	-0.006	0.023	0.023	3
		0.400	0.383	0.145	0.369	0.397	-0.017	0.021	0.021	4
		0.600	0.555	0.127	0.544	0.569	-0.045	0.016	0.018	7
		0.400	0.383	0.139	0.369	0.399	-0.017	0.019	0.019	4
		0.200	0.191	0.148	0.180	0.208	-0.009	0.022	0.022	4
		0.490	0.459	0.138	0.446	0.473	-0.031	0.019	0.020	6
	3	0.300	0.296	0.087	0.289	0.305	-0.004	0.008	0.008	1
		0.700	0.692	0.071	0.683	0.700	-0.008	0.005	0.005	1
		0.200	0.188	0.090	0.182	0.194	-0.012	0.008	0.008	6
		0.300	0.290	0.084	0.283	0.297	-0.010	0.007	0.007	3
		0.400	0.390	0.090	0.381	0.399	-0.010	0.008	0.008	3
		0.361	0.354	0.087	0.345	0.363	-0.007	0.008	0.008	2
500	1	0.500	0.493	0.071	0.486	0.500	-0.007	0.005	0.005	1
		0.300	0.293	0.077	0.287	0.301	-0.007	0.006	0.006	2
		0.400	0.398	0.077	0.391	0.406	-0.002	0.006	0.006	0
		0.300	0.291	0.080	0.281	0.299	-0.009	0.006	0.006	3
		0.600	0.595	0.069	0.591	0.600	-0.005	0.005	0.005	1
		0.224	0.209	0.078	0.204	0.215	-0.015	0.006	0.006	7
	2	0.200	0.191	0.105	0.184	0.198	-0.009	0.011	0.011	4
		0.400	0.389	0.098	0.378	0.402	-0.011	0.010	0.010	3
		0.600	0.587	0.085	0.580	0.595	-0.013	0.007	0.007	2
		0.400	0.385	0.096	0.375	0.393	-0.015	0.009	0.009	4
		0.200	0.189	0.105	0.181	0.195	-0.011	0.011	0.011	6
		0.490	0.477	0.092	0.467	0.487	-0.013	0.009	0.009	3
	3	0.300	0.300	0.063	0.295	0.306	0.000	0.004	0.004	0
		0.700	0.696	0.047	0.692	0.701	-0.004	0.002	0.002	1
		0.200	0.196	0.068	0.189	0.205	-0.004	0.005	0.005	2
		0.300	0.301	0.062	0.295	0.307	0.001	0.004	0.004	0
		0.400	0.391	0.062	0.385	0.395	-0.009	0.004	0.004	2
		0.361	0.359	0.061	0.355	0.364	-0.002	0.004	0.004	0

Table 6.21: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with eight indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.200	0.178	0.232	0.155	0.204	-0.022	0.054	0.054	11
		0.300	0.245	0.227	0.227	0.267	-0.055	0.051	0.054	18
		0.400	0.340	0.223	0.313	0.368	-0.060	0.050	0.054	15
		0.500	0.413	0.222	0.397	0.435	-0.087	0.049	0.057	17
		0.400	0.325	0.225	0.310	0.341	-0.075	0.051	0.056	19
		0.300	0.254	0.234	0.235	0.274	-0.046	0.055	0.057	15
		0.200	0.158	0.223	0.142	0.171	-0.042	0.050	0.051	21
		0.412	0.327	0.222	0.313	0.341	-0.085	0.049	0.057	21
	2	0.300	0.211	0.273	0.188	0.240	-0.089	0.075	0.082	30
		0.300	0.219	0.283	0.197	0.242	-0.081	0.080	0.087	27
		0.400	0.268	0.268	0.237	0.304	-0.132	0.072	0.089	33
		0.400	0.286	0.269	0.254	0.312	-0.114	0.072	0.085	28
		0.200	0.151	0.271	0.134	0.171	-0.049	0.074	0.076	24
		0.300	0.228	0.267	0.196	0.254	-0.072	0.071	0.076	24
		0.500	0.379	0.256	0.359	0.397	-0.121	0.065	0.080	24
		0.346	0.234	0.274	0.202	0.262	-0.112	0.075	0.088	32
	3	0.400	0.358	0.191	0.343	0.378	-0.042	0.036	0.038	11
		0.500	0.420	0.182	0.400	0.439	-0.080	0.033	0.040	16
		0.500	0.453	0.192	0.440	0.467	-0.047	0.037	0.039	9
		0.300	0.262	0.203	0.239	0.286	-0.038	0.041	0.043	13
		0.200	0.187	0.198	0.174	0.203	-0.013	0.039	0.039	6
		0.100	0.087	0.197	0.065	0.107	-0.013	0.039	0.039	13
		0.300	0.258	0.189	0.244	0.274	-0.042	0.036	0.037	14
		0.332	0.295	0.198	0.277	0.316	-0.037	0.039	0.041	11
100	1	0.200	0.183	0.178	0.167	0.201	-0.017	0.032	0.032	9
		0.300	0.259	0.171	0.245	0.274	-0.041	0.029	0.031	14
		0.400	0.378	0.160	0.364	0.389	-0.022	0.026	0.026	5
		0.500	0.440	0.157	0.423	0.458	-0.060	0.025	0.028	12
		0.400	0.355	0.159	0.344	0.367	-0.045	0.025	0.027	11
		0.300	0.269	0.173	0.253	0.284	-0.031	0.030	0.031	10
		0.200	0.190	0.171	0.174	0.209	-0.010	0.029	0.029	5
		0.412	0.367	0.161	0.353	0.383	-0.045	0.026	0.028	11
	2	0.300	0.235	0.210	0.214	0.255	-0.065	0.044	0.048	22
		0.300	0.253	0.203	0.237	0.270	-0.047	0.041	0.043	16
		0.400	0.339	0.197	0.319	0.359	-0.061	0.039	0.043	15
		0.400	0.345	0.199	0.329	0.359	-0.055	0.040	0.043	14
		0.200	0.164	0.201	0.143	0.182	-0.036	0.040	0.042	18
		0.300	0.261	0.208	0.236	0.281	-0.039	0.043	0.045	13
		0.500	0.425	0.194	0.409	0.441	-0.075	0.038	0.043	15
		0.346	0.289	0.213	0.264	0.306	-0.057	0.045	0.049	17
	3	0.400	0.377	0.138	0.364	0.391	-0.023	0.019	0.020	6
		0.500	0.465	0.130	0.456	0.473	-0.035	0.017	0.018	7
		0.500	0.470	0.130	0.458	0.482	-0.030	0.017	0.018	6
		0.300	0.284	0.143	0.272	0.296	-0.016	0.020	0.021	5
		0.200	0.186	0.140	0.175	0.197	-0.014	0.020	0.020	7
		0.100	0.092	0.136	0.079	0.106	-0.008	0.018	0.019	8
		0.300	0.278	0.138	0.265	0.292	-0.022	0.019	0.019	7
		0.332	0.315	0.142	0.300	0.331	-0.016	0.020	0.020	5

Table 6.22: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with eight indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.200	0.189	0.117	0.176	0.198	-0.011	0.014	0.014	6
		0.300	0.289	0.108	0.276	0.305	-0.011	0.012	0.012	4
		0.400	0.379	0.103	0.370	0.389	-0.021	0.011	0.011	5
		0.500	0.482	0.103	0.474	0.492	-0.018	0.011	0.011	4
		0.400	0.388	0.104	0.380	0.396	-0.012	0.011	0.011	3
		0.300	0.289	0.107	0.275	0.304	-0.011	0.012	0.012	4
		0.200	0.185	0.109	0.175	0.194	-0.015	0.012	0.012	7
		0.412	0.394	0.105	0.384	0.405	-0.019	0.011	0.011	4
	2	0.300	0.285	0.142	0.273	0.299	-0.015	0.020	0.020	5
		0.300	0.279	0.139	0.265	0.292	-0.021	0.019	0.020	7
		0.400	0.363	0.127	0.355	0.373	-0.037	0.016	0.018	9
		0.400	0.370	0.133	0.358	0.383	-0.030	0.018	0.019	7
		0.200	0.188	0.141	0.171	0.207	-0.012	0.020	0.020	6
		0.300	0.269	0.134	0.254	0.285	-0.031	0.018	0.019	10
		0.500	0.470	0.130	0.455	0.485	-0.030	0.017	0.018	6
		0.346	0.323	0.142	0.308	0.335	-0.023	0.020	0.021	7
	3	0.400	0.393	0.086	0.384	0.402	-0.007	0.007	0.008	2
		0.500	0.488	0.085	0.478	0.497	-0.012	0.007	0.007	2
		0.500	0.484	0.083	0.472	0.493	-0.016	0.007	0.007	3
		0.300	0.293	0.086	0.285	0.300	-0.007	0.007	0.007	2
		0.200	0.195	0.090	0.187	0.205	-0.005	0.008	0.008	2
		0.100	0.095	0.092	0.088	0.104	-0.005	0.008	0.008	5
		0.300	0.291	0.085	0.282	0.299	-0.009	0.007	0.007	3
		0.332	0.321	0.080	0.314	0.328	-0.011	0.006	0.007	3
500	1	0.200	0.196	0.079	0.187	0.206	-0.004	0.006	0.006	2
		0.300	0.296	0.077	0.289	0.303	-0.004	0.006	0.006	1
		0.400	0.390	0.075	0.382	0.398	-0.010	0.006	0.006	2
		0.500	0.488	0.069	0.481	0.494	-0.012	0.005	0.005	2
		0.400	0.384	0.076	0.376	0.392	-0.016	0.006	0.006	4
		0.300	0.294	0.078	0.288	0.301	-0.006	0.006	0.006	2
		0.200	0.206	0.082	0.197	0.214	0.006	0.007	0.007	-3
		0.412	0.408	0.074	0.402	0.414	-0.005	0.005	0.005	1
	2	0.300	0.295	0.104	0.288	0.303	-0.005	0.011	0.011	2
		0.300	0.293	0.101	0.285	0.304	-0.007	0.010	0.010	2
		0.400	0.385	0.094	0.376	0.395	-0.015	0.009	0.009	4
		0.400	0.377	0.096	0.368	0.384	-0.023	0.009	0.010	6
		0.200	0.193	0.107	0.181	0.205	-0.007	0.012	0.012	4
		0.300	0.289	0.104	0.279	0.299	-0.011	0.011	0.011	4
		0.500	0.482	0.093	0.474	0.490	-0.018	0.009	0.009	4
		0.346	0.332	0.097	0.324	0.341	-0.014	0.009	0.010	4
	3	0.400	0.399	0.058	0.394	0.405	-0.001	0.003	0.003	0
		0.500	0.494	0.055	0.490	0.499	-0.006	0.003	0.003	1
		0.500	0.493	0.056	0.487	0.498	-0.007	0.003	0.003	1
		0.300	0.298	0.058	0.292	0.304	-0.002	0.003	0.003	1
		0.200	0.202	0.062	0.196	0.209	0.002	0.004	0.004	-1
		0.100	0.097	0.062	0.092	0.100	-0.003	0.004	0.004	3
		0.300	0.293	0.060	0.288	0.298	-0.007	0.004	0.004	2
		0.332	0.328	0.062	0.323	0.333	-0.003	0.004	0.004	1

Table 6.23: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with two indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.414	0.091	0.409	0.420	-0.086	0.008	0.016	17
	0.400	0.321	0.111	0.310	0.333	-0.079	0.012	0.019	20
	0.600	0.501	0.084	0.492	0.509	-0.099	0.007	0.017	17
100	0.500	0.419	0.065	0.414	0.427	-0.081	0.004	0.011	16
	0.400	0.338	0.070	0.330	0.345	-0.062	0.005	0.009	16
	0.600	0.508	0.059	0.502	0.513	-0.092	0.003	0.012	15
250	0.500	0.425	0.038	0.422	0.429	-0.075	0.001	0.007	15
	0.400	0.338	0.042	0.335	0.342	-0.062	0.002	0.006	15
	0.600	0.508	0.037	0.506	0.509	-0.092	0.001	0.010	15
500	0.500	0.423	0.028	0.421	0.426	-0.077	0.001	0.007	15
	0.400	0.341	0.028	0.339	0.344	-0.059	0.001	0.004	15
	0.600	0.511	0.026	0.508	0.513	-0.089	0.001	0.009	15

Table 6.24: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with four indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.421	0.086	0.414	0.429	-0.079	0.007	0.014	16
	0.400	0.328	0.091	0.318	0.337	-0.072	0.008	0.013	18
	0.600	0.512	0.076	0.505	0.519	-0.088	0.006	0.014	15
100	0.500	0.435	0.055	0.431	0.442	-0.065	0.003	0.007	13
	0.400	0.344	0.058	0.338	0.351	-0.056	0.003	0.006	14
	0.600	0.535	0.054	0.531	0.539	-0.065	0.003	0.007	11
250	0.500	0.453	0.035	0.450	0.457	-0.047	0.001	0.003	9
	0.400	0.364	0.037	0.360	0.366	-0.036	0.001	0.003	9
	0.600	0.542	0.032	0.540	0.545	-0.058	0.001	0.004	10
500	0.500	0.460	0.024	0.458	0.462	-0.040	0.001	0.002	8
	0.400	0.363	0.025	0.360	0.367	-0.037	0.001	0.002	9
	0.600	0.550	0.023	0.548	0.552	-0.050	0.001	0.003	8

Table 6.25: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with six indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.411	0.079	0.404	0.419	-0.089	0.006	0.014	18
	0.400	0.323	0.094	0.309	0.336	-0.077	0.009	0.015	19
	0.600	0.491	0.074	0.483	0.497	-0.109	0.005	0.017	18
100	0.500	0.429	0.054	0.425	0.434	-0.071	0.003	0.008	14
	0.400	0.342	0.058	0.338	0.346	-0.058	0.003	0.007	15
	0.600	0.524	0.053	0.518	0.530	-0.076	0.003	0.009	13
250	0.500	0.452	0.033	0.450	0.455	-0.048	0.001	0.003	10
	0.400	0.360	0.035	0.356	0.365	-0.040	0.001	0.003	10
	0.600	0.542	0.030	0.540	0.545	-0.058	0.001	0.004	10
500	0.500	0.459	0.023	0.456	0.461	-0.041	0.001	0.002	8
	0.400	0.368	0.023	0.366	0.371	-0.032	0.001	0.002	8
	0.600	0.552	0.022	0.550	0.554	-0.048	0.000	0.003	8

Table 6.26: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with eight indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.405	0.080	0.399	0.411	-0.095	0.006	0.015	19
	0.400	0.315	0.089	0.308	0.321	-0.085	0.008	0.015	21
	0.600	0.482	0.074	0.476	0.489	-0.118	0.005	0.019	20
100	0.500	0.425	0.049	0.420	0.430	-0.075	0.002	0.008	15
	0.400	0.339	0.058	0.333	0.344	-0.061	0.003	0.007	15
	0.600	0.521	0.049	0.515	0.525	-0.079	0.002	0.009	13
250	0.500	0.453	0.032	0.450	0.455	-0.047	0.001	0.003	9
	0.400	0.358	0.036	0.354	0.361	-0.042	0.001	0.003	11
	0.600	0.548	0.032	0.546	0.552	-0.052	0.001	0.004	9
500	0.500	0.466	0.024	0.464	0.469	-0.034	0.001	0.002	7
	0.400	0.370	0.024	0.368	0.372	-0.030	0.001	0.001	7
	0.600	0.563	0.020	0.561	0.565	-0.037	0.000	0.002	6

Table 6.27: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with two indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.7	0.863	0.041	0.859	0.867	0.163	0.002	0.028	-23
	0.8	0.897	0.026	0.895	0.899	0.097	0.001	0.010	-12
100	0.7	0.866	0.027	0.864	0.869	0.166	0.001	0.028	-24
	0.8	0.900	0.016	0.898	0.901	0.100	0.000	0.010	-12
250	0.7	0.866	0.018	0.864	0.868	0.166	0.000	0.028	-24
	0.8	0.899	0.010	0.899	0.900	0.099	0.000	0.010	-12
500	0.7	0.867	0.011	0.866	0.868	0.167	0.000	0.028	-24
	0.8	0.899	0.007	0.898	0.900	0.099	0.000	0.010	-12

Table 6.28: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with four indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.712	0.080	0.702	0.722	0.112	0.006	0.019	-19
	0.7	0.793	0.055	0.788	0.799	0.093	0.003	0.012	-13
	0.8	0.858	0.035	0.856	0.860	0.058	0.001	0.005	-7
	0.9	0.903	0.021	0.900	0.905	0.003	0.000	0.000	0
100	0.6	0.715	0.057	0.710	0.721	0.115	0.003	0.016	-19
	0.7	0.793	0.038	0.790	0.796	0.093	0.001	0.010	-13
	0.8	0.856	0.024	0.854	0.859	0.056	0.001	0.004	-7
	0.9	0.903	0.014	0.902	0.904	0.003	0.000	0.000	0
250	0.6	0.708	0.035	0.705	0.711	0.108	0.001	0.013	-18
	0.7	0.792	0.024	0.790	0.794	0.092	0.001	0.009	-13
	0.8	0.860	0.015	0.859	0.861	0.060	0.000	0.004	-7
	0.9	0.903	0.009	0.902	0.904	0.003	0.000	0.000	0
500	0.6	0.710	0.025	0.708	0.712	0.110	0.001	0.013	-18
	0.7	0.795	0.017	0.793	0.796	0.095	0.000	0.009	-14
	0.8	0.858	0.011	0.857	0.859	0.058	0.000	0.003	-7
	0.9	0.903	0.006	0.903	0.904	0.003	0.000	0.000	0

Table 6.29: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with six indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.677	0.086	0.670	0.685	0.077	0.007	0.013	-13
	0.7	0.764	0.059	0.757	0.772	0.064	0.003	0.008	-9
	0.8	0.835	0.040	0.832	0.838	0.035	0.002	0.003	-4
	0.9	0.895	0.022	0.893	0.897	-0.005	0.001	0.001	1
	0.6	0.676	0.084	0.665	0.684	0.076	0.007	0.013	-13
	0.7	0.763	0.059	0.759	0.767	0.063	0.003	0.007	-9
100	0.6	0.681	0.055	0.676	0.685	0.081	0.003	0.010	-13
	0.7	0.763	0.043	0.759	0.766	0.063	0.002	0.006	-9
	0.8	0.837	0.027	0.835	0.840	0.037	0.001	0.002	-5
	0.9	0.895	0.016	0.893	0.897	-0.005	0.000	0.000	1
	0.6	0.675	0.062	0.669	0.682	0.075	0.004	0.009	-13
	0.7	0.761	0.039	0.757	0.765	0.061	0.002	0.005	-9
250	0.6	0.678	0.037	0.675	0.682	0.078	0.001	0.008	-13
	0.7	0.764	0.025	0.762	0.766	0.064	0.001	0.005	-9
	0.8	0.837	0.017	0.836	0.838	0.037	0.000	0.002	-5
	0.9	0.895	0.010	0.894	0.896	-0.005	0.000	0.000	1
	0.6	0.675	0.036	0.672	0.678	0.075	0.001	0.007	-13
	0.7	0.762	0.025	0.759	0.764	0.062	0.001	0.004	-9
500	0.6	0.678	0.027	0.675	0.681	0.078	0.001	0.007	-13
	0.7	0.763	0.017	0.761	0.765	0.063	0.000	0.004	-9
	0.8	0.837	0.012	0.836	0.838	0.037	0.000	0.001	-5
	0.9	0.895	0.007	0.895	0.896	-0.005	0.000	0.000	1
	0.6	0.679	0.025	0.677	0.681	0.079	0.001	0.007	-13
	0.7	0.762	0.019	0.761	0.764	0.062	0.000	0.004	-9

Table 6.30: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case B, models with eight indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.650	0.080	0.644	0.657	0.050	0.006	0.009	-8
	0.7	0.745	0.062	0.740	0.750	0.045	0.004	0.006	-6
	0.8	0.828	0.040	0.824	0.831	0.028	0.002	0.002	-4
	0.9	0.900	0.022	0.899	0.903	0.000	0.000	0.000	0
	0.6	0.664	0.079	0.658	0.670	0.064	0.006	0.010	-11
	0.7	0.753	0.063	0.747	0.759	0.053	0.004	0.007	-8
	0.8	0.831	0.038	0.828	0.834	0.031	0.001	0.002	-4
	0.9	0.904	0.020	0.903	0.906	0.004	0.000	0.000	0
100	0.6	0.653	0.059	0.648	0.660	0.053	0.003	0.006	-9
	0.7	0.745	0.044	0.740	0.750	0.045	0.002	0.004	-6
	0.8	0.829	0.026	0.826	0.831	0.029	0.001	0.002	-4
	0.9	0.900	0.016	0.898	0.902	0.000	0.000	0.000	0
	0.6	0.656	0.059	0.649	0.661	0.056	0.004	0.007	-9
	0.7	0.749	0.041	0.746	0.752	0.049	0.002	0.004	-7
	0.8	0.830	0.026	0.828	0.832	0.030	0.001	0.002	-4
	0.9	0.900	0.016	0.898	0.902	0.000	0.000	0.000	0
250	0.6	0.658	0.035	0.655	0.662	0.058	0.001	0.005	-10
	0.7	0.747	0.027	0.745	0.750	0.047	0.001	0.003	-7
	0.8	0.830	0.017	0.828	0.831	0.030	0.000	0.001	-4
	0.9	0.901	0.009	0.900	0.902	0.001	0.000	0.000	0
	0.6	0.657	0.037	0.654	0.660	0.057	0.001	0.005	-10
	0.7	0.748	0.027	0.746	0.750	0.048	0.001	0.003	-7
	0.8	0.828	0.018	0.827	0.830	0.028	0.000	0.001	-4
	0.9	0.901	0.010	0.901	0.902	0.001	0.000	0.000	0
500	0.6	0.657	0.026	0.655	0.659	0.057	0.001	0.004	-9
	0.7	0.748	0.018	0.747	0.750	0.048	0.000	0.003	-7
	0.8	0.830	0.012	0.829	0.831	0.030	0.000	0.001	-4
	0.9	0.901	0.007	0.900	0.902	0.001	0.000	0.000	0
	0.6	0.657	0.025	0.654	0.659	0.057	0.001	0.004	-9
	0.7	0.747	0.018	0.745	0.748	0.047	0.000	0.003	-7
	0.8	0.830	0.012	0.829	0.830	0.030	0.000	0.001	-4
	0.9	0.901	0.006	0.900	0.902	0.001	0.000	0.000	0

Table 6.31: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with two indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.8	0.812	0.181	0.790	0.835	0.012	0.033	0.033	-2
		0.5	0.511	0.264	0.491	0.529	0.011	0.070	0.070	-2
	2	0.4	0.445	0.348	0.425	0.475	0.045	0.121	0.123	-11
		0.8	0.798	0.252	0.771	0.818	-0.002	0.063	0.063	0
	3	0.1	0.103	0.297	0.074	0.138	0.003	0.088	0.088	-3
		0.9	0.946	0.103	0.936	0.952	0.046	0.011	0.013	-5
100	1	0.8	0.827	0.129	0.815	0.842	0.027	0.017	0.017	-3
		0.5	0.521	0.192	0.505	0.535	0.021	0.037	0.037	-4
	2	0.4	0.449	0.260	0.421	0.479	0.049	0.068	0.070	-12
		0.8	0.837	0.178	0.821	0.851	0.037	0.032	0.033	-5
	3	0.1	0.106	0.173	0.090	0.126	0.006	0.030	0.030	-6
		0.9	0.980	0.035	0.975	0.983	0.080	0.001	0.008	-9
250	1	0.8	0.845	0.078	0.839	0.851	0.045	0.006	0.008	-6
		0.5	0.518	0.116	0.509	0.529	0.018	0.013	0.014	-4
	2	0.4	0.444	0.156	0.433	0.458	0.044	0.024	0.026	-11
		0.8	0.879	0.086	0.869	0.888	0.079	0.007	0.014	-10
	3	0.1	0.111	0.103	0.099	0.122	0.011	0.011	0.011	-11
		0.9	0.989	0.016	0.987	0.990	0.089	0.000	0.008	-10
500	1	0.8	0.848	0.053	0.844	0.851	0.048	0.003	0.005	-6
		0.5	0.525	0.081	0.515	0.534	0.025	0.007	0.007	-5
	2	0.4	0.434	0.115	0.423	0.449	0.034	0.013	0.014	-9
		0.8	0.892	0.057	0.886	0.897	0.092	0.003	0.012	-12
	3	0.1	0.110	0.077	0.105	0.116	0.010	0.006	0.006	-10
		0.9	0.991	0.011	0.990	0.992	0.091	0.000	0.008	-10

Table 6.32: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with four indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.2	0.190	0.260	0.165	0.217	-0.010	0.068	0.068	5
		0.3	0.301	0.246	0.276	0.323	0.001	0.061	0.061	0
		0.5	0.480	0.232	0.447	0.512	-0.020	0.054	0.054	4
		0.7	0.668	0.208	0.650	0.691	-0.032	0.043	0.045	5
	2	0.2	0.187	0.313	0.156	0.216	-0.013	0.098	0.098	7
		0.4	0.393	0.294	0.371	0.416	-0.007	0.086	0.086	2
		0.6	0.533	0.295	0.509	0.561	-0.067	0.087	0.091	11
		0.5	0.458	0.293	0.433	0.485	-0.042	0.086	0.088	8
	3	0.3	0.286	0.265	0.267	0.309	-0.014	0.070	0.070	5
		0.5	0.495	0.230	0.479	0.512	-0.005	0.053	0.053	1
		0.7	0.666	0.207	0.648	0.682	-0.034	0.043	0.044	5
		0.2	0.184	0.263	0.156	0.214	-0.016	0.069	0.069	8
100	1	0.2	0.231	0.198	0.216	0.244	0.031	0.039	0.040	-15
		0.3	0.303	0.187	0.282	0.321	0.003	0.035	0.035	-1
		0.5	0.511	0.160	0.494	0.528	0.011	0.026	0.026	-2
		0.7	0.704	0.134	0.691	0.715	0.004	0.018	0.018	-1
	2	0.2	0.207	0.228	0.185	0.227	0.007	0.052	0.052	-4
		0.4	0.416	0.217	0.394	0.438	0.016	0.047	0.047	-4
		0.6	0.610	0.187	0.590	0.634	0.010	0.035	0.035	-2
		0.5	0.508	0.199	0.485	0.528	0.008	0.040	0.040	-2
	3	0.3	0.309	0.142	0.294	0.326	0.009	0.020	0.020	-3
		0.5	0.517	0.129	0.510	0.525	0.017	0.017	0.017	-3
		0.7	0.727	0.109	0.716	0.738	0.027	0.012	0.013	-4
		0.2	0.208	0.151	0.191	0.222	0.008	0.023	0.023	-4
250	1	0.2	0.218	0.114	0.209	0.226	0.018	0.013	0.013	-9
		0.3	0.314	0.109	0.304	0.324	0.014	0.012	0.012	-5
		0.5	0.528	0.105	0.519	0.538	0.028	0.011	0.012	-6
		0.7	0.738	0.086	0.732	0.745	0.038	0.007	0.009	-5
	2	0.2	0.218	0.152	0.200	0.233	0.018	0.023	0.023	-9
		0.4	0.446	0.140	0.433	0.461	0.046	0.020	0.022	-11
		0.6	0.635	0.121	0.622	0.650	0.035	0.015	0.016	-6
		0.5	0.536	0.137	0.523	0.550	0.036	0.019	0.020	-7
	3	0.3	0.312	0.093	0.303	0.322	0.012	0.009	0.009	-4
		0.5	0.530	0.081	0.523	0.538	0.030	0.007	0.008	-6
		0.7	0.745	0.068	0.740	0.751	0.045	0.005	0.007	-6
		0.2	0.212	0.098	0.204	0.219	0.012	0.010	0.010	-6
500	1	0.2	0.216	0.083	0.209	0.222	0.016	0.007	0.007	-8
		0.3	0.320	0.084	0.312	0.329	0.020	0.007	0.007	-7
		0.5	0.533	0.073	0.526	0.539	0.033	0.005	0.006	-7
		0.7	0.741	0.061	0.734	0.749	0.041	0.004	0.005	-6
	2	0.2	0.223	0.111	0.214	0.234	0.023	0.012	0.013	-11
		0.4	0.439	0.099	0.430	0.446	0.039	0.010	0.011	-10
		0.6	0.653	0.083	0.645	0.659	0.053	0.007	0.010	-9
		0.5	0.549	0.090	0.541	0.557	0.049	0.008	0.011	-10
	3	0.3	0.324	0.067	0.316	0.330	0.024	0.005	0.005	-8
		0.5	0.534	0.059	0.529	0.540	0.034	0.004	0.005	-7
		0.7	0.744	0.047	0.739	0.748	0.044	0.002	0.004	-6
		0.2	0.211	0.062	0.206	0.217	0.011	0.004	0.004	-6

Table 6.33: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with six indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.500	0.461	0.227	0.438	0.483	-0.039	0.052	0.053	8
		0.300	0.297	0.258	0.275	0.318	-0.003	0.067	0.067	1
		0.400	0.366	0.219	0.345	0.389	-0.034	0.048	0.049	8
		0.300	0.293	0.231	0.269	0.315	-0.007	0.054	0.054	2
		0.500	0.461	0.235	0.438	0.486	-0.039	0.055	0.057	8
		0.100	0.088	0.238	0.064	0.108	-0.012	0.056	0.057	12
	2	0.200	0.158	0.283	0.138	0.186	-0.042	0.080	0.082	21
		0.400	0.331	0.295	0.304	0.362	-0.069	0.087	0.092	17
		0.600	0.504	0.265	0.480	0.530	-0.096	0.070	0.080	16
		0.400	0.331	0.267	0.309	0.352	-0.069	0.071	0.076	17
		0.200	0.159	0.301	0.133	0.182	-0.041	0.091	0.092	20
		0.300	0.238	0.277	0.216	0.261	-0.062	0.077	0.081	21
	3	0.300	0.268	0.237	0.244	0.290	-0.032	0.056	0.057	11
		0.600	0.582	0.218	0.563	0.598	-0.018	0.048	0.048	3
		0.200	0.174	0.248	0.152	0.202	-0.026	0.062	0.062	13
		0.300	0.288	0.234	0.270	0.316	-0.012	0.055	0.055	4
		0.400	0.390	0.237	0.372	0.407	-0.010	0.056	0.056	3
		0.200	0.179	0.253	0.157	0.206	-0.021	0.064	0.065	10
100	1	0.500	0.498	0.155	0.484	0.512	-0.002	0.024	0.024	0
		0.300	0.302	0.173	0.283	0.321	0.002	0.030	0.030	-1
		0.400	0.396	0.170	0.379	0.413	-0.004	0.029	0.029	1
		0.300	0.316	0.177	0.300	0.337	0.016	0.031	0.032	-5
		0.500	0.505	0.160	0.491	0.525	0.005	0.025	0.025	-1
		0.100	0.100	0.178	0.087	0.117	0.000	0.032	0.032	0
	2	0.200	0.185	0.224	0.159	0.208	-0.015	0.050	0.050	7
		0.400	0.359	0.218	0.338	0.379	-0.041	0.048	0.049	10
		0.600	0.558	0.196	0.543	0.576	-0.042	0.038	0.040	7
		0.400	0.380	0.222	0.363	0.395	-0.020	0.049	0.050	5
		0.200	0.179	0.229	0.154	0.203	-0.021	0.053	0.053	11
		0.300	0.296	0.206	0.278	0.319	-0.004	0.042	0.042	1
	3	0.300	0.336	0.143	0.325	0.347	0.036	0.020	0.022	-12
		0.600	0.642	0.119	0.628	0.654	0.042	0.014	0.016	-7
		0.200	0.217	0.147	0.204	0.230	0.017	0.022	0.022	-9
		0.300	0.329	0.144	0.315	0.342	0.029	0.021	0.021	-10
		0.400	0.426	0.135	0.413	0.439	0.026	0.018	0.019	-7
		0.200	0.209	0.148	0.197	0.221	0.009	0.022	0.022	-4

Table 6.34: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with six indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.500	0.523	0.105	0.513	0.533	0.023	0.011	0.012	-5
		0.300	0.325	0.113	0.315	0.338	0.025	0.013	0.013	-8
		0.400	0.416	0.110	0.407	0.427	0.016	0.012	0.012	-4
		0.300	0.318	0.109	0.306	0.333	0.018	0.012	0.012	-6
		0.500	0.520	0.105	0.510	0.532	0.020	0.011	0.011	-4
		0.100	0.106	0.118	0.093	0.121	0.006	0.014	0.014	-6
	2	0.200	0.213	0.152	0.198	0.229	0.013	0.023	0.023	-7
		0.400	0.410	0.141	0.397	0.422	0.010	0.020	0.020	-2
		0.600	0.609	0.121	0.599	0.621	0.009	0.015	0.015	-1
		0.400	0.415	0.139	0.403	0.427	0.015	0.019	0.020	-4
		0.200	0.205	0.150	0.190	0.220	0.005	0.023	0.023	-2
		0.300	0.307	0.146	0.292	0.324	0.007	0.021	0.021	-2
	3	0.300	0.333	0.088	0.326	0.341	0.033	0.008	0.009	-11
		0.600	0.665	0.074	0.659	0.674	0.065	0.006	0.010	-11
		0.200	0.217	0.090	0.209	0.226	0.017	0.008	0.008	-9
		0.300	0.342	0.084	0.331	0.350	0.042	0.007	0.009	-14
		0.400	0.447	0.087	0.439	0.457	0.047	0.008	0.010	-12
		0.200	0.219	0.088	0.213	0.226	0.019	0.008	0.008	-10
500	1	0.500	0.536	0.070	0.529	0.543	0.036	0.005	0.006	-7
		0.300	0.319	0.080	0.312	0.325	0.019	0.006	0.007	-6
		0.400	0.429	0.079	0.421	0.437	0.029	0.006	0.007	-7
		0.300	0.322	0.079	0.316	0.328	0.022	0.006	0.007	-7
		0.500	0.532	0.071	0.525	0.539	0.032	0.005	0.006	-6
		0.100	0.108	0.083	0.102	0.116	0.008	0.007	0.007	-8
	2	0.200	0.207	0.109	0.202	0.214	0.007	0.012	0.012	-3
		0.400	0.431	0.101	0.422	0.441	0.031	0.010	0.011	-8
		0.600	0.630	0.083	0.622	0.637	0.030	0.007	0.008	-5
		0.400	0.422	0.096	0.413	0.430	0.022	0.009	0.010	-6
		0.200	0.213	0.104	0.200	0.225	0.013	0.011	0.011	-6
		0.300	0.311	0.102	0.299	0.321	0.011	0.010	0.011	-4
	3	0.300	0.343	0.062	0.337	0.350	0.043	0.004	0.006	-14
		0.600	0.672	0.050	0.667	0.676	0.072	0.002	0.008	-12
		0.200	0.225	0.063	0.220	0.229	0.025	0.004	0.005	-12
		0.300	0.338	0.063	0.333	0.342	0.038	0.004	0.005	-13
		0.400	0.446	0.059	0.441	0.451	0.046	0.003	0.006	-12
		0.200	0.221	0.059	0.216	0.227	0.021	0.003	0.004	-11

Table 6.35: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with eight indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.300	0.291	0.225	0.271	0.309	-0.009	0.051	0.051	3
		0.300	0.107	0.231	0.086	0.126	-0.193	0.053	0.091	64
		0.400	0.397	0.212	0.374	0.416	-0.003	0.045	0.045	1
		0.300	0.303	0.224	0.278	0.328	0.003	0.050	0.050	-1
		0.400	0.389	0.221	0.372	0.406	-0.011	0.049	0.049	3
		0.300	0.299	0.236	0.280	0.317	-0.001	0.056	0.056	0
		0.200	0.189	0.228	0.161	0.217	-0.011	0.052	0.052	6
		0.300	0.199	0.243	0.185	0.215	-0.101	0.059	0.069	34
	2	0.300	0.275	0.268	0.256	0.295	-0.025	0.072	0.073	8
		0.300	0.059	0.275	0.039	0.082	-0.241	0.075	0.134	80
		0.400	0.330	0.267	0.304	0.351	-0.070	0.071	0.076	17
		0.400	0.346	0.265	0.317	0.371	-0.054	0.070	0.073	14
		0.200	0.155	0.274	0.128	0.183	-0.045	0.075	0.077	22
		0.300	0.260	0.269	0.236	0.284	-0.040	0.073	0.074	13
		0.400	0.341	0.276	0.324	0.359	-0.059	0.076	0.080	15
		0.200	0.106	0.267	0.076	0.133	-0.094	0.071	0.080	47
	3	0.400	0.187	0.232	0.170	0.205	-0.213	0.054	0.099	53
		0.500	0.410	0.236	0.389	0.427	-0.090	0.056	0.064	18
		0.400	0.416	0.220	0.393	0.440	0.016	0.048	0.049	-4
		0.300	0.311	0.238	0.295	0.329	0.011	0.057	0.057	-4
		0.200	0.197	0.234	0.178	0.219	-0.003	0.055	0.055	1
		0.100	0.111	0.243	0.087	0.130	0.011	0.059	0.059	-11
		0.300	0.305	0.230	0.280	0.331	0.005	0.053	0.053	-2
		0.200	0.214	0.231	0.198	0.238	0.014	0.053	0.054	-7
100	1	0.300	0.325	0.169	0.312	0.337	0.025	0.029	0.029	-8
		0.100	0.114	0.174	0.100	0.130	0.014	0.030	0.030	-14
		0.400	0.425	0.159	0.410	0.438	0.025	0.025	0.026	-6
		0.300	0.326	0.172	0.308	0.339	0.026	0.030	0.030	-9
		0.400	0.440	0.168	0.424	0.454	0.040	0.028	0.030	-10
		0.300	0.324	0.169	0.303	0.347	0.024	0.028	0.029	-8
		0.200	0.223	0.172	0.211	0.240	0.023	0.030	0.030	-11
		0.200	0.222	0.168	0.207	0.240	0.022	0.028	0.029	-11
	2	0.300	0.279	0.205	0.261	0.303	-0.021	0.042	0.043	7
		0.100	0.111	0.223	0.087	0.135	0.011	0.050	0.050	-11
		0.400	0.394	0.198	0.378	0.413	-0.006	0.039	0.039	2
		0.400	0.410	0.208	0.387	0.439	0.010	0.043	0.043	-3
		0.200	0.194	0.214	0.178	0.209	-0.006	0.046	0.046	3
		0.300	0.291	0.213	0.272	0.311	-0.009	0.046	0.046	3
		0.400	0.390	0.204	0.371	0.413	-0.010	0.042	0.042	3
		0.100	0.090	0.213	0.073	0.113	-0.010	0.046	0.046	10
	3	0.200	0.224	0.138	0.213	0.236	0.024	0.019	0.020	-12
		0.400	0.476	0.126	0.460	0.493	0.076	0.016	0.022	-19
		0.400	0.479	0.132	0.467	0.494	0.079	0.017	0.024	-20
		0.300	0.354	0.131	0.344	0.365	0.054	0.017	0.020	-18
		0.200	0.235	0.135	0.223	0.249	0.035	0.018	0.020	-18
		0.100	0.113	0.139	0.098	0.127	0.013	0.019	0.019	-13
		0.300	0.359	0.130	0.350	0.368	0.059	0.017	0.020	-20
		0.200	0.231	0.139	0.221	0.240	0.031	0.019	0.020	-15

Table 6.36: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with eight indicators per construct, 500 runs

N	Block	True Weight	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.300	0.317	0.110	0.307	0.328	0.017	0.012	0.012	-6
		0.300	0.317	0.106	0.310	0.324	0.017	0.011	0.012	-6
		0.400	0.422	0.108	0.414	0.431	0.022	0.012	0.012	-5
		0.300	0.323	0.100	0.315	0.331	0.023	0.010	0.011	-8
		0.400	0.430	0.106	0.419	0.442	0.030	0.011	0.012	-8
		0.300	0.312	0.116	0.303	0.325	0.012	0.014	0.014	-4
		0.200	0.217	0.105	0.207	0.229	0.017	0.011	0.011	-8
		0.300	0.322	0.113	0.309	0.336	0.022	0.013	0.013	-7
	2	0.300	0.304	0.142	0.296	0.314	0.004	0.020	0.020	-1
		0.300	0.306	0.141	0.297	0.315	0.006	0.020	0.020	-2
		0.400	0.404	0.137	0.394	0.416	0.004	0.019	0.019	-1
		0.400	0.413	0.134	0.398	0.427	0.013	0.018	0.018	-3
		0.200	0.205	0.143	0.193	0.219	0.005	0.021	0.021	-2
		0.300	0.300	0.141	0.289	0.313	0.000	0.020	0.020	0
		0.400	0.407	0.134	0.394	0.421	0.007	0.018	0.018	-2
		0.200	0.192	0.145	0.179	0.205	-0.008	0.021	0.021	4
	3	0.400	0.424	0.082	0.418	0.431	0.024	0.007	0.007	-6
		0.500	0.534	0.079	0.526	0.542	0.034	0.006	0.007	-7
		0.400	0.428	0.082	0.422	0.437	0.028	0.007	0.008	-7
		0.300	0.320	0.082	0.313	0.328	0.020	0.007	0.007	-7
		0.200	0.210	0.088	0.203	0.217	0.010	0.008	0.008	-5
		0.100	0.108	0.092	0.099	0.118	0.008	0.008	0.009	-8
		0.300	0.318	0.084	0.309	0.326	0.018	0.007	0.007	-6
		0.200	0.215	0.099	0.205	0.225	0.015	0.010	0.010	-7
500	1	0.300	0.330	0.080	0.322	0.337	0.030	0.006	0.007	-10
		0.300	0.329	0.078	0.323	0.337	0.029	0.006	0.007	-10
		0.400	0.428	0.073	0.422	0.436	0.028	0.005	0.006	-7
		0.300	0.329	0.074	0.323	0.336	0.029	0.006	0.006	-10
		0.400	0.433	0.078	0.426	0.439	0.033	0.006	0.007	-8
		0.300	0.324	0.085	0.317	0.333	0.024	0.007	0.008	-8
		0.200	0.211	0.084	0.201	0.219	0.011	0.007	0.007	-5
		0.300	0.327	0.079	0.318	0.334	0.027	0.006	0.007	-9
	2	0.300	0.314	0.103	0.306	0.323	0.014	0.011	0.011	-5
		0.300	0.314	0.101	0.306	0.321	0.014	0.010	0.010	-5
		0.400	0.415	0.095	0.407	0.425	0.015	0.009	0.009	-4
		0.400	0.424	0.097	0.415	0.432	0.024	0.009	0.010	-6
		0.200	0.214	0.105	0.205	0.222	0.014	0.011	0.011	-7
		0.300	0.327	0.102	0.316	0.338	0.027	0.010	0.011	-9
		0.400	0.423	0.098	0.414	0.433	0.023	0.010	0.010	-6
		0.200	0.210	0.102	0.200	0.219	0.010	0.010	0.011	-5
	3	0.400	0.431	0.058	0.425	0.436	0.031	0.003	0.004	-8
		0.500	0.544	0.057	0.538	0.551	0.044	0.003	0.005	-9
		0.400	0.423	0.063	0.418	0.429	0.023	0.004	0.004	-6
		0.300	0.326	0.063	0.318	0.333	0.026	0.004	0.005	-9
		0.200	0.217	0.063	0.211	0.223	0.017	0.004	0.004	-8
		0.100	0.103	0.064	0.098	0.109	0.003	0.004	0.004	-3
		0.300	0.327	0.061	0.322	0.334	0.027	0.004	0.004	-9
		0.200	0.214	0.062	0.208	0.220	0.014	0.004	0.004	-7

Table 6.37: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with two indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.421	0.095	0.412	0.428	-0.079	0.009	0.015	16
	0.400	0.334	0.104	0.323	0.343	-0.066	0.011	0.015	17
	0.500	0.412	0.096	0.402	0.421	-0.088	0.009	0.017	18
100	0.500	0.422	0.061	0.416	0.428	-0.078	0.004	0.010	16
	0.400	0.336	0.068	0.330	0.342	-0.064	0.005	0.009	16
	0.600	0.502	0.057	0.497	0.508	-0.098	0.003	0.013	16
250	0.500	0.421	0.040	0.418	0.424	-0.079	0.002	0.008	16
	0.400	0.341	0.040	0.338	0.346	-0.059	0.002	0.005	15
	0.600	0.510	0.034	0.507	0.513	-0.090	0.001	0.009	15
500	0.500	0.425	0.027	0.423	0.427	-0.075	0.001	0.006	15
	0.400	0.340	0.028	0.337	0.342	-0.060	0.001	0.004	15
	0.600	0.510	0.025	0.508	0.513	-0.090	0.001	0.009	15

Table 6.38: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with four indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.428	0.088	0.420	0.437	-0.072	0.008	0.013	14
	0.400	0.333	0.115	0.322	0.342	-0.067	0.013	0.018	17
	0.500	0.432	0.088	0.425	0.439	-0.068	0.008	0.012	14
100	0.500	0.438	0.055	0.434	0.443	-0.062	0.003	0.007	12
	0.400	0.354	0.058	0.349	0.359	-0.046	0.003	0.005	12
	0.600	0.526	0.050	0.521	0.532	-0.074	0.002	0.008	12
250	0.500	0.454	0.031	0.452	0.457	-0.046	0.001	0.003	9
	0.400	0.362	0.034	0.359	0.365	-0.038	0.001	0.003	10
	0.600	0.542	0.032	0.539	0.546	-0.058	0.001	0.004	10
500	0.500	0.457	0.024	0.455	0.459	-0.043	0.001	0.002	9
	0.400	0.365	0.024	0.363	0.367	-0.035	0.001	0.002	9
	0.600	0.552	0.021	0.550	0.554	-0.048	0.000	0.003	8

Table 6.39: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with six indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.417	0.086	0.410	0.423	-0.083	0.007	0.014	17
	0.400	0.333	0.100	0.324	0.343	-0.067	0.010	0.014	17
	0.500	0.429	0.090	0.420	0.437	-0.071	0.008	0.013	14
100	0.500	0.433	0.057	0.428	0.438	-0.067	0.003	0.008	13
	0.400	0.339	0.064	0.332	0.345	-0.061	0.004	0.008	15
	0.600	0.520	0.052	0.515	0.523	-0.080	0.003	0.009	13
250	0.500	0.451	0.032	0.448	0.455	-0.049	0.001	0.003	10
	0.400	0.358	0.036	0.354	0.362	-0.042	0.001	0.003	11
	0.600	0.542	0.030	0.539	0.545	-0.058	0.001	0.004	10
500	0.500	0.458	0.022	0.456	0.461	-0.042	0.000	0.002	8
	0.400	0.368	0.024	0.366	0.370	-0.032	0.001	0.002	8
	0.600	0.552	0.022	0.550	0.555	-0.048	0.000	0.003	8

Table 6.40: True path coefficients, mean path coefficient estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with eight indicators per latent variable, 500 runs

N	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.500	0.421	0.086	0.415	0.427	-0.079	0.007	0.014	16
	0.400	0.331	0.107	0.323	0.340	-0.069	0.011	0.016	17
	0.500	0.411	0.085	0.402	0.419	-0.089	0.007	0.015	18
100	0.500	0.423	0.054	0.418	0.428	-0.077	0.003	0.009	15
	0.400	0.328	0.060	0.323	0.333	-0.072	0.004	0.009	18
	0.600	0.521	0.049	0.517	0.526	-0.079	0.002	0.009	13
250	0.500	0.454	0.033	0.451	0.457	-0.046	0.001	0.003	9
	0.400	0.360	0.034	0.357	0.363	-0.040	0.001	0.003	10
	0.600	0.546	0.030	0.543	0.549	-0.054	0.001	0.004	9
500	0.500	0.465	0.021	0.463	0.468	-0.035	0.000	0.002	7
	0.400	0.372	0.024	0.369	0.374	-0.028	0.001	0.001	7
	0.600	0.560	0.020	0.558	0.562	-0.040	0.000	0.002	7

Table 6.41: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with two indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.7	0.866	0.042	0.860	0.871	0.166	0.002	0.029	-24
	0.8	0.899	0.028	0.897	0.902	0.099	0.001	0.011	-12
100	0.7	0.866	0.027	0.864	0.869	0.166	0.001	0.028	-24
	0.8	0.899	0.018	0.898	0.902	0.099	0.000	0.010	-12
250	0.7	0.865	0.017	0.864	0.867	0.165	0.000	0.028	-24
	0.8	0.899	0.010	0.898	0.900	0.099	0.000	0.010	-12
500	0.7	0.866	0.012	0.864	0.867	0.166	0.000	0.028	-24
	0.8	0.900	0.007	0.899	0.900	0.100	0.000	0.010	-12

Table 6.42: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with four indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.713	0.075	0.706	0.720	0.113	0.006	0.018	-19
	0.7	0.792	0.055	0.787	0.795	0.092	0.003	0.011	-13
	0.8	0.858	0.034	0.855	0.860	0.058	0.001	0.004	-7
	0.9	0.901	0.022	0.899	0.904	0.001	0.000	0.001	0
100	0.6	0.714	0.057	0.710	0.718	0.114	0.003	0.016	-19
	0.7	0.789	0.039	0.786	0.792	0.089	0.002	0.010	-13
	0.8	0.857	0.025	0.855	0.859	0.057	0.001	0.004	-7
	0.9	0.902	0.015	0.901	0.904	0.002	0.000	0.000	0
250	0.6	0.714	0.036	0.711	0.718	0.114	0.001	0.014	-19
	0.7	0.793	0.023	0.791	0.796	0.093	0.001	0.009	-13
	0.8	0.857	0.016	0.856	0.859	0.057	0.000	0.004	-7
	0.9	0.903	0.009	0.902	0.904	0.003	0.000	0.000	0
500	0.6	0.711	0.024	0.708	0.712	0.111	0.001	0.013	-18
	0.7	0.794	0.016	0.793	0.794	0.094	0.000	0.009	-13
	0.8	0.857	0.011	0.856	0.858	0.057	0.000	0.003	-7
	0.9	0.903	0.007	0.902	0.903	0.003	0.000	0.000	0

Table 6.43: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with six indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.680	0.087	0.670	0.689	0.080	0.007	0.014	-13
	0.7	0.762	0.062	0.756	0.767	0.062	0.004	0.008	-9
	0.8	0.835	0.039	0.831	0.839	0.035	0.001	0.003	-4
	0.9	0.893	0.023	0.891	0.896	-0.007	0.001	0.001	1
	0.6	0.673	0.087	0.665	0.682	0.073	0.008	0.013	-12
	0.7	0.761	0.058	0.758	0.765	0.061	0.003	0.007	-9
100	0.6	0.671	0.063	0.663	0.677	0.071	0.004	0.009	-12
	0.7	0.763	0.042	0.759	0.766	0.063	0.002	0.006	-9
	0.8	0.836	0.028	0.833	0.838	0.036	0.001	0.002	-4
	0.9	0.895	0.016	0.894	0.897	-0.005	0.000	0.000	1
	0.6	0.678	0.056	0.673	0.684	0.078	0.003	0.009	-13
	0.7	0.763	0.039	0.759	0.767	0.063	0.002	0.006	-9
250	0.6	0.675	0.033	0.672	0.679	0.075	0.001	0.007	-13
	0.7	0.764	0.025	0.761	0.766	0.064	0.001	0.005	-9
	0.8	0.835	0.017	0.834	0.837	0.035	0.000	0.002	-4
	0.9	0.895	0.010	0.894	0.896	-0.005	0.000	0.000	1
	0.6	0.674	0.035	0.671	0.677	0.074	0.001	0.007	-12
	0.7	0.763	0.024	0.760	0.766	0.063	0.001	0.005	-9
500	0.6	0.678	0.025	0.675	0.680	0.078	0.001	0.007	-13
	0.7	0.762	0.017	0.761	0.764	0.062	0.000	0.004	-9
	0.8	0.836	0.012	0.835	0.837	0.036	0.000	0.001	-5
	0.9	0.895	0.007	0.894	0.896	-0.005	0.000	0.000	1
	0.6	0.678	0.024	0.675	0.680	0.078	0.001	0.007	-13
	0.7	0.762	0.019	0.760	0.764	0.062	0.000	0.004	-9

Table 6.44: True loadings, mean loading estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative biases for case C, models with eight indicators per construct, 500 runs

N	True Loading	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.660	0.083	0.652	0.667	0.060	0.007	0.010	-10
	0.7	0.749	0.060	0.744	0.756	0.049	0.004	0.006	-7
	0.8	0.830	0.038	0.825	0.833	0.030	0.001	0.002	-4
	0.9	0.900	0.023	0.899	0.902	0.000	0.001	0.001	0
	0.6	0.650	0.091	0.642	0.658	0.050	0.008	0.011	-8
	0.7	0.749	0.062	0.745	0.753	0.049	0.004	0.006	-7
	0.8	0.827	0.039	0.823	0.830	0.027	0.002	0.002	-3
	0.9	0.902	0.021	0.901	0.904	0.002	0.000	0.000	0
100	0.6	0.658	0.059	0.652	0.665	0.058	0.003	0.007	-10
	0.7	0.749	0.041	0.747	0.752	0.049	0.002	0.004	-7
	0.8	0.829	0.028	0.827	0.831	0.029	0.001	0.002	-4
	0.9	0.901	0.014	0.900	0.903	0.001	0.000	0.000	0
	0.6	0.659	0.060	0.654	0.665	0.059	0.004	0.007	-10
	0.7	0.746	0.043	0.742	0.751	0.046	0.002	0.004	-7
	0.8	0.829	0.029	0.827	0.832	0.029	0.001	0.002	-4
	0.9	0.901	0.015	0.900	0.902	0.001	0.000	0.000	0
250	0.6	0.655	0.036	0.651	0.659	0.055	0.001	0.004	-9
	0.7	0.749	0.026	0.747	0.751	0.049	0.001	0.003	-7
	0.8	0.831	0.018	0.829	0.832	0.031	0.000	0.001	-4
	0.9	0.901	0.009	0.900	0.901	0.001	0.000	0.000	0
	0.6	0.658	0.034	0.656	0.661	0.058	0.001	0.005	-10
	0.7	0.747	0.028	0.745	0.750	0.047	0.001	0.003	-7
	0.8	0.830	0.018	0.829	0.832	0.030	0.000	0.001	-4
	0.9	0.901	0.010	0.900	0.902	0.001	0.000	0.000	0
500	0.6	0.657	0.026	0.655	0.659	0.057	0.001	0.004	-10
	0.7	0.748	0.018	0.746	0.750	0.048	0.000	0.003	-7
	0.8	0.830	0.012	0.828	0.831	0.030	0.000	0.001	-4
	0.9	0.901	0.006	0.901	0.902	0.001	0.000	0.000	0
	0.6	0.658	0.026	0.655	0.660	0.058	0.001	0.004	-10
	0.7	0.748	0.018	0.747	0.750	0.048	0.000	0.003	-7
	0.8	0.829	0.012	0.828	0.830	0.029	0.000	0.001	-4
	0.9	0.901	0.007	0.901	0.902	0.001	0.000	0.000	0

Chapter 7

Toward the Definition of a Structural Equation Model of Patent Value

Abstract. This chapter aims to propose a structural equation model which relates the variables that determine the patent value. Even though some patent indicators have been directly used to infer the private or social value of innovations, the results suggest that patent value is a more complex variable that may be modeled as an endogenous unobservable variable in a first- and in a second-order model, and which depends respectively on three and four constructs. Such variables include the knowledge used by companies to create their inventions, the technological scope of the inventions, the international scope of protection, and the technological usefulness of the inventions. The model allows the conceptualization of patent value into a potential and a recognized value of intangible assets, aiming toward an index construction approach. Partial Least Squares (PLS) Path Modelling is performed as an exploratory model-building procedure. We use a sample of 2,901 patents granted in the United States in the field of renewable energy.

7.1 Introduction

Patents are one of the main sources of technological information. A patent is an exclusive right granted to inventors by a state only when the invention fulfils three basic requirements: the invention is new, it involves an inventive activity and it is useful for industry. Until now research involving patent data has been associated with the analysis of information contained in the patent document, such as backward and forward citations or number of claims, and the relationship between patents and research and development (R&D), innovation or economic growth. In recent years, patent indicators have been used to study the economical value of patents. In most cases, analytical approaches have been based on standard econometric analysis techniques such as probit or logit models, and survey analysis. However, patent value may be seen as a complex construct depending on a variety of elements. General and specific market conditions, countries' legal frameworks, geographic proximity or accumulated scientific and technological knowledge are different dimensions that have shown to affect patent value.

This paper proposes that a holistic and multidimensional model may offer a robust understanding of the different variables that determine patent value. For the moment, and considering patent document information, two path models are built considering five dimensions represented by five constructs. They are: patent value, technological usefulness of the invention, knowledge stock used by the company to create the technology, technological scope of the invention, and international scope of protection. The models are strongly based on the theory developed by the technological change scientific community and a thorough review of the literature on patent valuation. Each construct is associated with a set of observable variables. So, they can be estimated by these indicators. Manifest variables are mainly built from information contained in patent documents. A set of patents granted in the United States (U.S.) in the area of renewable energies was retrieved from Delphion database. The proposed path models are replicable because they could be repeated for different technological fields or countries. Moreover, the models may allow one to distinguish between: (a) those variables related to patent value at the time of application, i.e. those variables that could deliver a measure of potential value of patents, and (b) those that determine the value after the patent's application.

In the literature, research that addresses patent value using a structural equation model (SEM) approach is quite scarce. Moreover, rather traditional methods based on multivariate normal distribution assumption have been implemented. The advantage of SEM is flexibility in working with theory and data, approaching the whole phenomenon, and a more complete representation of the complex theory. Additionally, and contrary to a covariance-based approach such as the linear structural relation model (LISREL), PLS Path Modelling is theory-building-oriented and causal-predictive-oriented. Therefore, the exploratory nature of this procedure allows for the first formulation of a structural model of patent value. Finally, the PLS Path Modelling algorithm is a powerful technique for the analysis of skewed or long-tail data, such as patent data. Therefore, we also attempt to show the benefits of PLS Path Modelling as a tool for exploration and prediction of skewed data.

In this research, the models specification is made from a PLS perspective. So, we are posing PLS models. Section 7.2 provides background on patent indicators and constructs, and section 7.3 reviews the PLS Path Modelling procedure for hierarchical component models with repeated manifest variables and formative constructs. Section 7.4 addresses the first- and second-order model formulation, while also postulating on the indicators, latent variables (LVs) and causal relationships among variables. In particular, formative and reflective relationships among manifest and latent variables are justified. A description of patent data is given in Section 7.5. Section 7.6 reports the results, and shows the performance and effectiveness of PLS Path Modelling when working with patent data characterized by long tails. Finally, section 7.7 gives final remarks and some directions for future research.

7.2 Background

7.2.1 Patent Indicators and Constructs

Patent indicators have been used by scientific communities to study phenomena such as technological change or the growth of science and technology. Forward citations, i.e. the number of times that each patent has been cited by another patent, are the most widely used indicator to measure the value or importance of patents. Nevertheless, other indicators have also been introduced as a measure of value, such as family size, number of claims, number of international patent classification (IPC) codes where the patent is classified, and backward citations. Here, family size refers to the number of countries where a patent is sought for the same invention (Lanjouw, 1998). As a general patenting strategy, companies protect their inventions in their local countries first and then in other jurisdictions. Patents with a large family size tend to be more valuable or important (Harhoff *et al.*, 2003), although Guellec et al. reported that this relationship might sometimes be inaccurate and "may reflect a lack of maturity of the applicant" (Guellec & van Pottelsberghe, 2000, p. 114). Even so, family size may be proposed as a proxy variable for the international scope of patent rights, and as a measure of patent value. The number of backward citations or references in a patent represents "all of the important prior art upon which the issued patent improves" (Narin *et al.*, 1997, p. 318), and allows one to demonstrate that the invention is genuinely new. Claims are made in a special section in the patent document, where the thing that is being protected is specified. The claims section consists of a numbered list. Therefore, the number of claims is in fact the number of inventions protected (Tong & Frame, 1994, p. 134). Patents with a large number of claims have a higher likelihood of being litigated, so they can be considered more valuable (Harhoff & Reitzig, 2004; Lanjouw & Schankerman, 2001; Reitzig, 2004a). International patent classification classes were introduced as a proxy variable for the scope of protection by Lerner (1994). An invention with a larger technological scope should be more valuable due to its broader potential applications. The number of inventors and the number of applicants have also been used as indicators of the patent value (Reitzig, 2004a).

Most patent indicators have been used to explain a conceptual variable or a construct. The relationship between patent citations and patent value has been deeply studied (Albert *et al.*, 1991; Carpenter *et al.*, 1981; Guellec & van Pottelsberghe, 2000; Harhoff *et al.*, 1999, 2003; Reitzig, 2003, 2004a; Trajtenberg, 1990b). Carpenter *et al.* (1981), Albert *et al.* (1991) and Harhoff *et al.* (1999) have successfully shown that those patents that are related to important technological developments are most highly cited. Harhoff *et al.* (2003) was the first to use backward and forward citations together as proxy variables for patent value, and Trajtenberg (1990b) established the role of citations as an indicator of the value of innovations. Patent citations and patent value have also been associated with market value and/or the R&D expenditures of companies (Connolly & Hirschey, 1988; Griliches, 1981; Hall *et al.*, 2005; Lerner, 1994). The relationship among patent value and patenting strategy, technological diversity (through the IPC), domestic and international R&D collaborations and/or co-applications (analyzing the country of residence of the authors) and the mix of designated states for protection (through the family size), have been studied by Guellec & van Pottelsberghe (2000). Reitzig (2003, 2004a) studied the factors that determine an individual patent value. Analyzing the results of a questionnaire, he found that novelty and inventive activity are the most important factors in patents that are used as "bargaining chips." Connolly & Hirschey (1988) showed that patent statistics are significantly related to companies' market value. In addition, Griliches (1981) found a significant relation among companies' market value, the book value of R&D expenditures and the number of patents. He based his research on a time-series cross-section analysis of United States firm data. Lerner (1994) reported that patent scope has a significant impact on the valuation of firms, while Hall *et al.* (2005) investigated the trend in US patenting activities over the last 30 years, finding that the ratios of R&D to asset stock, patents to R&D, and citations to patents significantly affect companies' market value.

On the other hand, some of these indicators have been related to other constructs. The number of inventors and applicants, backward citations and the number of claims have been related to patent novelty, i.e. the technological distance between a protected invention and prior art. A patent's protection level or its technological scope or breadth can be measured by the number of claims or number of IPC classes into which the patent is classified (Lerner, 1994). Furthermore, patent stocks or knowledge stocks have been associated with the economic growth of a country as well as the economical activity (Griliches, 1990), research and development results (Lanjouw & Schankerman, 2004) and the value of innovation (Sherry & Teece, 2004) and technological performance (Tong & Frame, 1994). In this last case, the researchers found that the number of claims is a better indicator than the number of patents in the national technological capacity.

Finally, little research has reported on the structural relationship among latent variables which influence patent value using a multidimensional approach. The recent investigations of Harhoff *et al.* (2003); Harhoff & Reitzig (2004) and Reitzig (2003, 2004a) used a large number of indicators of patent value aimed mainly at estimating the probability of opposition to a patent. In most cases, analytical approaches have been based on standard econometric analysis techniques (probit or logit models) or survey analysis. One reason that could explain why a multi-dimensional and structural approach has not been applied to technology/patent valuation is that more general structural models are based on maximum likelihood estimation and the multivariate normal distribution of data. Patent indicators are very heterogeneous and asymmetric, and, in general, they exhibit a large variance and skew. Consequently, assuming that this type of data has a multivariate normal distribution may lead to biased results. As seen below, PLS Path Modelling overcomes this drawback because it is an iterative algorithm that makes no assumptions about data distribution. Moreover, unlike other methods such as probit or logit models, it allows researchers to depict the relationship among a set of latent variables. Thus, we have the possibility of modelling the patent value as an unobservable variable.

7.2.2 Patent Value

Patents are intellectual assets that do not necessarily have an immediate return. A patent may protect a product that can be manufactured and sold. But a patent may also protect technologies which, together with other technologies, enable the manufacture of a final product. In both cases, to obtain an economic value from patents may be extremely difficult. In studying patent value, different approaches have been taken throughout the literature. Some of the approaches focus on the private value of a patent while others concentrate on a patent's social value. Lanjouw (1998, p. 407) defined the private value of a patent in terms of "the difference in the returns that would accrue to the innovation with and without patent protection." The magnitude of this difference would be crucial in applying or renewing the protection. Reitzig (2004a) also focused on the private value of patents, and specifies the need to consider the patent value as a construct. Technical experts were surveyed and, according to them, the research showed that the factors that determine patent value are: state of the art (existing technologies), novelty, inventiveness, breadth, difficulty of inventing, disclosure and dependence on complementary assets[1]. Additionally, Trajtenberg (1990b) showed that patent data was highly correlated with some indicators of the social benefits of innovations. Guellec & van Pottelsberghe (2000) presented a value scale proposing that technology increases its own value as it passes through different stages: from invention to application, examination, publication and decision to grant, and finally to the high value stage if the patent is granted. The distinction is made between the intrinsic value of the patent simply for being granted (and thereby having proven novelty, inventive activity and applicability) and the potential value of technology (dependent on its potential for generating future returns).

Some patent indicators have been used to directly infer the patent's value, such as forward citations or family size (see Table 7.1). Even though this may be useful and may give an approximation of the patent value, many elements may affect the invention and protection process. We consider some of these factors based on the presented background, and represent their interactions proposing a multidimensional analysis of the problem. It is worth noting that this research does not seek to determine the value of an individual patent or to obtain a monetary value of the assets. Rather, the patent value is proposed in terms of the technological usefulness of the inventions. This model, however, allows us to compare and rank the value of company's patent portfolios. We address the question of what variables determine the patent value and how they relate to each other. These variables are modeled as unobserved variables. So, they and their relationships set up a structural equation model.

7.3 The PLS Path Modelling Approach for Model Formulation

PLS Path Modelling is a component-based procedure for estimating a sequence of latent variables developed by the statistician and econometrician Herman Wold (1980, 1982, 1985). During the last few years, it has proved to be useful for estimating structural models, in marketing and information system research in particular, and in the social sciences in general (Chin, 1998a; Esposito Vinzi, 2007; Henseler *et al.*, 2009; Hulland, 1999; Marcoulides, 2003; Tenenhaus *et al.*, 2005). Some of its features have encouraged its use, such as: (1) it is an iterative algorithm that offers an explicit estimation of the latent variables, and their relationships, (2) it works with few cases and makes no assumptions about data distribution -in contrast with LISREL that makes strong assumptions about data distribution and where hundreds of cases are necessary for its application, and (3) it overcomes the

[1]We attempt to consider these variables as constructs in the proposed structural model. However, recall that in this research, the manifest variables are mainly obtained from the patent document. So, latent and manifest variables are subject to this constraint.

Table 7.1: Brief summary of approaches used to study the patent value

Author	**Construct**	**Indicators**	**Dependent variable**	**Method**
Trajtenberg (1990)	Social value of innovations	Patent count weighted by citations	Consumer surplus	Multinomial logit model
Guellec et al. (2000)	Patent value, patenting strategy, technological diversity, R&D collaboration	Number of IPC, family size, dummy variables, etc.	Probability that a EPO patent application is granted	Probit model
Reitzig (2003)	Patent value, novelty, inventive activity, invent around, disclosure	-	'present patent value'	Survey, probit model
Harhoff et al. (2003)	Private value of patents, value of renewed patent protection and asset value of patent right	Survey of patent-holders, backward and forward citations, family size, IPC, outcome of opposition proceedings	Patent right as a price to sell the patent right	Survey, probit model
Hall et al. (2005)	Market value	Patent citations, R&D expenditures, total assets	Tobin's q	Tobin's Q equation

identification problems when formative measurement models are included. Wold (1985) emphasizes that "using prior knowledge and intuition the investigator is free to specify the LVs, to design the inner relations, and to compile a selection of indicators for each LV" [p. 582]. The path model "is usually tentative since the model construction is an evolutionary process. The empirical content of the model is extracted from the data, and the model is improved by interactions through the estimation between the model and the data and the reactions of the researcher" (Wold, 1980, p. 70).

In a PLS Path Modelling approach, the structural model or inner model –also called the inner relations and substantive theory– depicts the relationship among latent variables as multiple regressions (Equation 7.1). The arrangement of the structural model is strongly supported by theory at the model specification stage. So, PLS Path Modelling is used to explore if these relationships hold up or whether other theory-based specifications, that may be proposed, help in providing a better explanation for a particular phenomenon.

$$\xi_j = \beta_{j0} + \sum_i \beta_{ji}\xi_i + \nu_j \tag{7.1}$$

ξ_j and ξ_i are the endogenous and exogenous latent variables, respectively. β_{ji} are called path coefficients, and they measure the relationship among constructs.

The condition imposed is $E(\xi_j/\xi_i) = \sum_i \beta_{ji}\xi_i$. There is no linear relationship between predictor and residual, $E(\nu_j/\forall\xi_i) = 0$ and $cov(\nu_j, \xi_i) = 0$.

The measurement model or outer model –also called the outer relations– describes the relationship between latent (ξ_i) and manifest (x_{ih}) variables in two different ways: Mode A and Mode B. "Mode A is often used for an endogenous LV and Mode B for an exogenous one. Mode A is appropriate for a block with a reflective measurement model and Mode B for a formative one" (Tenenhaus *et al.*, 2005, p. 268). Reflective relationships seek to represent variance and covariances between the manifest variables that are generated or caused by a latent variable. So, observed variables are treated as an effect of unobserved variables (Bollen & Lennox, 1991; Cohen *et al.*, 1990). In a reflective measurement model, the manifest variables are measured with error. Alternatively, formative relationships are used to minimize residuals in the structural relationships (Fornell & Bookstein, 1982), and here, manifest variables are treated as forming the unobserved variables. MacCallun & Browne (1993) said that observed variables in a formative model are exogenous measured variables. In a formative outer model the manifest variables are presumed to be error-free and the unobserved variable is estimated as a linear combination of the manifest variables plus a disturbance term, so they are not true latent variables (as in the traditional factorial approach). As in this case all variables forming the construct should be considered, the disturbance term represents all those non-modeled causes.

In Mode A or in reflective relationships, manifest and latent variables relationships are described by ordinary least square regressions (Equation 7.2). The parameters π_h are called loadings. The condition imposed is $E(x_h/\xi) = \pi_{h0} + \pi_h\xi$, ϵ_h with zero mean and uncorrelated with ξ. Loadings indicate the extent to which each indicator reflects the construct, and represent the correlation between indicators and component scores.

$$x_{ih} = \pi_{ih0} + \pi_{ih}\xi_i + \epsilon_{ih} \tag{7.2}$$

In Mode B or in formative relationships, unobserved variables are generated by their own manifest variables as a linear function of them and a residual (Equation 7.3). The parameters w_h are called weights, and allow us to determine the extent to which each indicator contributes to the formation of the constructs. Each block of manifest variables may be multidimensional. The condition imposed is $E(\xi/x_h) = \sum_h w_h x_h$. This implies that the residuals δ_i have zero mean and they are uncorrelated with the manifest variables x_i.

$$\xi_i = \sum_h w_{ih}x_{ih} + \delta_i \tag{7.3}$$

Wold's basic-design of PLS Path Modelling (Wold, 1980, 1982, 1985) does not consider higher-order latent variables. Therefore, in Wold's algorithm each construct must be related to a set of observed variables in order to be estimated. However, Lohmöller (1989) proposed a procedure for the case of hierarchical constructs; that is to say, for cases where there is a construct that does not have a block of measurement variables, or more simply: it is only related to other constructs. In hierarchical component modelling, manifest variables of first-order latent variables are repeated for the second-order latent variable. So, a set of "auxiliary" variables is introduced for estimation purposes. After that, the model is estimated using PLS Path Modelling in the usual way. Hence, the specification of PLS has an additional equation that Lohmöller (1989) called the cross-level relation (see Equation 7.4). The condition imposed is $E(\xi_j \epsilon_{jl}) = 0$. We are interested in this type of model because, as seen below, the patent value construct may be modeled as a second-order latent variable, i.e. the value can only be estimated through linear relations with other latent variables.

$$y_{jl} = \pi_{jl0} + \pi_{jl}\xi_j + \epsilon_{jl} \tag{7.4}$$

Reliability of reflective measurement models is evaluated by examining loadings. A rule of thumb generally accepted is 0.7 or more. This implies that "there is more shared variance between construct and variable than error variance" (Hulland, 1999, p. 198). A low value in a loading factor suggests that the indicator has little relation to the associated construct. All indicators of a block of variables must reflect the same construct. Therefore, there should be high collinearity within each block of variables. Thus, the internal consistency of a reflective measurement model is related to the coherence between constructs and their measurement variables. The unidimensionality of the block of variables may be assessed by using Cronbach's alpha coefficient (should be > 0.7), and composite reliability (should be > 0.7). According to Chin (1998a, p. 320) "alpha tends to be a lower bound estimate of reliability whereas composite reliability is a closer approximation under the assumption that the parameter estimates are accurate."

To represent the extent to which measures of a given construct differ from measures of other constructs (discriminant validity), the average variance extracted (AVE) may be calculated. Therefore, as suggested by Fornell & Larcker (1981), the percentage of variance captured by the construct in relation to the variance due to random measurement error is computed (should be > 0.5). Likewise when models have more than two reflective constructs, cross loadings may be obtained by calculating the correlations between component scores and indicators associated with other reflective constructs. If an indicator has higher correlation with another latent variable instead of the associated latent variable, its position should be reconsidered in the model. Therefore, each indicator has to be more related to its construct than another one in the same model. To assess the significance of loadings, weights and path coefficients, standard errors and t-values may be computed by bootstrapping (200 samples; t-value > 1.65 significant at the 0.05 level; t-value > 2 significant at the 0.01 level).

The inner model is assessed by examining the path coefficients among latent variables. The value of path coefficients provides evidence regarding the strength of the association among latent variables. Moreover, the coefficient of determination (R-square) of each endogenous variable gives the overall fit of the model or the percentage of variance explained by the model. In this research, PLS Path Modelling and bootstrapping were carried out in SmartPLS (Ringle *et al.*, 2005b) with a centroid weighting scheme.

7.3.1 A Brief Overview of Formative and Reflective Outer Models

The distinction between reflective and formative measurement models for structural equation models is an issue that has been addressed by several scientific communities. Major contributions have been made by researchers from statistics (Cohen *et al.*, 1990), psychology and sociology (Bollen & Lennox, 1991; Bollen & Ting, 2000), information science (Petter *et al.*, 2007), and business and marketing research (Diamantopoulos & Winklhofer, 2001; Fornell & Bookstein, 1982). There are some decision rules criteria to determine if a relationship should be modeled as formative or reflective (Mode B or Mode A in the Wold's PLS approach). The guidelines can be summarized in five points as follows (Cohen *et al.*, 1990; Fornell & Bookstein, 1982; McDonald, 1996). (1) The strong theory and the previous knowledge of a phenomenon under study should help to clarify the generative nature of the construct. When a formative relationship is considered, manifest variables must cover the entire scope of construct. (2) Correlations among manifest variables. In a reflective outer model, manifest variables have to be highly correlated; in contrast this condition must not be applied in a formative outer model. (3) Within-construct correlations versus between-construct correlations. This is a common practice in the model specification stage by means of cross-validation; the applied rule is that the former should be greater than the latter. However, Bollen & Lennox (1991) show that this may lead to an incorrect indicator selection for reflective and formative outer models, because this rule may have exceptions. So, the condition must be applied with caution. (4) Sample size and multicollinearity affect the stability of indicator coefficients, and they are a frequent problem in multiple regressions. So, multicollinearity will influence the quality of the estimates in formative relationships. (5) Interchangeability. This concept refers to whether or not the manifest variables share the same concept

(Diamantopoulos & Winklhofer, 2001; Jarvis *et al.*, 2003). All manifest variables in a reflective model explain the same construct. So, removing an indicator from the block of variables should not have a significant effect on the construct. The situation is completely different when considering formative outer models. The indicators do not have to be interchangeable or share the same concept. That is what Bollen & Lennox (1991) called "sampling facets of a construct"; in other words manifest variables of a formative block of variables should represent all the aspects that form the concept. Finally, Gudergan *et al.* (2008) recently proposed a procedure based on tetrad analysis to distinguish between a reflective and formative measurement model in a component-based approach. However, when an outer model has less than four observed variables, this procedure requires adding manifest variables from other outer models. Therefore, the discussion on the reflective and formative nature of the constructs studied here is based mainly on the five rules presented previously.

7.4 Patent Value Models

Two models were tested. First of all, we are interested in knowing the relationships among patent indicators, patent value, and different constructs which up to now have been studied and identified as patent value determinants[2]. In previous research, these constructs have not been modeled as unobservable variables, such as in a structural equation model approach. So, the model formulation began by defining the patent value as an endogenous latent variable, since it is the primary variable to be estimated in the model. Summarizing the results of previous researchers, three unobserved variables related to the dependent variable were identified as exogenous: the knowledge stock of the patent, the technological scope of the invention, and the international scope of the protection (see Figure 7.1). We took into account all of the measurement variables found in the state of the art, and which can be computed from information contained in the patent document. Nevertheless, indicators constructed from the patent text, such as from the abstract or technical description, are excluded from this study.

The knowledge stock represents the base of knowledge that was used by the applicant to create an invention. This would be the content domain. This existing knowledge encourages the inventive activity and may come from within or outside the company. We would like to find those indicators that are value determinants, and that companies may use to make decisions. Since we are considering the patent document as the main data source, the applicants and inventors –that have contributed their knowledge to the creation of the invention– may be considered as

[2]It is worth noting that we are not interested in explaining the variance and covariance among manifest variables as in a covariance-based approach, at least not at this stage.

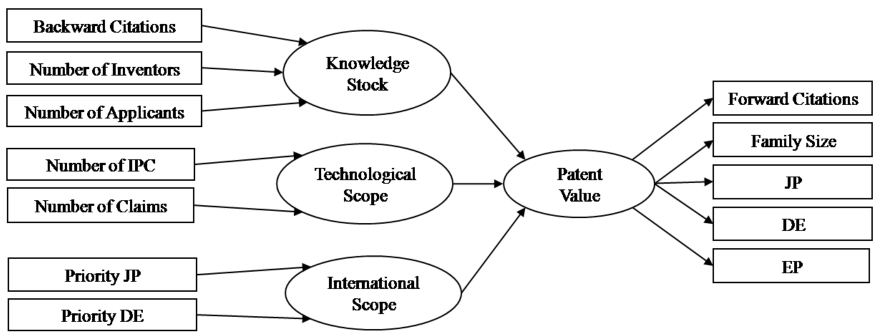


Figure 7.1: First-order model of patent value; patent value is an endogenous latent variable; knowledge stock, technological scope, and international scope are formative exogenous constructs.

forming this construct. The same applies to the backward citations. The previous works, cited in the patent document, are the scientific and technical knowledge units that must exist before the creation of an invention, and they may be used as knowledge inputs within the invention process. Moreover, backward citations represent the prior art, and demonstrate that the invention had not been protected before. These three indicators have been related to the patent value for other authors (see for instance Reitzig (2004a)). However, they still have not been used to estimate an unobserved variable as they are in a structural equation model.

From a theoretical standpoint, the knowledge stock is an exogenous latent variable, and affects the value of a patent. Keeping in mind the backward citations, it seems reasonable to think that an invention that is protected in an area where a lot of inventions are applied –hence with a large knowledge stock– will have less value than a potential radical innovation or a breakthrough invention, and therefore having a smaller knowledge stock. The number of inventors and applicants are revealed first in time, and cause a change on the knowledge stock, and not vice-versa. Additionally, it is not difficult to see that there is no covariance among backward citations, and the number of inventors and applicants. For instance, a patent may contain a large number of references, but the invention may be created only by one inventor or by one applicant. So, a reflective approach would fail to meet the unidimensionality condition. For this construct, however, multicollinearity would not be a problem. Hence, a formative mode is suitable for modelling the relationship between the indicators and the knowledge stock.

The technological scope of the invention is related to the potential utility of an invention in some technological fields. So, the manifest variables for this construct are the number of four-digit IPC classes where the patent is classified, and the number of claims of the patent. The IPC classes allow us to know the technical fields related to the invention, and therefore the number of potential application fields. This does not mean that an invention ultimate use is restricted to a determined area. A company may protect an invention for strategic purposes, for example to prevent its being used by a competitor. Here, the underlying issue is that the larger the number of classification codes, the larger the number of potential application fields, and hence, the greater the technological scope of the patent. On the other hand, and according to Tong & Frame (1994, p. 134), "each claim represents a distinct inventive contribution, so patents are, in effect, bundles of inventions." Claims are a description of what the inventors actually claim to have invented and describe the potential application of the invention. As seen in the literature review, the number of claims should reflect the inventive activity of the invention. So, under the assumption that a highly sophisticated invention will require much inventiveness, the patent will also have a considerable amount of claims. Thus, this variable will also give information about the technological scope of the patents. It is arguable that this is not always so. Probably there are sophisticated inventions that have not required a large number of claims to be protected. But this may be unusual in the renewable energy field. As seen in Table 1 below, the number of claims is a skewed variable (skewness = 4.29, kurtosis = 43.65), with median 14. Following the rules presented before to distinguish between formative and reflective outer models, in this case, the manifest variables are revealed first, and cause a change in the technological scope of the inventions. When defining the manifest variables determining the technological scope. Probably, inventors have an idea of the applicability of the invention long before the time of protecting it. But, it is the patent value, therefore the protected invention, that is being analyzed here. So, a formative relationship is modeled between the indicators and the constructs. Additionally, as with the knowledge stock, there is no collinearity among manifest variables, and the block of variables is not one-dimensional.

The international scope refers to the geographic zones where the invention is protected. Inventions are usually protected in the local country first and then in others, as part of the companies' patenting strategy. All the patents considered in the sample are granted in the U.S. So, we defined two dummy variables that consider whether the invention had been protected in Japan (priority JP) or in Germany (priority DE) during the priority period. Japan and Germany are large producers of renewable energy technologies. Hence, it is interesting to examine whether these variables affect the patent value. Variables indicating whether inventions have been protected through the European Patent Office (EPO) or by the World Intellectual Property Organization (WIPO) have been excluded from the analysis because they provide little information. This means that for the international scope, not all the variables that could form the construct are being considered. So, higher disturbance terms are expected in this case. The international scope is clearly caused by the manifest variables. Here, again there is no collinearity among manifest variables, the block of variables is not one-dimensional. Therefore, formative relationships are considered in this block of variables.

On the other hand, the importance of a patent for future technological developments will be reflected in the number of times that the patent is cited, since the patent is useful for the development of other technologies (Guellec & van Pottelsberghe, 2000), and in the patenting strategy pursued by the company over time. The latter is measured by taking into account the size of the patent family or the number of countries where the protection is sought. For the block of variables of patent value, a reflective relationship is considered between manifest and latent variables. As in this case all the indicators should explain the same construct (aside from the variables that have traditionally been used to infer the patent value), dummy variables are defined by considering whether the patent has been protected in Japan (JP), Germany (DE) or through the European Patent Office (EP). So, in this research, the first analyzed case is a first-order model composed by four constructs: knowledge stock, technological scope, international scope, and patent value (Figure 7.1).

It is worth noting that the first three constructs –knowledge stock, technological scope, and international scope– give an "a priori" value of patents. Thus, the intrinsic characteristics of the patent at the time of its application, along with the patenting strategy of the company in the priority period, may give a preliminary idea of patent value. In contrast, patent value estimated through forward citations and family size gives an "a posteriori" value for patents. This value (recognized value) is obtained over time and is given by others through the number of times that the patent is cited and the number of countries where the protection is sought. Estimating the patent value only through these manifest variables seems too ambitious. Rather, it is reasonable to think that the patent value is jointly given by those variables that determine the "a priori" and the "a posteriori" patent value. Using this approach, the influence of the "a posteriori" relative to the "a priori" patent value may also be assessed. Hence, the indicators that were initially related to the patent value are also associated with a fifth underlying latent variable related to the potential usefulness of the patent. The more useful a patent is, the more it is cited by others and the more important it is to the company's patenting strategy. We call this latent variable "technological usefulness." From a methodological standpoint, this means that the patent value is not directly related to a block of observed variables. So, this construct is regarded as a second-order latent variable that is influenced by all of the other constructs in a second-order model. The proposed model is shown in Figure 7.2. We explore the veracity of the assumptions with PLS Path Modelling.

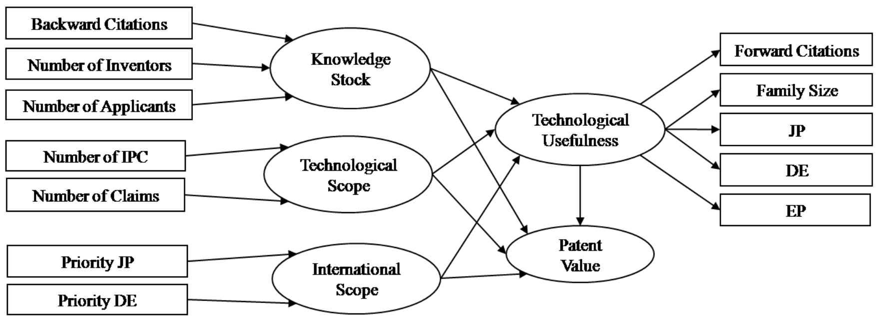


Figure 7.2: Hierarchical component model of patent value; patent value is an endogenous second-order latent variable; technological usefulness is a reflective endogenous latent variable; knowledge stock, technological scope, and international scope are formative exogenous constructs.

7.5 Patent Data

Renewable energy patents include wind, solar, geothermal, wave / tide, biomass, and waste energy. To select suitable patent data, we use the IPC classes for renewable energies listed by Johnstone *et al.* (2007). The sample comprises a total of 2,901 patents (sample 1), published in 1990-1991, 1995-1996, 1999-2000 and 2005-2006, and granted in the U.S. (source: Delphion database). We retrieved these data, and the indicators described above were computed. The number of claims was collected manually for each patent.

Table 7.2 provides descriptive statistics for patent indicators. The results indicate that some variables are very heterogeneous and asymmetric, and they also exhibit large variance. So, normality is not a good assumption. Positive values of skewness indicate positive/right skew (notice how the medians are always smaller than the means). Likewise, positive kurtosis indexes show distributions that are sharper than the normal peak.

Table 7.2: Descriptive statistics of patent data

Manifest Variable	**Mean**	**Standard Deviation**	**Minimum**	**Mediam**	**Maximum**	**Skewness**	**Kurtosis**
Number of applicants	1.04	0.29	1	1	9	12.85	260.81
Number of inventors	2.21	1.58	1	2	14	1.76	4.23
Backward citations	15.36	18.97	0	11	327	5.54	50.79
Number of IPC	6.28	4.52	1	5	48	2.09	7.71
Number of claims	17.02	15.08	1	14	279	4.29	43.65
Priority JP	0.19	0.39	0	0	1	1.54	0.37
Priority DE	0.08	0.27	0	0	1	3.09	7.55
Forward citations	5.63	10.16	0	2	158	5.3	46.83
Family size	8.53	11.62	1	6	202	5.58	51.27
Dummy JP	0.44	0.49	0	0	1	0.23	-1.95
Dummy DE	0.32	0.46	0	0	1	0.75	-1.44
Dummy EP	0.43	0.49	0	0	1	0.25	-1.94

Additionally, the priority countries of these patents are U.S. (59%), Japan (19%), Germany (9%), Great Britain (2%), France (1%) and so on. Patents belong to 1,581 applicants. Patents have been granted to companies (69%), individuals (25%) and universities, research centers or governmental institutions (6%). Due to the manner in which the sample was selected, the sample is homogenous in terms of technological area and the country where the patents were granted. However, the sample is heterogeneous in terms of the type of applicant or the industry in which the companies are classified, and this heterogeneity could affect the results. This also means that there are companies belonging to different industries that are interested in developing renewable energy innovations. At any rate, it is worth noting that at this stage, the patent value model is being tested in general at the level of renewable energy technologies. We estimate the model using the total sample (2,901 patents, sample 1). However, providing that time is an important factor that may affect the findings, three additional samples were taken. Patent indicator matrices were selected in the following application years: 1990-1991 (N = 129, sample 2), 1995-1996 (N = 128, sample 3) and 1999-2000 (N = 536, sample 4). So, in order to analyze whether it is possible to find a pattern in the parameter estimates, the proposed models were estimated with all data, and with time-period data (notice that cases are different in each time-period).

7.6 Results

The internal consistency of reflective outer models, technological usefulness and patent value was assessed by using Cronbach's alpha and composite reliability. For the first-order, the Cronbach's alpha coefficients for patent value are 0.68, 0.79, 0.76 and 0.68 for samples 1, 2, 3 and 4, respectively. Moreover, composite reliability coefficients are 0.77, 0.85, 0.84 and 0.79 for each sample, respectively. So, the patent value is unidimensional. AVE scores are 0.48, 0.56, 0.54 and 0.48 for patent value and for samples 1, 2, 3 and 4, respectively. So, the constructs capture on average more than 50% of the variance in relation to the amount of variance due to measurement error. In the second-order model, technological usefulness has the same Cronbach's alpha and composite reliability coefficients that patent value has in the first-order model. Cronbach's alpha coefficients for the patent value are 0.59, 0.68, 0.7 and 0.58 for samples 1, 2, 3 and 4, respectively. Composite reliability coefficients are 0.72, 0.76, 0.79 and 0.71 for each sample, respectively. Therefore, both technological usefulness and patent value are unidimensional. The technological usefulness captures on average a 54% of the variance in relation to the amount of variance due to measurement error (see the AVE scores for patent value in the first-order model). However, AVE scores for patent value (second-order latent variable) are quite different, 0.24, 0.29, 0.3 and 0.22 for samples 1, 2, 3 and 4, respectively. So, this block of variables is unidimensional, and the latent variable captures on average a 26% of the variance in relation to the amount of variance due to measurement error. This low percentage may be because reflective and formative indicators have been repeated for the second-order latent variable.

Table 7.3 reports the cross loadings for the reflective block of variables in the second-order model of patent value in the three analyzed time-periods. Forward citations, family size and dummy variables JP, DE and EP are slightly more correlated in the three time-periods, with the technological usefulness of the patents rather than the patent value itself. In regards to other indicators, quite the opposite happens: the correlation between indicators and patent value are always higher than the correlation between indicators and technological usefulness. This is adequate even though patent value indicators are used as auxiliary variables in order to estimate the model. It is worth noting that cross loadings of some variables are very similar over time, suggesting a pattern. This phenomenon is interesting because it indicates that the number of inventors; the number of IPC classes; dummy variables JP, DE and EP; forward citations and family size are strongly and constantly correlated with the patent value and its technological usefulness throughout time. This empirical evidence supports the relationships between latent and manifest variables as proposed in the models.

Table 7.3: Cross loadings between indicators for reflective block of variables

	1990-1991		**1995-1996**		**1999-2000**	
Manifest Variable	**Patent Value**	**Technological Usefulness**	**Patent Value**	**Technological Usefulness**	**Patent Value**	**Technological Usefulness**
Number inventors	0.572	0.279	0.611	0.424	0.492	0.135
Backward citations	0.064	0.129	0.092	0.067	0.141	0.091
Number of IPC	0.587	0.387	0.465	0.357	0.495	0.228
Number of claims	-0.074	-0.027	0.403	0.257	0.131	0.048
Dummy priority JP	0.527	0.258	0.391	0.253	0.414	0.162
Dummy priority DE	0.205	0.127	0.103	0.127	0.154	0.136
Forward citations	0.229	0.292	0.295	0.29	0.085	0.085
Family size	0.775	0.894	0.741	0.825	0.714	0.859
Dummy JP	0.816	0.836	0.818	0.833	0.727	0.774
Dummy DE	0.692	0.775	0.754	0.808	0.559	0.681
Dummy EP	0.666	0.818	0.739	0.799	0.658	0.809

Tables 7.4 and 7.5 present the standardized loadings and weights by PLS estimation and t-values by bootstrapping for the first- and second-order models, respectively. Loadings and weights reveal the strength of the relationship between manifest and latent variables. The number of inventors, the number of IPC classes and the dummy priority variables JP and DE are strongly and significantly related to their constructs in all cases in the first- and in the second-order models. Some authors (Cassel *et al.*, 1999; Chin & Newsted, 1999; Vilares *et al.*, 2010) have studied the performance of the PLS Path Modelling algorithm using Monte Carlo

simulations. Among others, the factors analyzed have been the sample size and the number of manifest variables per latent variable. In general, researchers agree and recommend having at least three indicators per construct. However, only (Chin *et al.*, 2003) considered in their study the case of two observed variables per latent variables in their study of interaction effects with reflective outer models. However, as a result of their simulation study, Vilares *et al.* (2010, p. 13) reported that "PLS always produces good estimates for perceived value loadings [a latent variable with two indicators, the author]. This is an interesting result, since PLS is presented as being 'consistent at large' ..." In the formative outer models analyzed here, there are few indicators available per construct. However, the magnitudes of the weights are large enough to infer that there may be a formative relationship between indicators and constructs. Additionally, these results suggest that the patent value and the technological usefulness are evident since the patent is applied. Therefore, the value can be assessed at an early stage. The number of claims shows a weaker association with the technological scope than the number of IPC classes. Perhaps this indicator is more related to the "quality" of the invention, not in the sense of how inventions have an impact on different technological fields (scope) but rather on how important this impact is in a given technological field. Regarding the international scope, this variable seems to be formed by its indicators. The manifest variables are statistically significant in all cases in the two analyzed models. So, this could mean that in the renewable energy field, besides protecting the invention in the U.S., it is important as a value determinant for early protection of the inventions which originate in the other two largest producers of these technologies: Japan and Germany.

On the other hand, patent value and technological usefulness are always strongly and significantly reflected in their explanatory variables. Forward citations, patent family and dummy variables constantly reflect patent value in the first-order model and technological usefulness in the second-order model. The forward citations are not significant in the models evaluated in 1999-2000. But, this may be due to the fact that in recent years patents have been cited less, and the variable is less informative than in previous years. Moreover, loadings for the relationship between forward citations and technological usefulness are smaller than, for instance, loadings for the relationship between family size and technological usefulness. These results may mean that the longitudinal nature of this variable – citations that are received throughout the time– is an important factor that should be taken into account when considering this indicator in the models. The quality of each outer model is measured through the communality index, i.e. the proportion of variance in the measurement variables accounted for by the latent variable. For the second-order model, communality indexes for patent value are 0.29, 0.30 and 0.22 for the 1990-1991, 1995-1996 and 1999-2000 models, respectively. Therefore, indicators have approximately 30% of the variance in common with its latent variable. As seen above, this low percentage may be because reflective and formative indicators have been repeated for the second-order latent variable. The communality indexes for technological usefulness are 0.57, 0.55 and 0.49 for each time-period, also giving evidence of an important percentage of shared variance.

Table 7.4: Standardized loadings and weights for outer models for the first-order model of the patent value, t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level

Construct	**Indicator**	**Sample 1**	**1990-1991**	**1995-1996**	**1999-2000**
Knowledge stock	Backward citations	0.541*	0.420*	0.128	0.499*
		(1.860)	(1.688)	(0.791)	(1.670)
	Number of inventors	0.807**	0.920**	0.988**	0.872**
		(3.054)	(4.937)	(9.086)	(2.794)
Technological scope	Number of IPC	0.966**	0.997**	0.803**	0.985**
		(5.935)	(13.746)	(5.455)	(4.502)
	Number of claims	0.176	-0.058	0.529	0.103**
		(0.756)	(0.364)	(1.432)	(0.354)
International scope	Priority JP	0.802**	0.909**	0.904**	0.847**
		(3.662)	(5.492)	(7.844)	(3.630)
	Priority DE	0.725**	0.512**	0.502**	0.660**
		(2.814)	(2.043)	(2.479)	(2.422)
Patent value	Forward citations	-0.108	0.274**	0.299*	0.096
		(0.940)	(2.041)	(1.693)	(0.524)
	Family size	0.840**	0.893**	0.813**	0.845**
		(9.464)	(36.017)	(15.126)	(5.297)
	Dummy JP	0.777**	0.843**	0.841**	0.802**
		(6.593)	(19.572)	(21.277)	(4.549)
	Dummy DE	0.690**	0.777**	0.811**	0.671**
		(5.530)	(11.126)	(18.389)	(4.087)
	Dummy EP	0.780**	0.808**	0.794**	0.786**
		(7.921)	(11.975)	(12.513)	(5.272)

Tables 7.6 and 7.7 show the findings for the inner relationships (standardized beta coefficients, significance levels and coefficients of determination) for the first- and second-order models respectively. Path coefficient of knowledge stock, technological scope and international scope as related to patent value are significant at 0.01 levels in almost all cases. Therefore, the patent value may be formed by constructs estimated from reliable patent indicators. The first-order model allows us to obtain an estimate of the patent value "in time equal to zero." As showed in the second-order model, the knowledge stock, the technological scope and the international scope are also related to technological usefulness. Moreover, technological usefulness and patent value are significantly related, indicating how the former is an important variable in the prediction of the latter. The second-order model allows us to obtain the patent value as the sum of the value in time equal to zero, and the value given by others, that is the technological usefulness.

The determination coefficient for patent value is 0.9 in the second-order mod-

Table 7.5: Standardized loadings and weights for outer models for the second-order model of the patent value, t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level

Construct	**Indicator**	**Sample 1**	**1990-1991**	**1995-1996**	**1999-2000**
Knowledge stock	Backward citations	0.439	0.248	0.122	0.357
		(1.619)	(1.103)	(0.991)	(1.114)
	Number of inventors	0.871**	0.976**	0.989**	0.938**
		(3.828)	(8.060)	(24.728)	(3.214)
Technological scope	Number of IPC	0.952**	0.995**	0.761**	0.974**
		(6.544)	(18.078)	(4.633)	(4.140)
	Number of claims	0.220	-0.078	0.584**	0.150
		(1.028)	(0.546)	(3.139)	(0.516)
International scope	Priority JP	0.867**	0.931**	0.947**	0.915**
		(4.090)	(10.601)	(7.863)	(4.096)
	Priority DE	0.639**	0.465**	0.401*	0.548*
		(2.422)	(2.709)	(1.701)	(1.943)
Technological usefulness	Forward citations	0.762**	0.836**	0.834**	0.774**
		(6.833)	(22.739)	(24.167)	(5.177)
	Family size	0.795**	0.818**	0.799**	0.809**
		(10.667)	(11.800)	(18.126)	(11.499)
	Dummy JP	0.705**	0.775**	0.809**	0.681**
		(7.983)	(11.891)	(18.318)	(6.256)
	Dummy DE	-0.052	0.292**	0.290**	0.085
		(0.488)	(2.280)	(2.190)	(0.616)
	Dummy EP	0.853**	0.894**	0.825**	0.859**
		(13.577)	(36.226)	(21.104)	(11.526)
Patent value	Backward citations	0.232	0.064	0.092	0.141
		(1.511)	(0.564)	(1.005)	(0.735)
	Number of inventors	0.476**	0.572**	0.611**	0.492**
		(3.477)	(5.964)	(8.825)	(3.016)
	Number of IPC	0.549**	0.587**	0.465**	0.495**
		(5.909)	(7.837)	(4.820)	(3.420)
	Number of claims	0.185	-0.074	0.403**	0.131
		(1.296)	(0.748)	(3.193)	(0.810)
	Priority JP	0.387**	0.527**	0.391**	0.414**
		(2.723)	(5.466)	(3.604)	(2.461)
	Priority DE	0.202**	0.205**	0.103	0.154
		(5.318)	(2.262)	(1.269)	(1.191)
	Forward citations	-0.085	0.229*	0.295**	0.085
		(0.861)	(1.944)	(2.453)	(0.659)
	Family size	0.730**	0.775**	0.741**	0.714**
		(8.250)	(15.351)	(11.612)	(5.952)
	Dummy JP	0.711**	0.816**	0.818**	0.727**
		(6.083)	(20.264)	(18.295)	(4.349)
	Dummy DE	0.586**	0.692**	0.754**	0.559**
		(5.318)	(8.318)	(11.977)	(4.349)
	Dummy EP	0.672**	0.666**	0.739**	0.658**
		(7.196)	(6.650)	(13.752)	(6.341)

els, i.e. the model fit the data in an acceptable way. This result is not surprising; it confirms the aforementioned findings and indicates how the data is better explained by second-order models as compared with first-order models. However, we must consider this result carefully, because the patent value is estimated con-

Table 7.6: Standardized path coefficients for the first-order model of patent value, t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level

Latent Variable	Sample 1	1990-1991	1995-1996	1999-2000
Knowledge stock to Patent value	0.115 (1.248)	0.202* (1.987)	0.306** (2.263)	0.091 (1.040)
Technological scope to Patent value	0.238** (2.892)	0.314** (4.221)	0.335** (3.084)	0.200** (2.278)
International scope to Patent value	0.243** (3.199)	0.154* (1.998)	0.251** (3.044)	0.220** (2.420)
R^2 of patent value	0.161	0.234	0.35	0.114

Table 7.7: Standardized path coefficients for the second-order model of patent value, t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level

Latent Variable	Sample 1	1990-1991	1995-1996	1999-2000
Knowledge stock to Patent value	0.280** (9.979)	0.226** (9.510)	0.229** (12.349)	0.293** (8.281)
Technological scope to Patent value	0.278** (8.811)	0.227** (10.737)	0.226** (8.870)	0.271** (7.620)
International scope to Patent value	0.212** (5.505)	0.232** (11.314)	0.166** (7.659)	0.236** (5.160)
Knowledge stock to Technological usefulness	0.104 (1.162)	0.180* (1.752)	0.299** (3.771)	0.072 (0.783)
Technological scope to Technological usefulness	0.237** (2.686)	0.315** (3.290)	0.334** (3.387)	0.207** (2.133)
International scope to Technological usefulness	0.225** (2.486)	0.142 (1.376)	0.236** (3.042)	0.200** (2.252)
Technological usefulness to Patent value	0.683** (14.511)	0.668** (16.951)	0.697** (20.558)	0.698** (11.207)
R^2 of patent value	0.998	0.998	0.999	0.997
R^2 of usefulness	0.148	0.219	0.338	0.103

sidering all the measurement variables of the models. Another explanation for this is that in the second-order models, the contribution of the recognized value of patents (technological usefulness) is considered, and this would help fit the data better. Unlike patent value, technological usefulness has a moderate coefficient of determination. Perhaps other indicators should help to better explain the model, or again the longitudinal nature of the forward citations is an important factor to be considered. However, we think that the results are acceptable, taking into account the literature review and the goodness of fit obtained using other models in the analysis of patent data. It is worth noting that the structural relationships are significant.

7.7 Final Remarks

This research relates manifest variables that come from information contained in the patent document with latent variables into a single replicable model. The magnitude of this relationship and the importance of each construct are known, including the influence of knowledge stock, the technological and international scope in the value of the technology. In the first-order model, the variables that most affect the patent value are the technological and the international scope. In the second-order model, the technological usefulness is also important.

A distinction between two patent values can be made: an “a priori” and intrinsic value, which the patent has at the moment of its application (the potential value of the patent); and an “a posteriori” value that the patent acquires over time through the actions of a company or others (the value that is recognized). The potential value depends on the characteristics of the patent at the time of application -such as the patenting strategy of a company, the technological applicability of the patents in different technological fields and the base of knowledge that is necessary for the creation of a new invention. As time passes, the patent potentiality is recognized and reflected in the number of times that it is cited and in the number of countries where it is protected. This recognition is a reflection of its technological usefulness. Even though companies can assess the importance or impact of their inventions, these results and the procedure for obtaining them are becoming a tool for improving the strategy of developing new products and inventions, improving intellectual property policy and for comparing technologies with other competitors. The stability of results over time augur that this may be possible.

In order to assess companies’ patent portfolios using a model that can be replicated, a follow-up to this research will study patent value evolution as well as the market-patent relationship and its implications. Furthermore, there are other indicators related to patent value that have been previously studied, but they cannot be computed from the information contained in the patent documents, such as the number of renewals and the number of opposition cases. Nevertheless, these variables could be related to another latent variable in the model, or be a reflection of the technological usefulness of an invention. Finally, PLS Path Modelling has proven to be a suitable approach for analyzing patent data.

Chapter 8

The Longitudinal Nature of Patent Value and Technological Usefulness: Exploring PLS Structural Equation Models

Abstract. The purpose of this chapter is to investigate the evolution of patent value and technological usefulness over time using longitudinal structural equation models. The variables are modeled as endogenous unobservable variables which depend on three exogenous constructs: the knowledge stock used by companies to create their inventions, the technological scope of the inventions and the international scope of protection. Two set-ups are explored. The first longitudinal model includes time-dependent manifest variables and the second includes time-dependent unobservable variables. The structural equation models are estimated using Partial Least Squares Path Modelling. We showed that there is a trade-off between the exogenous latent variables and technological usefulness over time. This means that the former variables become less important and the latter more important as time passes.

8.1 Introduction

In this paper we explore a predictive dynamic model that considers patent value as an unobservable variable. The patent value model has been previously presented as a first and a second-order structural equation model (Martínez-Ruiz & Aluja-Banet, 2008, 2009). The structural equation model (SEM) was based on the theoretical background and an extensive review of the literature. The patent value was modeled as an endogenous unobserved variable depending on the following four exogenous constructs: the knowledge stock used by a firm to create the invention, the technological scope of protection, the international scope of protection and the technological usefulness of the inventions. The latter is also an endogenous latent variable depending on the first three exogenous constructs. Each latent variable was estimated using a set of manifest variables or indicators constructed mainly from the information contained in patent documents.

Now we introduce the dynamic aspects of the model, since patent value has an intrinsic life cycle. This means that over time, the value first increases and then reaches a stage where the size of the increments become smaller and smaller. We attempt to capture this phenomenon and to estimate the evolution of the value over time. We explored two models. The first one considers time-dependent indicators for technological usefulness; this allows us to obtain a global estimation of the patent value as a weighted sum of the manifest variables at different time points. In the second model, technological usefulness is modeled as an unobservable variable which changes over time (time-dependent latent variable); and each of these constructs is estimated by a group of measured manifest variables within the corresponding time period. Both models allow the analysis of loadings and path coefficients over time, but the second also measures the changes of the lagged endogenous latent variable. This approach has been followed by Jöreskog & Wold (1982b) to model longitudinal data in structural equation models, and we use it in a PLS Path Modelling framework.

8.2 Patent Value

Patents are intellectual assets that do not necessarily have an immediate return. A patent may protect a product that can be manufactured and sold. But a patent may also protect technologies which, together with other technologies, enable the manufacture of a final product. In both cases, obtaining an economic value of patents may be extremely difficult. In studying patent value, different approaches have been taken throughout the literature. Some of the approaches focus on the private value of a patent while others concentrate on a patent's social value. Lanjouw *et al.* (1998, p. 407) defined the private value of a patent in terms of "the difference in the returns that would accrue to the innovation with and without patent protection." The magnitude of this difference would be crucial in applying or renewing the protection. Reitzig (2004a) also focused on the private value of patents, and specifies the need to consider the patent value as a construct. Technical experts were surveyed and, according to them, the research showed that the factors that determine patent value are: state of the art (existing technologies), novelty, inventiveness, breadth, difficulty of inventing, disclosure and dependence on complementary assets. Additionally, Trajtenberg (1990b) showed that patent data was highly correlated with some indicators of the social benefits of innovations. Guellec & van Pottelsberghe (2000) presented a value scale proposing that technology increases its own value as it passes through different stages: from invention to application, examination, publication and decision to grant, and finally to the high value stage if the patent is granted. The distinction is made between the intrinsic value of the patent simply for being granted (and thereby having proven novelty, inventive activity and applicability) and the potential value of technology (dependent on its potential for generating future returns). On the other hand, Lanjouw & Schankerman (2004) constructed an index for patent quality, emphasizing "both the technological and value dimensions of an innovation" (p. 443). Using factorial analysis, the researchers model patent quality as an underlying construct that explains a set of patent indicators (forward citations, backward citations, family size and number of claims). The latent variable is computed as a "linear combination of the set of indicators, where weights depend on the factor loadings" (p. 449). One of the main results of this research is that the use of a latent variable model significantly reduces the variability of the construct.

Some patent indicators have been used to directly infer the patent's value, such as forward citations or family size. Even though this may be useful and may give an approximation of the patent value, many elements may affect the invention and protection process. We consider some of these factors based on the background, and represent their interactions proposing a multidimensional analysis of the problem. It is worth noting that this research does not seek to determine the value of an individual patent or to obtain a monetary value of the assets. Rather, the patent value is proposed in terms of the technological usefulness of the inventions. This model, however, allows us to compare and rank the value of a company's patent portfolio. We addressed the question of what variables determine the patent value and how they relate to each other. These variables are modeled as unobserved variables. So, they and their relationships set up a structural model. Little research has reported on the structural relationship among latent variables which influence patent value using a multidimensional approach. The recent investigations of Harhoff *et al.* (2003), Harhoff & Reitzig (2004), Reitzig (2003), and Reitzig (2004a) used a large number of indicators of patent value which were aimed mainly at estimating the probability of opposition to a patent. In most cases, analytical approaches have been based on standard econometric analysis techniques (probit or logit models) or survey analysis. One reason that could explain why a multidimensional and structural approach has not been applied to technology/patent value is that the more general structural models are based on maximum likelihood estimation and the multivariate normal distribution of data. Patent indicators are very heterogenous and asymmetric, and, in general, they exhibit a large variance and skew. Consequently, assuming that this type of data has a multivariate normal distribution may lead to biased results. PLS Path Modelling overcomes this drawback because it is an iterative algorithm that makes no assumptions about data distribution.

8.3 Longitudinal Nature of Patent Indicators

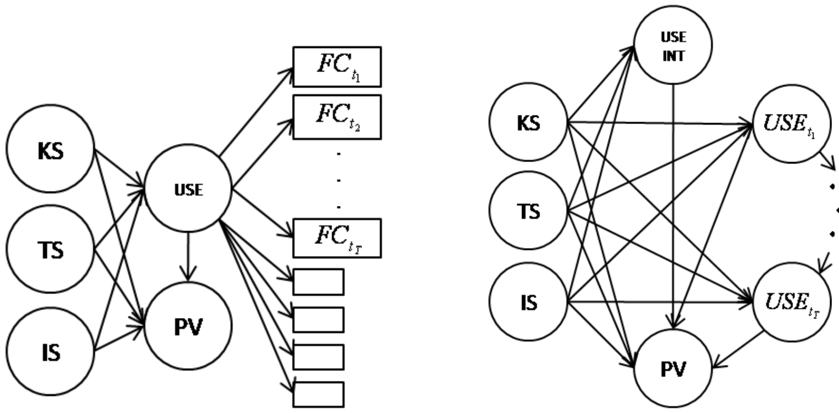
A fundamental feature of longitudinal data is that the same measurement is obtained on different occasions for the same individuals. So, the aim of a longitudinal study is to assess the changes between occasions and explain these changes based on theoretical grounds. It is important to emphasize that the patent indicators described above have a temporary nature. The number of inventors, applicants, cited patents, claims and IPC codes are determined at the time of the patent filing or during the patent examination process. We may assume that they are determined at the instant zero. However, this assumption is not valid for forward citations and family size. Both, the family size and the forward citations, are variables with a longitudinal character. Usually, the companies first protect their inventions in their local countries and then in others within a period of time. So, the family size is an indicator that may change over time. Harhoff *et al.* (2003, p. 1360) said that this indicator "may be available around the time of application."

In addition, it is known that the number of forward citations is an indicator that may vary over time, since a patent may receive citations over a long period. As a first attempt –and given the complexity of recovering some data– we retrieve the number of yearly citations received for patents belonging to the sample. This allows us to assess the implications on the results by considering the longitudinal nature of the data in the estimated models. We are also aware that longitudinal data have an intrinsic autoregressive nature. So, in this way, we also explore the robustness of the proposed structural models.

8.4 Patent Value Models for Longitudinal Data

In order to model the patent value over time, we explore two longitudinal models (see Figure 8.1). The first longitudinal model A (Figure 8.1(a)) considers three exogenous constructs –knowledge stock (KS), technological scope (TS) and international scope (IS)– at time point zero. The measurement variables of these constructs are the indicators described in Chapter 7[1]. The endogenous latent variable, technological usefulness (USE), is measured by time-dependent manifest variables. The indicators of this variable are the number of forward citations

[1]Number of inventors, number of applicants and backward citations for KS; the number of claims and the number of IPC codes for TS; two dummy variables, priority JP and priority DE, for IS.



(a) Model A: Time-dependent manifest variables

(b) Model B: Time-dependent latent variables

Figure 8.1: Patent value structural models for longitudinal data. In model A, the forward citations measured at different time points (FC_{t_i}), the family size and the dummy variables (JP, DE, EP) are manifest variables of the latent variable, technological usefulness. In model B, technological usefulness is a time-dependent latent variable, each measured by a set of indicators.

per year. So, by capturing the longitudinal nature of forward citations through an unobservable variable, we "average" the contribution of the longitudinal indicators. For USE, we also consider the previously defined dummy variables (JP, DE and EPO). The patent value (PV) is modeled as an endogenous latent variable formed by the weighted contribution of knowledge stock, technological scope, international scope and the technological usefulness. Hence, this model gives an overall measure of the patent value.

The second longitudinal model B (Figure 8.1(b)) considers: (1) the same aforementioned exogenous constructs, (2) an auxiliary endogenous construct[2] (USE-INT) clustering the family size and the dummy variables JP, DE and EPO, and (3) the technological usefulness (USE) as a set of time-dependent latent variables, each one measured by blocks of observed variables at different time points. We modeled seven different time periods: 1992-1993, 1994-1995, 1996-1997, 1998-1999, 2000-2001, 2002-2003, and 2004-2005. In model B, the patent value (PV) is also formed by the weighted sum of all the constructs and latent variables, but now the model allows for the analysis of changes in the technological usefulness over different time periods.

[2]This latent variable groups the measures related to family size and the international protection of patents. Initially, when the model did not consider variations in time, the variables were considered as indicators of technological usefulness. However, it is now necessary to rethink this formulation.

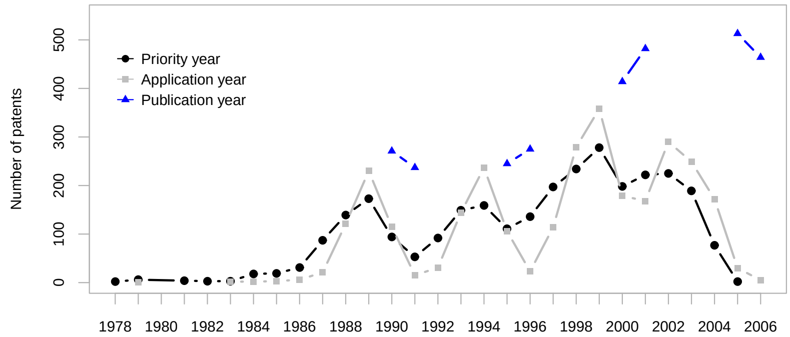


Figure 8.2: Number of patents by priority, application and publication year

8.5 The Patent Sample

To estimate the proposed models, we retrieved a random sample of patent data from the Delphion database (N=2,901). We established some criteria for retrieving data. We used the International Patent Classification (IPC) codes for renewable energies listed by Johnstone *et al.* (2007). Hence, patents are classified in codes related to wind, solar, geothermal, wave/tide, biomass and waste energies. All patents were granted in the U.S. and were published in 1990-1991, 1995-1996, 1999-2000, and 2005-2006. We chose these time periods arbitrarily. Since "whenever possible, the application date should be used as the relevant time placer for patents" (Hall *et al.*, 2001, p. 10), we reorganized the data by application year and the indicators were computed. Figure 8.2 shows the number of patents by priority, application and publication year. Additionally, the priority countries of these patents are U.S. (59%), Japan (19%), Germany (9%), Great Britain (2%), France (1%) and so on. Patents belong to 1,581 applicants. Patents have been granted to companies (69%), individuals (25%) and universities, research centers or governmental institutions (6%).

In order to analyze whether it is possible to find a pattern in the parameter estimates, the proposed models were estimated with time-period data. We arbitrarily chose three patent indicator matrices: (1) a set of 359 patents applied for in the years 1989-1990-1991 (sample 1), (2) a set of 129 patents applied for in the years 1995-1996 (sample 2), and (3) a set of 179 patents applied for in 2000 (sample 3). Applicants of these patents are companies. According to the Delphion database, these data sets represent 41.74%, 35.15%, and 51.29% of all patents applied for in the U.S. in the field of renewable energy during the selected time periods, respectively. However, we do not known this percentage in relation to

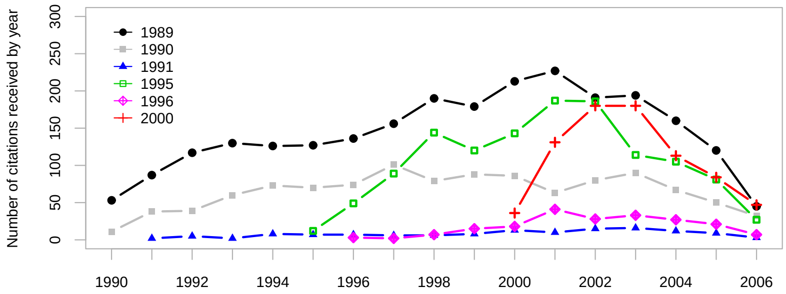


Figure 8.3: Number of citations received by year, patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000

patents applied for by companies. Due to the manner in which the sample was selected, the sample is homogenous in terms of technological area and the country where the patents were granted. However, the sample is heterogeneous in terms of the type of industry in which the companies are classified, and this heterogeneity could affect the results. This also means that there are companies belonging to different industries that are interested in developing renewable energy innovations. At any rate, it is worth noting that at this stage, the patent value model is being tested in general at the level of renewable energy technologies.

Table 8.4 in section 8.9 provides descriptive statistics for patent indicators for each patent data set. The results indicate that some variables are very heterogeneous and asymmetric, and they also exhibit large variance. So, normality is not a recommended assumption. Positive values of skewness indicate positive/right skew (notice how the medians are always smaller than the means). Likewise, positive kurtosis indexes show distributions that are sharper than the normal peak.

All forward citations received by patents per year were retrieved from the United States Patent and Trademark Office (USPTO) database from 1992 to 2005. Figures 8.3 and 8.4 show the number of citations received by year and the accumulated citations received by year, respectively, for patent applications in 1989, 1990, 1991, 1995, 1996, and 2000. These figures show an increase in the number of citations over time. Figure 8.3 shows that the number of citations reaches a peak then decreases. The patents applied for in 1989 are the most cited. The patents less cited are those applied for in 1991 and 1996.

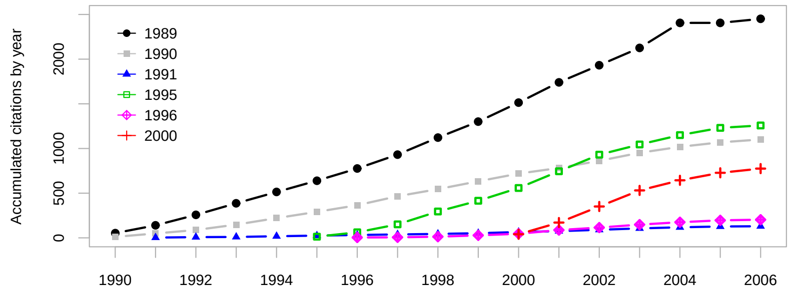


Figure 8.4: Accumulated citations received by year, patents applied for in 1989, 1990, 1991, 1995, 1996 and 2000

8.6 PLS Path Modelling for Longitudinal Data

PLS Path Modelling is a component-based procedure for estimating a sequence of latent variables developed by the statistician and econometrician Herman Wold. During the last few years, it has proved to be useful for estimating structural models, in marketing and information systems research in particular, and in the social sciences in general. Some of its features have encouraged its use, such as: (1) it is an iterative algorithm that offers an explicit estimation of the latent variables and their relationships, (2) it works with fewer cases and makes no assumptions about data distribution, and (3) it overcomes the identification problems when formative measurement models are included. Wold's basic-design of PLS Path Modelling does not consider higher-order latent variables. Therefore, in Wold's algorithm each construct must be related to a set of observed variables in order to be estimated. However, Lohmöller (1989) proposed a procedure for the case of hierarchical constructs; that is to say, for cases where there is a construct with no block of manifest variables, or more simply: it is only related to other constructs. In hierarchical component modelling, manifest variables of first-order latent variables are repeated for the second-order latent variable. So, a set of "auxiliary" variables is introduced for estimation purposes. After that, the model is estimated using PLS Path Modelling in the usual way.

Estimations of structural equation models with longitudinal data –a widespread practice in econometrics– have traditionally been made with simultaneous equation models (Kmenta, 1986; Aigner *et al.*, 1987) and LISREL (Jöreskog & Wold, 1982a). From the point of view of Partial Least Squares (PLS) Path Modelling (Wold, 1982), there are still few theoretical developments and applications. Jan Lohmöller and Herman Wold (Lohmöller, 1989) were the first to address the issue in a PLS framework. The researchers proposed a procedure called Latent Variables Three-mode Path Analysis (LVP3), which is based on three pivotal points: a three-way or three-mode data set[3], a three-mode factor model (FA3 model) as outer model, and a three-mode path model as inner model. In LVP3, loadings and path matrices are specified as Kronecker matrices, so the LVP algorithm includes additional steps in the iterative stage to satisfy the implications of the Kronecker structures (Lohmöller, 1989) and calculate these matrices. Lohmöller (1989) presented this algorithm and warned that the convergence properties and the reliability of the numerical results were still to be explored. The main advantage of the procedure is that "the path model for the time points assumes a second-order autoregression process, i.e. each time point t is regressed to the one directly preceding it, $t-1$, and to the time point before that $t-2$" (Lohmöller, 1989, p. 236).

In a more traditional way, structural equation models with longitudinal data consider the repetition of the structural and measurement models in each of the years under study. Therefore, when the model is tested, the whole model is estimated at the same time. So, both background variables and initial measures, as well as the final status are included in the model. Wold (1982) and Scepi & Esposito Vinzi (2003) followed this approach, but the latter also introduced multi-table analysis with an aim toward identifying temporal components in the data structure. We addressed the more traditional approach in order to estimate the longitudinal patent value models. As a PLS Path Modelling procedure is used, the aforementioned autoregressive process is also implemented.

8.7 Results

Model assessment. We first assess the internal consistency of reflective outer models by using Cronbach's alpha coefficient (should be > 0.7). All reflective measurement models are unidimensional. Cronbach's alpha coefficients of technological usefulness are 0.91, 0.94 and 0.82 for models A with samples 1, 2, and 3, respectively. Cronbach's alpha coefficient for the auxiliary latent variable in model B is 0.80 and Cronbach's alpha range is from 0.94 to 0.99 for the different time points of technological usefulness. We computed the average variance extracted (AVE) to assess the extent to which measures of a given construct differ from measures of other constructs (discriminant validity). As suggested by Fornell & Larcker (1981), the percentage of variance captured by the construct in relation to the variance due to random measurement errors should be greater than 0.5. The AVE of technological usefulness is 0.58, 0.65, and 0.53 for samples 1, 2, and 3, respectively, in models with time-dependent manifest variables. Thus, the latent

[3]This terminology was introduced by Tucker (1963, 1964).

variable is capturing on average more than 50% of the variance in relation to the amount of variance due to measurement error. For the model with time-dependent latent variables, the AVE of technological usefulness ranges from 0.95 to 0.99 at the different time points.

Tables 8.5 and 8.6 in section 8.9 report the cross loadings –or correlations between manifest variables and constructs– for a reflective block of variables in models A and B in the three analyzed time-periods. As shown, each observed variable is correlated more with its corresponding construct. Thus, for instance, in the longitudinal model with time-dependent latent variables, the family size and the dummy variables JP and DE are more related with the auxiliary variable USE-Int than with the technological usefulness in the different time-periods. This empirical evidence supports the relationships between latent and manifest variables as proposed in the models.

Loading estimates for reflective models A and B are reported in Tables 8.7 and 8.8 in section 8.9, respectively. Loadings indicate how much variance each indicator shares with the latent variable (reliability). A rule of thumb generally accepted is 0.7 or more (Hulland, 1999). A low value in a loading factor suggests that the indicator has little relation to the associated construct. As shown in the tables, all loading estimates are significant at the 0.01 level[4]. For models A, the loadings of time-dependent manifest variables, that is, the forward citations for the different time periods, range from 0.776 to 0.955. Thus, the technological usefulness is reflected in reliable time-dependent indicators, and the latent variable explains the correlations among the manifest variables. Although significant, the loadings for the family size and the dummy variables (Germany and Japan) are less than 0.7. This situation changes when the time-dependent latent variable model is considered. For model B, loadings are always greater than 0.7. In this case, the family size and the dummy variables (DE and JP) are reliable indicators of the auxiliary latent variable USE-Int, and the construct explains the correlation among the indicators. We have not given a definitive name to this latent variable. "International patenting strategy" would describe the concept formed by the family size and the dummy variables JP and DE. These indicators measure whether the inventions have been protected internationally, particularly in the major countries which produce renewable energy technologies. Hence, these variables may change over time. This is what makes the construct different in regards to an international scope.

The reliability of formative outer models –knowledge stock, technological scope, international scope– was assessed by examining weight estimates and the correlations between the constructs and their corresponding manifest variables. Manifest

[4]The t-values were computed by bootstrapping with 200 bootstrap resamples; t-value > 1.65 significant at the 0.05 level; t-value > 2 significant at the 0.01 level.

variables in formative measurement models do not have to be intercorrelated. In fact, Pearson correlations between patent indicators are small and medium,[5] ranging from 0.04 to 0.25[6]. However, the indicators should be correlated with the constructs which are related, because the manifest variables are supposed to contribute to the formation of the unobservable variable. Tables 8.5 and 8.6 show the correlations between knowledge stock, technological scope and international scope, along with their corresponding manifest variables (models A and B, and samples 1, 2, and 3). Tables 8.7 and 8.8 in section 8.9 show the estimates for outer relationships. For models A, of which there are three, the weight estimates are in line with the correlations between constructs and indicators. For samples 1 and 2, there are medium and large correlations between the knowledge stock and its indicators. The weight estimates are positive, and the number of inventors indicates a significant relationship with the construct. The same happens with the technological scope and international scope. For model A with sample 3, the weight estimates are negative as well as the correlations between the constructs and their corresponding indicators. Thus, weight estimates and correlations between each formative construct and its corresponding indicators vary in the same way, validating the formative constructs. Negative correlations are attributed to the fact that the sample corresponds to patents applied for in 2000. As discussed below, forward citations influence the model estimates in a meaningful way. These variables are less informative for sample 3, affecting the estimation of the unobservable variables for that year. The results for model B suggest that a model with time-dependent latent variables may reveal significant relationships in the formative outer models. The number of IPC codes, the number of claims, and the dummy priority variables JP and DE are strongly and significantly related to their constructs. The same was found in Martínez-Ruiz & Aluja-Banet (2009). The relationship between knowledge stock and the number of inventors is also significant.

Since multicollinearity is a problem in multiple regression –and the basic design of PLS Path Modelling uses multiple regression to estimate inner relationships– we calculated the correlations between the estimated constructs and the variance inflation factor[7] so as to perform a collinearity diagnostic. The variance inflation factors for the regression coefficients of the technological usefulness range from 5.18 to 89.88 from 1992-1993 to 2002-2003, respectively. Moreover, we calculated the

[5]Cohen (1988) suggests that correlations of 0.1, 0.3, and 0.5 express small, medium and large effect sizes, respectively.

[6]Correlations between the number of inventors, backward citations, the number of IPC codes, the number of claims, and the dummy variables.

[7]The square root of the variance inflation factor tells you how much large the standar error is, compared with what it would be if that variable were uncorrelated with the other independent variables in the equation. A common rule of thumb is that if VIP(regression coefficient) > 5 then multicollinearity is high.

mean communalities to test for the discriminant validity of unobservable variables. Tables 8.9 and 8.10 in section 8.9 report the results. The mean communalities of each construct are larger than the correlations between the construct and other unobservable variables (models A and B, samples 1, 2, and 3). So, the constructs share more variance with its block of indicators than with another construct representing a different block of manifest variables. In addition, there was no evidence of collinearity between knowledge stock, technological scope, and international scope, nor between these constructs and the auxiliary latent variable USE-Int and technological usefulness. However, and as expected, technological usefulness for the different time periods is highly correlated. Therefore, we estimated the inner relationships by using multiple regression and PLS regression. The latter is recommended to avoid multicollinearity problems among inputs.

Multiple regression to estimate structural relationships. Once the latent variables were obtained with PLS Path Modelling, we estimated the structural relationships by using multiple regression in the usual way. Tables 8.1 and 8.2 show the standardized path coefficients of the longitudinal model with time-dependent manifest variables for the three-analyzed samples, and with time-dependent latent variables for sample 1, respectively. These tables also report the significance of each estimate.

The relationships between patent value and the exogenous constructs are all significant at the 0.01 level (models A and B, samples 1, 2, and 3). The magnitude of the regression coefficients and the t statistic reveal the contribution of each variable to the patent value. Path coefficients of model B show how technological usefulness is reflected over time ($\beta_{92-93} = 0.131$ $t - value = 23.16$,...,$\beta_{04-05} = 0.145$ $t - value = 19.73$) while the regression coefficient in model A ($\beta_{sample3} = 0.856$, $t - value = 18.53$) is "averaging" the contribution of the change in forward citation over time. In addition, when considering forward citations as longitudinal manifest variables of technological usefulness, the regression coefficient between technological scope and technological usefulness is 0.049 ($t - value = 2.28$, sample 1). However, this value changes when the technological usefulness is modeled as a time-dependent latent variable. In model B, the regression coefficient between technological scope and USE 92-93 is 0.210 ($t - value = 2.08$). This relationship is smaller in subsequent years. So, this result suggests that the relative contribution of the technological scope of protection –determined when the invention is classified in some IPC codes and the number of claims– is larger in the first stage of the life cycle of patents, and then it declines. The knowledge stock and the international scope also appear to add more to the patent value in an early stage. The inner relationship between the auxiliary latent variable USE-Int and the patent value is also significant.

Table 8.1: Standardized path coefficients of the A-structural model (longitudinal model with time-dependent manifest variables) for samples 1, 2, and 3; t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level.

	Sample 1: 1989-1990-1991		Sample 2: 1995-1996		Sample 3: 2000	
Construct	**Technological Usefulness**	**Patent Value**	**Technological Usefulness**	**Patent Value**	**Technological Usefulness**	**Patent Value**
Knowledge stock	0.140	0.146	0.239	0.121	0.193	0.192
	(1.4605)	(4.952**)	(2.299**)	(7.638**)	(1.860*)	(3.148**)
Technological scope	0.049	0.133	0.214	0.117	0.041	0.201
	(2.283**)	(4.944**)	(2.721**)	(7.494**)	(0.554)	(3.390**)
International scope	0.251	0.131	0.106	0.0779	0.169	0.166
	(0.637)	(4.267**)	(0.908)	(4.491**)	(1.921*)	(2.346**)
Technological usefulness		0.888		0.897		0.856
		(26.851**)		(52.024**)		(18.538**)

Table 8.2: Standardized path coefficients of the B-structural model (longitudinal model with time-dependent latent variables) for sample 1; t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level.

	Sample 1: 1989-1990-1991								
Construct	**USE-Int**	**USE 92-93**	**USE 94-95**	**USE 96-97**	**USE 98-99**	**USE 00-01**	**USE 02-03**	**USE 04-05**	**PV**
Knowledge stock	0.095	0.111	0.050	0.001	-0.040	0.003	-0.002	-0.005	0.028
	(1.110)	(1.303)	(1.348)	(0.032)	(1.328)	(0.155)	(0.151)	(0.387)	(3.710**)
Technological scope	0.186	0.210	0.022	0.054	-0.020	-0.029	-0.007	-0.011	0.034
	(1.990*)	(2.082**)	(0.607)	(1.396)	(0.758)	(1.289)	(0.427)	(0.737)	(4.017**)
International scope	0.340	-0.022	-0.031	0.032	0.007	-0.018	-0.019	-0.019	0.018
	(2.989**)	(0.320)	(1.071)	(0.750)	(0.189)	(0.912)	(1.147)	(1.376)	(2.390**)
Auxiliary construct USE-Int									0.069
									(3.612**)
Technol. Usefulness 92-93			0.883						0.131
			(35.097**)						(23.167**)
Technol. Usefulness 94-95				0.911					0.146
				(32.888**)					(25.194**)
Technol. Usefulness 96-97					0.952				0.160
					(51.654**)				(21.377**)
Technol. Usefulness 98-99						0.968			0.154
						(68.869**)			(14.214**)
Technol. Usefulness 00-01							0.980		0.163
							(93.416**)		(14.649**)
Technol. Usefulness 02-03								0.985	0.154
								(118.539**)	(15.350**)
Technol. Usefulness 04-05									0.145
									(19.738**)

These results are similar to those obtained in Martínez-Ruiz & Aluja-Banet (2009). However, the effects captured by the structural model are smaller, mainly among the formative constructs and the technological usefulness. This may be due to the fact that considering the longitudinal nature of forward citations helps to reveal the relative weight that this variable has on the estimate of patent value. Figure 8.5 shows the evolution of loadings, which describes the relationship between forward citations and patent value for model A and samples 1, 2 and 3. The Figure clearly shows how the patent value increases, stabilizes and then decreases over time. Figure 8.6 shows the evolution of the standardized path coefficient, which describes the relationship between knowledge stock, technological scope and international scope, and technological usefulness, as well as between the latter and

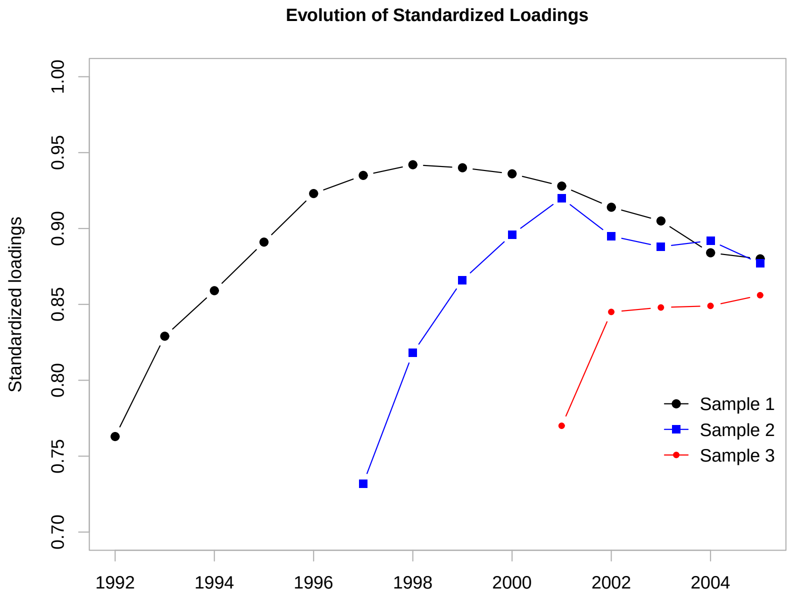


Figure 8.5: Evolution of standardized loadings of the longitudinal model with time-dependent manifest variables (model A) and samples 1, 2 and 3. The loadings describe the relationships between forward citations and patent value.

the patent value for model B. The results suggest a trade-off between exogenous constructs and technological usefulness. This means that the knowledge stock, the technological scope and the international scope contribute more to the patent value at time equal to zero while the technological usefulness determines the patent value in subsequent time-periods.

Finally, the determination coefficients R^2 for technological usefulness in models A are 0.09, 0.15, and 0.06 with samples 1, 2, and 3, respectively. The R^2 in model B is close to 0.8 for technological usefulness and 0.17 for the auxiliary latent variable. So, this suggests that the data is better explained by a longitudinal model with time-dependent latent variables.

PLS regression to estimate structural relationships. Since there is multicollinearity between the technological usefulness for the different time-periods, we also estimated the structural relationships by using PLS regression. The number of significant components t_h were determined by leave-one-out cross validation. The marginal contribution of each PLS component t_h to the predictive power of the regression model was estimated using the Q^2_h index and redundancies[8].

[8]For each h-component, the Q^2_h index is defined as $Q^2_h = 1 - \sum_{k=1}^{q} PRESS_{kh} / \sum_{k=1}^{q} RSS_{k(h-1)}$, where PRESS is the Predicted REsidual Sum of Squares, and RSS is the Residual Sum of Squares of the latent variable Y_k when the regression model

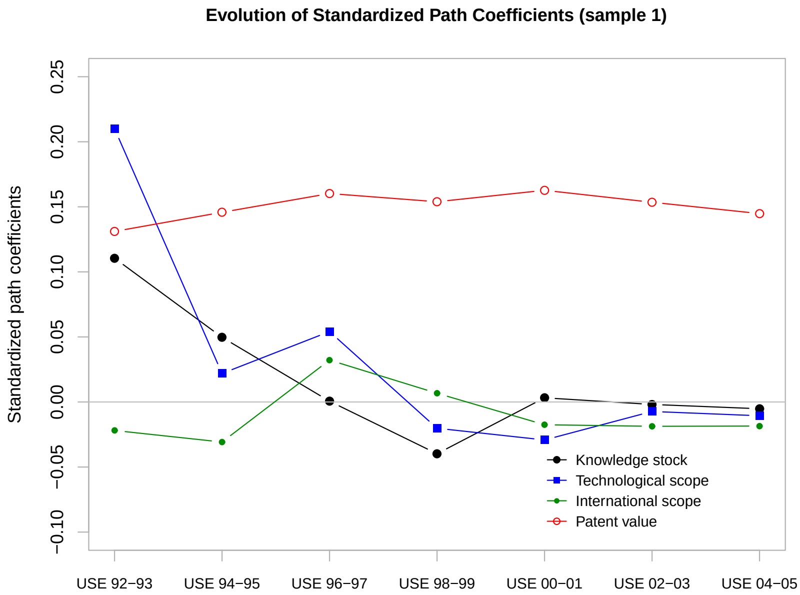


Figure 8.6: Evolution of standardized path coefficients of the longitudinal model with time-dependent latent variables (model B) and sample 1 (1989-1990-1991). The path coefficients describe the relationships between knowledge stock, technological scope and international scope, and the technological usefulness (USE), and between the latter and the patent value.

The patent value was regressed on the knowledge stock, the technological scope, the international scope, the technological usefulness for the different time periods and the auxiliary variable USE-Int, a total of 11 regressors. By default, the PLS regression holds as many components as there are independent variables in the model. For patent value PLS regression, however, two components can predict about 68% of the variation of the regressors (see Table 8.11 in section 8.9). Thus, a two-component model would be sufficient to describe the patent value in terms of the exogenous latent variables and constructs. Nonetheless, we report the results for models with one, two, three and four components. Figures 8.7 shows the correlations between latent variables and the first four PLS components. As shown in Figure 8.7(a) for instance, the patent value, the technological usefulness for the different time periods and the technological scope are highly correlated with the first component whereas the international scope and the auxiliary latent variables USE-int are correlated more with the second component. The knowledge stock

is estimated considering $h-1$ components. The rule is to retain the h-component when $Q_h^2 \geq 0.0975$. The redundancy coefficient measures the amount of explained variance in the indicators for the endogenous construct, explained by the set of manifest variables of the exogenous constructs. It is defined as, $Rd(Y,t_h) = \frac{1}{q}\sum_{k=1}^{q} cor^2(y_k,t_h)$, where q is the number of endogenous variables.

Table 8.3: PLS-regression coefficients for 1-component, 2-component and 3-component models, and variable importance in the projection (VIP index) for model B and sample 1.

	PLS Regression Path Coefficients			**Variable Importance in the Projection (VIP)**		
Construct	1 Comp	2 comp.	3 Comp.	1 Comp	2 comp.	3 Comp.
Knowledge stock	0.029	0.033	0.027	0.241	0.241	0.242
Technological scope	0.044	0.040	0.034	0.364	0.364	0.364
International scope	0.007	0.023	0.020	0.059	0.084	0.085
Auxiliary latent variable USE-Int	0.040	0.062	0.067	0.324	0.335	0.336
Technological usefulness 92-93	0.133	0.132	0.132	1.092	1.092	1.092
Technological usefulness 94-95	0.147	0.145	0.145	1.201	1.200	1.200
Technological usefulness 96-97	0.157	0.156	0.156	1.283	1.282	1.282
Technological usefulness 98-99	0.159	0.159	0.159	1.302	1.301	1.301
Technological usefulness 00-01	0.157	0.157	0.158	1.288	1.287	1.287
Technological usefulness 02-03	0.154	0.153	0.154	1.257	1.256	1.256
Technological usefulness 04-05	0.149	0.148	0.148	1.219	1.219	1.219

and also the technological scope are correlated with the third component. The forth component helps to explain the auxiliary latent variable USE-Int and the international scope.

Table 8.3 shows the PLS-regression coefficients considering one, two and three component models and the variable importance in the projection (VIP index) for model B and sample 1[9]. As shown, and according to the results of the 2-components PLS model, the regression coefficients are very similar to those obtained with multiple regression. However, the technological usefulness in the different time points is the variable that most contributes to the prediction of the patent value.

8.8 Final Remarks

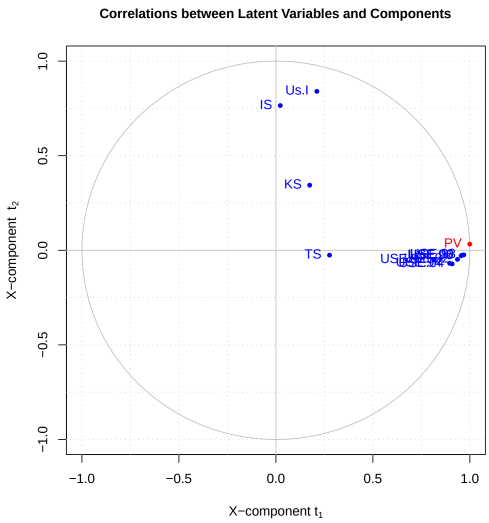
It seems reasonable to think that if a company has invested a lot of knowledge in the creation of an invention, this invention will tend to have a larger value. In the same way, a technology with multiple potential applications would be more valuable than one that can only be applied in a more limited area. The same applies for the international scope of protection. An invention with broader protection is presumably more valuable than one without it. The estimation results of the patent value models for longitudinal data suggest that the contribution of the knowledge stock used by companies to create their inventions, the technological scope of the inventions and the international scope of protection are variables that

[9]The VIP index reflects the influence of the explanatory variables in the h-component model. For a jth independent variable, the $VIP_{hj} = \sqrt{(\frac{p}{Rd(Y;t_1,\ldots,t_h)} \sum_{l=1}^{h} Rd(Y;t_l)w_{lj}^2)}$. The contribution of variable j to the construction of the component t_l is measured by the weights w_{lj}^2. The variables that have larger VIP (> 1) are more important for predicting the dependent variable.

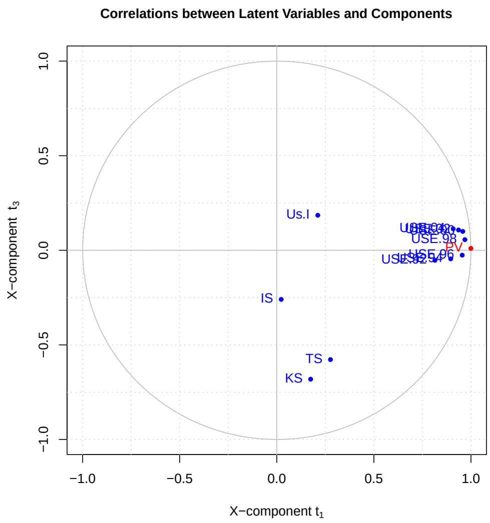
contribute little to the patent value when compared to the technological usefulness. As expected, this is more evident in model B than in model A. Based on the PLS regression findings, for the model with time-dependent manifest variables (model A), about 56.97% of the patent value is explained by the technological usefulness of patents and 43.03% by the other exogenous constructs (sample 1), whereas in the model with time-dependent latent variables (model B), these percentages are 89.40% and 10.60%, respectively.

Forward citations are the most widely used measures for assessing the importance or value of patents. This variable can be used as an indicator for a construct whose contribution to the patent value can be weighted –as was done in the case of technological usefulness. However, the value provided by this indicator is a later value or *a posteriori* value, which can be estimated once time has elapsed; and this could be too late if technology decisions must be made immediately. The benefit of using longitudinal models is that considering the time factor and the longitudinal nature of forward citations help to reveal when each of the construct and latent variables is important. Hence, exogenous constructs may be good indicators of value in the first stage of the life cycle of patents.

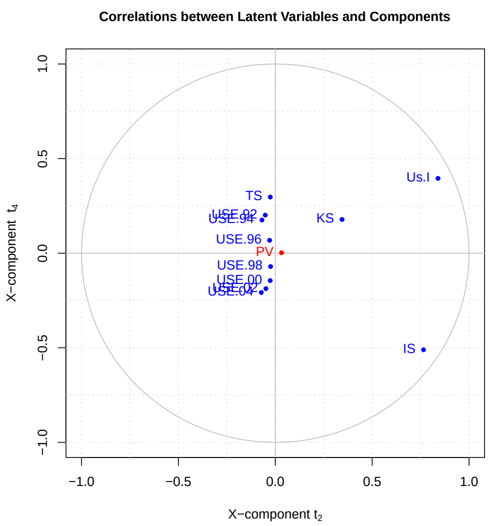
From a statistical standpoint, there are some aspects that have to be considered. First, these models were estimated with small samples. PLS Path Modelling is known for its ability to build a set of unobservable variables and estimate the structural relationships between them when small samples are available (Chin & Newsted, 1999; Tenenhaus & Hanafi, 2010). However, estimating the models with a larger sample, or even considering the population, would help to confirm the exploratory results presented here. In addition, it is well known that for consistency at large, PLS Path Modelling requires three or more indicators per construct –for reflective outer models at least. Simulation studies support this claim (Chin & Newsted, 1999). However, recent investigations have shown that the estimates of formative relationships with few indicators are fairly robust. On the other hand, considering longitudinal data requires caution when assessing results. As expected, the forward citations per year are highly correlated indicators; that is, the value of the variable at time t_i will influence the value of the variable at time t_{i+1}. This may not affect the estimates of the relationship in the outer model –because these models are modeled in a reflective mode– but this can affect the stability of the estimates of the structural relationships. This problem is solved using PLS regression instead multiple regression in the second stage of the PLS Path Modelling procedure.



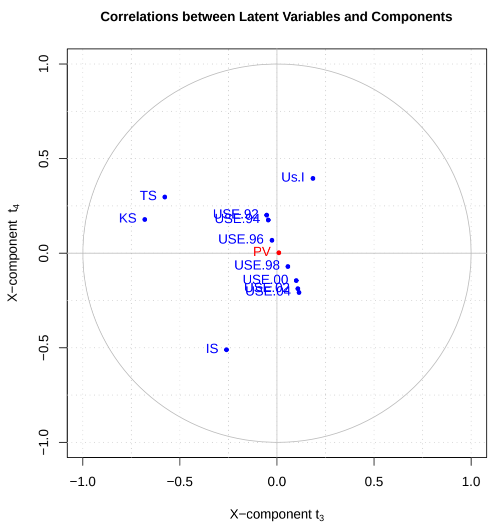
(a) X-components t_1 and t_2



(b) X-components t_1 and t_3



(c) X-components t_2 and t_4



(d) X-components t_3 and t_4

Figure 8.7: Correlations between latent variables and PLS components. The patent value (PV) is regressed on the knowledge stock (KS), the technological scope (TS), the international scope (IS), the technological usefulness (USE) for the different time periods and the auxiliary variable USE-Int (USE-Int).

8.9 Appendix: Tables

Table 8.4: Descriptive statistics of patent data.

	Sample 1: 1989-1990-1991							Sample 2: 1995-1996							Sample 3: 2000						
Manifest Variable	**Mean**	**Standard Deviation**	**Minimum**	**Mediam**	**Maximum**	**Skewness**	**Kurtosis**	**Mean**	**Standard Deviation**	**Minimum**	**Mediam**	**Maximum**	**Skewness**	**Kurtosis**	**Mean**	**Standard Deviation**	**Minimum**	**Mediam**	**Maximum**	**Skewness**	**Kurtosis**
Number of applicants	1.02	0.13	1	1	2	7.57	55.64	1.07	0.47	1	1	6	9.35	95.89	1.06	0.25	1	1	3	4.90	25.98
Number of inventors	1.74	1.17	1	1	8	2.11	5.78	2.04	1.35	1	1	6	1.09	0.04	2.26	1.64	1	2	9	1.62	2.75
Backward citations	8.44	6.25	1	7	32	1.51	2.35	12.58	11.71	1	10	89	3.42	17.05	14.10	13.06	1	11	88	3.09	12.42
Number of IPC	5.12	3.13	1	4	22	1.47	3.34	5.00	3.20	1	4	19	1.89	4.72	6.21	4.45	1	5	21	1.17	0.87
Number of claims	12.78	9.89	1	10	79	2.25	8.64	18.29	15.38	1	16	87	2.18	5.89	16.60	18.03	1	12	156	3.95	22.80
Priority JP	0.17	0.38	0	0	1	1.74	1.03	0.16	0.36	0	0	1	1.93	1.75	0.20	0.40	0	0	1	1.55	0.40
Priority DE	0.06	0.25	0	0	1	3.58	10.84	0.03	0.17	0	0	1	5.48	28.42	0.09	0.29	0	0	1	2.90	6.50
Forward citations	9.29	11.53	0	6	81	3.12	12.79	10.40	17.55	0	6	158	5.84	42.89	3.52	4.91	0	2	42	3.37	20.56
Family size	6.15	6.77	1	3	52	2.32	8.63	8.47	14.87	1	4	120	5.06	30.87	8.71	9.79	1	7	82	3.91	22.81
Dummy EP	0.30	0.46	0	0	1	0.87	-1.25	0.41	0.49	0	0	1	0.37	-1.90	0.51	0.50	0	1	1	-0.06	-2.02
Dummy JP	0.41	0.49	0	0	1	0.36	-1.88	0.41	0.49	0	0	1	0.37	-1.90	0.49	0.50	0	0	1	0.03	-2.02
Dummy DE	0.27	0.44	0	0	1	1.04	-0.92	0.31	0.46	0	0	1	0.83	-1.33	0.36	0.48	0	0	1	0.60	-1.66
Forward citations 1990	0.18	0.57	0	0	5	4.75	30.24	0.00	0.00	0	0	0	-	-	0.00	0.00	0	0	0	-	-
Forward citations 1991	0.53	1.06	0	0	7	3.10	12.26	0.00	0.00	0	0	0	-	-	0.00	0.00	0	0	0	-	-
Forward citations 1992	0.98	1.72	0	0	11	2.82	9.71	0.00	0.00	0	0	0	-	-	0.00	0.00	0	0	0	-	-
Forward citations 1993	1.52	2.41	0	1	18	2.92	11.26	0.00	0.00	0	0	0	-	-	0.00	0.00	0	0	0	-	-
Forward citations 1994	2.10	2.98	0	1	24	2.91	12.82	0.01	0.09	0	0	1	11.36	129.00	0.00	0.00	0	0	0	-	-
Forward citations 1995	2.66	3.62	0	1	28	2.80	11.92	0.11	0.34	0	0	2	3.15	9.98	0.00	0.00	0	0	0	-	-
Forward citations 1996	3.27	4.23	0	2	34	2.81	12.07	0.51	0.87	0	0	4	1.79	2.74	0.00	0.00	0	0	0	-	-
Forward citations 1997	4.00	5.20	0	2	38	2.95	12.14	1.22	2.56	0	0	24	5.92	49.09	0.00	0.00	0	0	0	-	-
Forward citations 1998	4.77	6.25	0	3	45	3.02	12.14	2.39	4.39	0	1	32	4.52	25.72	0.01	0.07	0	0	1	13.38	179.00
Forward citations 1999	5.53	7.30	0	3	48	3.04	11.49	3.43	5.48	0	2	39	4.15	21.90	0.02	0.15	0	0	1	6.52	40.94
Forward citations 2000	6.40	8.58	0	4	54	3.13	11.88	4.68	7.50	0	3	61	4.51	27.40	0.22	0.55	0	0	3	2.81	8.57
Forward citations 2001	7.24	9.46	0	4	65	3.14	12.30	6.45	9.71	0	4	75	4.19	23.06	0.96	1.53	0	0	7	2.15	4.90
Forward citations 2002	8.03	10.32	0	5	75	3.18	13.03	8.11	12.93	0	5	97	4.75	27.72	1.96	3.70	0	1	40	6.57	62.95
Forward citations 2003	8.87	11.32	0	6	81	3.13	12.65	9.25	14.31	0	5	102	4.36	23.60	2.97	4.64	0	2	45	4.72	37.34
Forward citations 2004	9.87	12.36	0	6	93	3.24	14.05	10.27	15.10	0	6	105	4.14	21.58	3.60	5.14	0	2	46	3.81	25.45
Forward citations 2005	10.03	12.63	0	6	93	3.30	14.78	11.06	15.97	0	7	117	4.18	22.31	4.07	5.61	0	2	49	3.59	22.68
Forward citations 2006	10.26	12.86	0	6	99	3.34	15.32	11.33	16.36	0	7	122	4.24	22.91	4.33	5.93	0	2	50	3.38	19.59

Table 8.5: Cross loadings of indicators for A-measurement models and samples 1, 2 and 3.

	1989-1990-1991				1995-1996				2000			
Indicator	**KS**	**TS**	**IS**	**USE**	**KS**	**TS**	**IS**	**USE**	**KS**	**TS**	**IS**	**USE**
Backward citations	0.277	0.107	-0.127	0.056	0.156	0.189	-0.088	0.044	-0.992	-0.162	0.048	-0.193
Number of inventors	0.932	0.088	0.364	0.142	0.992	0.283	0.133	0.316	-0.034	-0.210	0.031	0.009
Number of IPC codes	0.098	0.817	0.103	0.195	0.147	0.434	0.024	0.120	-0.178	-0.997	-0.111	-0.096
Number of claims	0.080	0.606	-0.154	0.159	0.278	0.936	-0.010	0.271	-0.168	-0.157	0.089	0.003
Priority JP	0.302	-0.008	0.994	0.048	0.181	0.038	0.772	0.091	0.139	-0.015	-0.055	-0.045
Priority DE	0.014	0.011	-0.013	0.002	-0.044	-0.049	0.574	0.094	0.023	-0.098	-0.978	-0.153
Family size	0.125	0.179	0.194	0.223	0.312	0.339	0.029	0.371	-0.241	-0.236	-0.089	-0.243
Dummy JP	0.252	0.186	0.365	0.316	0.401	0.241	0.370	0.462	-0.012	-0.147	-0.133	-0.198
Dummy DE	0.121	0.148	0.151	0.148	0.365	0.228	0.276	0.413	-0.112	-0.001	-0.297	-0.347
Forward citations 1992	0.130	0.217	0.030	0.755								
Forward citations 1993	0.127	0.222	-0.004	0.831								
Forward citations 1994	0.157	0.223	-0.011	0.866								
Forward citations 1995	0.160	0.227	0.008	0.900								
Forward citations 1996	0.178	0.257	0.030	0.931								
Forward citations 1997	0.140	0.260	0.043	0.947	0.157	0.112	0.133	0.766				
Forward citations 1998	0.116	0.232	0.025	0.958	0.175	0.180	0.084	0.855				
Forward citations 1999	0.108	0.210	0.016	0.957	0.184	0.211	0.071	0.906				
Forward citations 2000	0.098	0.187	0.010	0.954	0.191	0.213	0.079	0.936				
Forward citations 2001	0.112	0.182	0.012	0.943	0.234	0.250	0.083	0.952	0.149	0.084	0.092	0.820
Forward citations 2002	0.097	0.171	-0.006	0.930	0.214	0.223	0.057	0.932	0.175	0.042	0.054	0.921
Forward citations 2003	0.094	0.172	-0.014	0.920	0.218	0.226	0.030	0.927	0.155	0.017	0.084	0.928
Forward citations 2004	0.084	0.158	-0.028	0.899	0.257	0.232	0.020	0.925	0.120	0.021	0.120	0.929
Forward citations 2005	0.079	0.157	-0.032	0.896	0.262	0.249	0.001	0.907	0.103	0.067	0.141	0.925

Table 8.6: Cross loadings of indicators for B-measurement models and sample 1.

Indicator	KS	TS	IS	USE-Int	USE 92-93	USE 94-95	USE 96-97	USE 98-99	USE 00-01	USE 02-03	USE 04-05	PV
Backward citations	0.342	0.112	-0.149	-0.004	0.045	0.067	0.076	0.063	0.039	0.036	0.045	0.061
Number of inventors	0.905	0.087	0.343	0.222	0.120	0.141	0.138	0.093	0.095	0.085	0.068	0.171
Number of IPC codes	0.098	0.785	0.097	0.265	0.182	0.183	0.222	0.173	0.124	0.092	0.068	0.206
Number of claims	0.088	0.647	-0.136	-0.012	0.138	0.142	0.148	0.148	0.149	0.171	0.179	0.186
Priority JP	0.287	-0.018	0.824	0.284	0.014	0.001	0.040	0.022	0.013	-0.007	-0.026	0.056
Priority DE	0.008	0.011	0.464	0.193	-0.014	-0.026	-0.029	-0.008	-0.025	-0.026	-0.030	-0.005
Family size	0.125	0.172	0.252	0.895	0.095	0.094	0.126	0.121	0.108	0.085	0.059	0.187
Dummy JP	0.242	0.174	0.373	0.886	0.180	0.158	0.188	0.201	0.218	0.198	0.176	0.293
Dummy DE	0.121	0.146	0.265	0.746	0.009	0.028	0.076	0.077	0.059	0.047	0.028	0.118
Forward citations 1992	0.132	0.217	0.034	0.138	0.974	0.820	0.736	0.658	0.606	0.567	0.545	0.757
Forward citations 1993	0.126	0.221	-0.023	0.121	0.976	0.922	0.833	0.743	0.684	0.647	0.620	0.832
Forward citations 1994	0.157	0.223	-0.022	0.127	0.905	0.989	0.886	0.788	0.729	0.688	0.657	0.866
Forward citations 1995	0.162	0.227	-0.005	0.122	0.866	0.989	0.939	0.845	0.780	0.738	0.705	0.896
Forward citations 1996	0.180	0.256	0.008	0.158	0.825	0.944	0.988	0.901	0.836	0.789	0.753	0.925
Forward citations 1997	0.142	0.259	0.030	0.170	0.767	0.878	0.987	0.957	0.894	0.843	0.804	0.935
Forward citations 1998	0.118	0.232	0.021	0.174	0.733	0.838	0.951	0.994	0.939	0.896	0.857	0.948
Forward citations 1999	0.109	0.211	0.008	0.168	0.698	0.802	0.918	0.993	0.973	0.934	0.897	0.953
Forward citations 2000	0.099	0.188	0.001	0.180	0.671	0.772	0.886	0.970	0.996	0.966	0.932	0.955
Forward citations 2001	0.112	0.185	-0.005	0.163	0.649	0.748	0.860	0.947	0.996	0.985	0.957	0.950
Forward citations 2002	0.098	0.175	-0.018	0.152	0.628	0.727	0.834	0.926	0.983	0.998	0.975	0.940
Forward citations 2003	0.094	0.177	-0.024	0.145	0.616	0.712	0.816	0.911	0.970	0.998	0.988	0.932
Forward citations 2004	0.085	0.163	-0.038	0.123	0.602	0.691	0.790	0.885	0.949	0.984	0.999	0.913
Forward citations 2005	0.081	0.162	-0.043	0.124	0.595	0.686	0.786	0.880	0.947	0.983	0.999	0.910

Table 8.7: Standardized weights and loadings of the A-measurement models according to the type of constructs and for samples 1, 2, and 3; t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level.

Construct	Indicator	Sample 1 1989-1990-1991		Sample 2 1995-1996		Sample 3 2000	
Knowledge stock	Backward citations	0.269	(0.914)	0.125	(0.639)	-1.004	(3.225**)
	Number of inventors	0.987	(4.225**)	0.988	(11.432**)	-0.131	(0.402)
Technological scope	Number of IPC	0.872	(3.427**)	0.353	(1.549)	-0.991	(3.304**)
	Number of claims	0.459	(1.563)	0.905	(6.145**)	-0.081	(0.271)
International scope	Priority JP	0.984	(4.170**)	0.821	(3.078**)	-0.211	(0.682)
	Priority DE	0.330	(1.098)	0.637	(2.164**)	-1.011	(3.062**)
Technological usefulness	Forward citations 1992	0.776	(10.285**)				
	Forward citations 1993	0.837	(15.605**)				
	Forward citations 1994	0.867	(21.051**)				
	Forward citations 1995	0.886	(25.339**)				
	Forward citations 1996	0.912	(56.225**)				
	Forward citations 1997	0.942	(73.476**)	0.766	(17.085**)		
	Forward citations 1998	0.954	(101.785**)	0.855	(26.683**)		
	Forward citations 1999	0.955	(96.355**)	0.906	(35.718**)		
	Forward citations 2000	0.954	(91.283**)	0.936	(53.855**)		
	Forward citations 2001	0.944	(80.711**)	0.952	(66.227**)	0.819	(16.468**)
	Forward citations 2002	0.933	(66.253**)	0.932	(57.330**)	0.921	(23.904**)
	Forward citations 2003	0.924	(58.308**)	0.927	(56.315**)	0.928	(22.558**)
	Forward citations 2004	0.905	(41.677**)	0.925	(66.034**)	0.928	(21.922**)
	Forward citations 2005	0.902	(40.289**)	0.907	(51.682**)	0.925	(22.181**)
	Family size	0.355	(2.426**)	0.371	(4.065**)	-0.243	(1.742*)
	Dummy JP	0.439	(2.894**)	0.462	(4.634**)	-0.198	(1.377)
	Dummy DE	0.256	(1.929*)	0.413	(3.457**)	-0.347	(2.455**)
Patent value	Backward citations	0.046	(0.435)	0.076	(0.871)	-0.381	(2.549**)
	Number of inventors	0.361	(3.029**)	0.447	(4.299**)	-0.029	(0.179)
	Number of IPC	0.357	(2.377**)	0.183	(1.681)	-0.333	(2.377**)
	Number of claims	0.177	(1.496)	0.384	(5.268**)	-0.070	(0.512)
	Priority JP	0.259	(2.147**)	0.177	(1.783*)	0.018	(0.121)
	Priority DE	0.045	(0.493)	0.106	(0.998)	-0.317	(2.295**)
	Forward citations 1992	0.763	(14.561**)				
	Forward citations 1993	0.829	(18.958**)				
	Forward citations 1994	0.859	(22.806**)				
	Forward citations 1995	0.891	(30.297**)				
	Forward citations 1996	0.923	(46.023**)				
	Forward citations 1997	0.935	(54.846**)	0.732	(13.267**)		
	Forward citations 1998	0.942	(63.915**)	0.818	(20.821**)		
	Forward citations 1999	0.940	(61.484**)	0.866	(25.174**)		
	Forward citations 2000	0.936	(55.249**)	0.896	(30.605**)		
	Forward citations 2001	0.928	(49.439**)	0.920	(40.372**)	0.770	(7.959**)
	Forward citations 2002	0.914	(41.874**)	0.895	(36.354**)	0.845	(8.707**)
	Forward citations 2003	0.905	(37.838**)	0.888	(35.426**)	0.848	(9.107**)
	Forward citations 2004	0.884	(29.482**)	0.892	(41.528**)	0.849	(8.682**)
	Forward citations 2005	0.880	(28.566**)	0.877	(36.892**)	0.856	(8.126**)
	Family size	0.375	(2.526**)	0.406	(4.146**)	-0.299	(1.950*)
	Dummy JP	0.498	(2.926**)	0.519	(5.076**)	-0.194	(1.177)
	Dummy DE	0.277	(1.973*)	0.456	(3.655**)	-0.354	(2.382**)

Table 8.8: Standardized weights and loadings of the B-measurement models according to the type of constructs and for sample 1; t-values in parenthesis, ** at the 0.01 significance level, * at the 0.05 significance level.

Construct	**Indicator**	**Sample 1: 1989-1990-1991**
Knowledge stock	Backward citations	0.427
		(1.344)
	Number of inventors	0.943
		(3.411**)
Technological scope	Number of IPC codes	0.763
		(3.110**)
	Number of claims	0.619
		(2.172**)
International scope	Priority JP	0.893
		(3.107**)
	Priority DE	0.570
		(1.892*)
Usefulness Int.	Family size	0.895
		(5.129**)
	Dummy JP	0.886
		(4.382**)
	Dummy DE	0.746
		(2.606**)
Technological usefulness	Forward citations 1992	0.974
1992-1993		(100.067**)
	Forward citations 1993	0.976
		(146.971**)
Technological usefulness	Forward citations 1994	0.989
1994-1995		(275.169**)
	Forward citations 1995	0.989
		(301.685**)
Technological usefulness	Forward citations 1996	0.988
1996-1997		(159.570**)
	Forward citations 1997	0.987
		(158.284**)
Technological usefulness	Forward citations 1998	0.994
1998-1999		(447.793**)
	Forward citations 1999	0.993
		(437.170**)
Technological usefulness	Forward citations 2000	0.996
2000-2001		(806.005**)
	Forward citations 2001	0.996
		(818.647**)
Technological usefulness	Forward citations 2002	0.998
2002-2003		(1564.357**)
	Forward citations 2003	0.998
		(1559.498**)
Technological usefulness	Forward citations 2004	1.000
2004-2005		(8624.720**)
	Forward citations 2005	1.000
		(8526.497**)

Table 8.9: Correlations among constructs and mean communalities; longitudinal model with time-dependent manifest variables (model A) and samples 1, 2, and 3.

	1989-1990-1991				**1995-1996**				**2000**			
Construct	**KS**	**TS**	**IS**	**USE**	**KS**	**TS**	**IS**	**USE**	**KS**	**TS**	**IS**	**USE**
Knowledge stock	1	0.116	0.323	0.185	1	0.303	0.121	0.318	1	0.190	-0.052	0.192
Technological scope	0.116	1	0.021	0.269	0.303	1	0.000	0.287	0.190	1	0.103	0.095
International scope	0.323	0.021	1	0.100	0.121	0.000	1	0.134	-0.052	0.103	1	0.164
Technological usefulness	0.185	0.269	0.100	1	0.318	0.287	0.134	1	0.192	0.095	0.164	1
Mean Communalities (AVE)	0.480	0.516	0.469	0.583	0.504	0.532	0.463	0.654	0.492	0.509	0.480	0.540

Table 8.10: Correlations among constructs and mean communalities; longitudinal model with time-dependent latent variables and sample 1 (1989-1990-1991).

Construct	**KS**	**TS**	**IS**	**USE-Int**	**USE 92-93**	**USE 94-95**	**USE 96-97**	**USE 98-99**	**USE 00-01**	**USE 02-03**	**USE 04-05**
Knowledge stock	1	0.129	0.260	0.208	0.132	0.161	0.163	0.115	0.106	0.096	0.083
Technological scope	0.129	1	-0.010	0.195	0.225	0.227	0.261	0.223	0.187	0.176	0.163
International scope	0.260	-0.010	1	0.363	0.005	-0.014	0.019	0.015	-0.002	-0.021	-0.041
Usefulness-Int	0.208	0.195	0.363	1	0.133	0.126	0.166	0.172	0.172	0.149	0.123
Technological usefulness 92-93	0.132	0.225	0.005	0.133	1	0.895	0.806	0.720	0.662	0.623	0.599
Technological usefulness 94-95	0.161	0.227	-0.014	0.126	0.895	1	0.923	0.825	0.763	0.721	0.689
Technological usefulness 96-97	0.163	0.261	0.019	0.166	0.806	0.923	1	0.941	0.876	0.827	0.789
Technological usefulness 98-99	0.115	0.223	0.015	0.172	0.720	0.825	0.941	1	0.962	0.921	0.883
Technological usefulness 00-01	0.106	0.187	-0.002	0.172	0.662	0.763	0.876	0.962	1	0.979	0.948
Technological usefulness 02-03	0.096	0.176	-0.021	0.149	0.623	0.721	0.827	0.921	0.979	1	0.984
Technological usefulness 04-05	0.083	0.163	-0.041	0.123	0.599	0.689	0.789	0.883	0.948	0.984	1
Mean Communalities (AVE)	0.468	0.518	0.447	0.714	0.950	0.978	0.975	0.987	0.993	0.996	0.999

Table 8.11: Percentage of variation accounted for by partial least squares components, both individual and cumulative, and Q_h^2 index. The patent value is regressed on the knowledge stock, the technological scope, the international scope, the technological usefulness for the different time periods and the auxiliary variable USE-Int.

		Redundancy	
Component	**Q_h^2 index**	**Individual**	**Cumulative**
1	0.999	55.327	55.327
2	0.869	12.973	68.300
3	0.632	8.570	76.870
4	0.041	6.517	83.387

Chapter 9

Two-Step PLS Path Modelling Mode C to Estimate Nonlinear and Interaction Effects among Formative Constructs: Monte Carlo Simulations and Patent Value Models

Abstract. A Two-Step PLS Path Modelling Mode C (TsPLS) procedure is implemented to estimate nonlinear and interaction effects among formative constructs. A Monte Carlo simulation study is carried out in order to provide empirical evidence of its performance. Findings suggest that the TsPLS procedure offers a way to build "proper indices" for linear, nonlinear and interaction terms, all of which are unobservable, and to estimate the relationships between them. Inner linear, nonlinear and interaction effects are underestimated and outer relationships overestimated. Accuracy and precision increase with increasing sample size and number of observed variables per construct. In addition, a patent value model is used to illustrate the procedure.

9.1 Introduction

Nonlinearities in structural equation models (SEMs) can be addressed in different ways, either investigating nonlinear relationships between latent and manifest variables or among latent variables. Even though there is a wide literature on nonlinear covariance-based models[1], little attention has been paid to Nonlinear PLS Path Modelling. In the latter framework, most of the research has concentrated on modelling the nonlinear relationship between manifest and latent variables, and on studying the interaction effects among latent variables. Four procedures have been proposed to estimate interaction effects among latent variables when reflective outer models are modeled with Mode A: the product indicator approach (Chin *et al.*, 2003), the two-stage approach (Henseler & Fassott, 2005; Henseler *et al.*, 2009), Wold's approach (Wold, 1982), and the orthogonalizing approach (Little *et al.*, 2006). Nevertheless, these procedures cannot be used for estimating nonlinearities among formative constructs, "since formative indicators are not assumed to reflect the same underlying construct (i.e., can be independent of one another and measuring different factors), the product indicators between two sets of formative indicators will not necessarily tap into the same underlying interaction effect" (Chin *et al.*, 2003, p. 11, supplemental material).

The purpose of this chapter is two-fold. First, a Two-Step PLS Path Modelling Mode C (TsPLS) procedure is implemented to estimate nonlinear and interaction effects among formative constructs. Following Kenny and Judd's approach, power and cross-product terms of constructs are included in the structural relationships among unobserved variables. Thus, if ξ_1 and ξ_2 are formative exogenous constructs and η is a reflective endogenous latent variable, the following nonlinear and moderating constructs are related to η: ξ_1^2, ξ_2^2, and $\xi_1\xi_2$. The procedure considers the score estimation of linear constructs in the usual way (step one). Formative constructs are computed using PLS Mode B and reflective latent variables using PLS Mode A. This is called PLS Mode C in Wold's approach (Wold, 1982, p. 10). When the convergence is reached, scores of nonlinear and interaction terms are directly computed (step two). The dependent latent variable is regressed on the linear, nonlinear and moderate latent terms. A Monte Carlo simulation study is carried out in order to provide empirical evidence of the performance of the algorithm. Results are provided in terms of mean value of the estimates, mean standard deviations (dispersion), mean confidence intervals (reliability), biases (accuracy), variances (variability), mean square errors (precision) and mean square biases (accuracy).

The second purpose of this chapter is to study nonlinearities in a real case on patent valuation. The structural equation model of patent value considers three formative exogenous constructs: the knowledge stock used by the applicant to create an invention, the technological scope of the invention, and the international scope of protection (Martínez-Ruiz & Aluja-Banet, 2009). These variables are modeled as determinants of the patent value, which is defined as a reflective en-

[1]See, for instance, McDonald (1962); Busemeyer & Jones (1983); Kenny & Judd (1984); Bollen (1995); Rizopoulos & Moustaki (2008) and references therein.

dogenous latent variable. Manifest variables are built from information contained in patent documents from a sample of patents granted in the U.S. in the renewable energy field. In this research we are interested in knowing (1) if the knowledge stock is moderating the relationship between the international scope and patent value and (2) if there is a nonlinear effect of the international scope on patent value.

In this chapter, we briefly examine the different approaches for modelling nonlinear and interaction effects considered in a PLS Path Modelling framework. Then, we describe and report the TsPLS procedure and the Monte Carlo simulation study, respectively. Finally, we present a true case of patent valuation.

9.2 Background

Up to now and from a component-based approach, researchers have focused mainly on studying interaction effects between unobservable variables and nonlinear relationships between manifest and latent variables. Studies have only considered SEMs with reflective measurement models.

9.2.1 Nonlinear Relationships between Manifest and Latent Variables

Krämer (2005) studied the nonlinear relationship between manifest and latent variables, and replaced the inner products in the outer estimation of formative blocks of variables by kernel functions. Jakobowicz (2007), however, argues that this approach has important shortcomings; among them, the difficult interpretation of the weights and the selection of the kernel function are the most important failings. Jakobowicz & Saporta (2007) also examined the nonlinear relationships in outer models. They based their analysis on the similarities between latent variable scores and the first principal component. In order to maximize the explanatory power of the first principal component in PLS Path Modelling in a reflective approach, they proposed transforming the observed variables with nonlinear principal component analysis and monotic B-spline functions. Moreover, Jakobowicz (2007) proposed transforming the inner model by monotic B-spline functions, estimating the model parameters in two stages. Even though in this case the model relationships can be interpreted, the transformation properties have yet to be studied.

9.2.2 Approaches to Interaction Effects

A moderator variable is "a qualitative or quantitative variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable" (Baron & Kenny, 1986, p. 1174). In a SEM context, the idea is to study whether there is any unobserved variable that has a nonlinear or interaction effect on a given endogenous latent variable. Four approaches have been suggested to deal with this problem in a PLS framework: Wold's approach (Wold, 1982), the product indicator approach (Chin *et al.*, 2003), the two-stage approach (Henseler & Fassott, 2005; Henseler *et al.*, 2009), and the orthogonalizing approach (Little *et al.*, 2006). Despite the fact that none of these approaches can be used for formative outer models, to the best of our knowledge, contributions on these topics are rather scarce. So, we give a brief overview of the methods.

The product indicator approach. The product indicator procedure was introduced by Busemeyer & Jones (1983) and Kenny & Judd (1984) in a covariance-based framework, and by Chin *et al.* (2003) in PLS Mode A to estimate interaction effects. Cross-product terms are introduced and modeled as unobserved variables. Building a set of indicators for these terms, the entire structural model is estimated in the usual manner. To do this, each indicator of the moderator construct should be multiplied by each indicator of the predictor variable. The main drawback of this procedure is that there is no criterion for determining the number of indicators of the interaction term.

The two-step approach. The idea of a two-step procedure for SEM estimation comes from the covariance-based scientific community (Anderson & Gerbing, 1988), and it has been suggested by Chin *et al.* (2003) for estimating SEMs with formative constructs[2]. Henseler & Chin (2010) implemented the two-step approach by running the PLS Path Modelling algorithm to compute latent variable scores in the usual manner (step one). Once these are obtained, the interaction effects are calculated (step two) and the dependent latent variable is regressed on single and interaction effects –if a multiple linear regression procedure is followed. The researchers implemented the procedure for SEMs with reflective outer models. The main drawback of this procedure is that the interaction effects are not considered in the computation of unobservable variable scores (limited-information approach).

Wold's approach. The original solution of Herman Wold (1982) for estimating interaction effects was presented for a model with two latent variables, and it focused on modelling a nonlinear inner relationship through a quadratic term. In his research, the nonlinear latent variable is calculated as the power term of the construct within the PLS Path Modelling procedure, i.e. in the iterative stage. A

[2]Recall that in PLS Path Modelling, reflective and formative representations can be modeled. In the iterative stage, Mode A is usually used to upgrade the weights vector in reflective outer models; and Mode B in the case of formative outer models.

generalization of Wold's approach can be found in Henseler & Chin (2010). These researchers call it "the hybrid approach" because it combines elements from the product indicator and the two-step approaches. The key aspect of this procedure is the non-standardization of nonlinear and moderating variables, which allow the effects to be interpreted once they are calculated.

The orthogonal approach. This procedure[3] has been suggested in a covariance-based approach by Little *et al.* (2006). As the nonlinear terms are usually computed directly from latent variable scores, it is necessary to address multicollinearity among independent variables in the multiple regression of the endogenous on the exogenous latent variables. The orthogonal approach was introduced to overcome this drawback. Basically, the procedure consists of building indicators by multiplying the manifest variables of predictor and moderator variables as in a product indicator approach. The residuals from the regressions of these indicators on the variables of the predictor and moderator variables are used as indicators of the interaction term. Henseler *et al.* (2007) suggested that this approach is easily transferable to PLS path modelling and Henseler & Chin (2010) finally implemented the procedure.

9.2.3 Simulation studies

Simulation studies have been conducted to assess whether the PLS Path Modelling algorithm can detect the presence of nonlinear or interaction effects. These investigations have only worked with reflective measurement models, however. Through an extensive Monte Carlo simulation study, Chin *et al.* (2003) have compared the product indicator approach and the product of the sums approach. The latter was suggested by Cohen & Cohen (1983) in a covariance-based specification and considered by Chin *et al.* (2003) and Goodhue *et al.* (2007) in a PLS context. As seen above, in a product indicator approach, a moderating latent variable is created by considering all possible combinations obtained by multiplying the indicators from predictor and moderator variables. In a product of the sums approach, the interaction variable is calculated as the sum of the items of the first construct multiplied by the sum of the items of the second construct. Thus, the variables are aggregated in a single score (summed or averaged). In both approaches, the dependent variable is regressed on the predictors, the moderator variable and the interaction effect. Chin *et al.* (2003) showed that the product indicator approach is superior to regression with a product of the sums approach. The interaction path coefficient of product indicator approach "was larger and closer to the true

[3]This procedure is based on the so-called residual-centering approach that is widely used in multiple regression when there is multicollinearity between interaction terms and first-order terms (Lance, 1988).

parameter value." Goodhue *et al.* (2007) argued that this approach provides less statistical power than the product of the sums approach, and this problem increases with the number of indicators and with the sample size. They recommend that "if having a sufficient sample size to achieve statistical significance for an interaction path is a concern, regression or PLS with a product of sums approach would be preferred to PLS with product indicator approach" (p. 213). The results of these research efforts are not yet conclusive.

Recently, Henseler & Fassott (2005), Henseler *et al.* (2007), Henseler & Fassott (2010) and Henseler & Chin (2010) have compared the performance of the four interaction effects approaches described above for a PLS Path Modelling framework. Henseler & Chin (2010) concluded that:

1. The two-step approach and Wold's approach (or hybrid approach) place more emphasis on controlling type II errors (statistical power), the two-step approach being the recommended procedure for assessing the significance of an interaction effect.

2. For medium and large sample sizes, the product indicator approach provides the smaller biases.

3. For small sample sizes, the orthogonal approach allows for estimates closest to the true values (lower bias).

4. The product indicator approach and the orthogonal approach are recommended if the researcher is looking for small mean square errors (foci in precision and prediction).

5. As the prediction is the main foci of the PLS Path Modelling procedure, the orthogonal approach is strongly recommended for estimating interaction effects given that it provides accurate estimates for interaction and direct effects.

From another perspective, moderating effects have also been integrated in component-based models by splitting up the data into groups according to a categorical variable, and estimating the model for each data group (Henseler & Fassott, 2005). The underlying idea is that the differences between estimates moderate the relationships among the variables in the model. This approach has also been followed in covariance-based specifications by Ping (1995). Qureshi & Compeau (2009) focused on comparing covariance– and component–based SEM multigroup analysis for checking the ability of the methods to detect between-group differences. The researchers found three important results, summarized as follows:

1. The component-based approach outperforms the covariance-based approach when the sample size is small, data are normally distributed, and the exogenous variables are correlated.

2. The covariance-based approach performs similarly to the component-based approach with large samples and normal data.

3. In the case of non-normal data, "neither technique could consistently detect differences across the groups in two of the paths, suggesting that both techniques struggle with the prediction of a highly skewed and kurtotic dependent variable. Both techniques detected the differences in the other paths consistently under conditions of non-normality, with the component-based approach preferable at moderate effect sizes, particularly for smaller samples" (p. 197).

9.2.4 Formative Constructs

Recently there has been an increase in the use of formative outer models to model manifest and latent relationships (see Chapter 6). However, empirical evidence is still not conclusive enough to provide a roadmap to researchers for modelling formative constructs, at least in a PLS approach. Simulation studies considering formative block of variables have been carried out, but they have focused on investigating the PLS Path Modelling performance (Cassel *et al.*, 1999); the study of the effect of erroneously assuming a model as reflective when it is actually formative (Jarvis *et al.*, 2003; Petter *et al.*, 2007; Vilares *et al.*, 2010); the analysis of the PLS Path Modelling behavior against asymmetric data (Vilares *et al.*, 2010) and multicollinearity (Westlund *et al.*, 2008); and the comparison of the robustness of the covariance- and component-based approaches (Ringle *et al.*, 2009). So far, there is no evidence or guidelines to assess how well the PLS algorithm detects the presence or absence of nonlinear and/or interaction effects among formative constructs. So, using Monte Carlo simulations, this research is the first to investigate the performance of a Two-Step PLS Path Modelling Mode C (TsPLS) procedure for SEMs with reflective and formative measurement models.

Recall that formative outer models are used to minimize residuals in the relationships between latent and manifest variables. Manifest variables cause a change in the latent variable and they should be a census of indicators, not a sample (Bollen & Lennox, 1991). Additionally, a strong theory must support the decision to model a block of variables as formative (Fornell & Bookstein, 1982; Cohen *et al.*, 1990; Bollen & Lennox, 1991). Unlike the reflective measurement models, the manifest variables forming a construct do not have to be especially correlated. On the contrary, the multicollinearity affects the stability of the estimates.

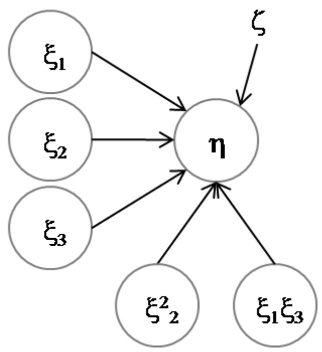


Figure 9.1: Path diagram of a structural equation model with linear (ξ_1, ξ_2 and ξ_3) nonlinear (ξ_2^2) and interaction ($\xi_1\xi_3$) effects.

9.3 Two-Step PLS Path Modelling Mode C (TsPLS)

For considering nonlinearities in a structural equation model, linear relationships between latent variables are replaced by a linear polynomial model. So, quadratic and cross-product terms of the constructs are introduced in the relationships between exogenous and endogenous unobservable variables. Thus, if ξ_1, ξ_2, and ξ_3 are exogenous constructs and η is an endogenous latent variable, the following nonlinear and interaction constructs may be related to η: ξ_2^2 and $\xi_1\xi_3$ (see Figure 9.1). Equation 9.1 describes the structural relationship between the dependent latent variable and the linear and nonlinear unobservable terms:

$$\eta = \beta_{j0} + \sum_j \beta_j \xi_j + \sum_j \alpha_j \xi_j^2 + \sum_j \sum_i \gamma_{ji} \xi_j \xi_i + \zeta \tag{9.1}$$

where ξ_j and η are the exogenous and endogenous unobservable variables respectively, β_j, α_j, and γ_j are regression coefficients, and ζ is the residual term. This relationship considers the main effects of unobservable variables, but also allowing for nonlinear and interaction effects of exogenous constructs.

To compute nonlinear terms, a two-step score construction procedure was implemented based on the Lohmöller specification for PLS Path Modelling. Unobservable variables are computed as usual in the iterative stage of the PLS algorithm (step one). Mode A is used to update the outer weights in reflective outer models, and Mode B in formative outer models. Once the algorithm converges, the quadratic and cross-product terms are directly computed from the value of the unobservable variables (step two)[4]. Then, the endogenous latent variable is regressed on the linear, nonlinear and moderate latent terms.

[4]Note that it is not explicitly considered a set of manifest variables to estimate nonlinear and interaction terms.

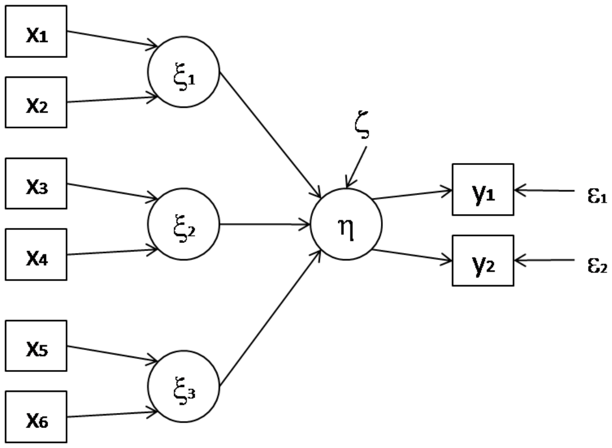


Figure 9.2: Structural and measurement models of the simulated setups; measurement models consider two, four, six, and eight indicators per construct.

9.4 Designing the Monte Carlo Simulation Study

A simulation study was performed to address several issues (Paxton *et al.*, 2001; Gentle, 2003). The aims were:

1. To analyze the performance of the TsPLS procedure to recover true values for linear, nonlinear and interaction effects when considering formative measurement models.

2. To analyze the performance of the TsPLS procedure when few indicators are considered per unobserved variable.

3. To analyze the performance of the TsPLS procedure when considering different sample sizes.

Based on a patent value model (Martínez-Ruiz & Aluja-Banet, 2009), the underlying PLS population model considered a simple structure with three formative exogenous constructs and one reflective endogenous latent variable (Figure 9.2). The experimental design considered models with two, four, six and eight indicators per construct, and four sample sizes (50, 100, 250, 500) were studied. Five hundred random data sets were generated for each of the 4×4 cells of the two-factor design. Five hundred replications (t) were made for each cell in the design. TsPLS algorithm with centroid scheme was implemented in R-project (R Development Core Team, 2007). To update the outer weights vectors, Mode A was used in reflective outer models and Mode B in formative outer models. We arbitrarily investigated the nonlinear effect of the second construct (ξ_2), and the moderating effect of the first construct (ξ_1) on the third unobservable variable (ξ_3). Results are provided in terms of the:

- Mean value of the estimates.

- Mean standard deviation (dispersion) and variance (variability).

- Mean confidence intervals (reliability).

- Bias (accuracy, $\frac{1}{t}\sum_{i=1}^{t} E[\theta_i] - \theta$) and mean relative bias (accuracy, MRB= $100 * \frac{1}{t}\sum_{i=1}^{t} \frac{\theta - E[\theta_i]}{\theta}$, Chin *et al.* (2003)).

- Mean square error (precision, MSE= $Bias^2 + Variance$).

9.4.1 Generating Data

Data were generated from a component-based model (Schneeweiss, 1991). For each formative exogenous construct ξ_j, we began generating standardized manifest variables x_{jh} as independent normal data. This is quite consistent with the literature review above, where manifest variables in a formative measurement model do not have to have a special type of relationship and should rather represent different facets of a construct. Once the manifest variables were generated, we computed the linear, nonlinear and interaction terms. The endogenous latent variable η was calculated as a linear combination of the exogenous unobserved variables, and the nonlinear and interaction terms. Disturbance terms were computed as random normal data with a zero mean and the corresponding standard deviation. They were distributed independently of unobservable variables. Standardized observed variables y_i of reflective outer models are generated in the usual way: by considering errors as random normal data with a zero mean, the corresponding standard deviation and they were uncorrelated with the latent variable. The variables are generated so that the variance of the errors and disturbance terms are positive. It is worth noting that in all cases, the generated exogenous constructs are not collinear.

9.4.2 Interpreting the Regression Coefficients: Some Comments on Standardization

We follow the recommendation already made for multiple regression (Allison, 1977; Bohrnstedt & Marwell, 1978)[5]. Therefore, independent and dependent variables, which are both linear and unobservable, are standardized; nonlinear and interaction latent terms are not standardized. If the regression coefficients are significant, this procedure ensures the interpretability of the coefficients. Henseler & Chin (2010) used this approach for estimating interaction effects in structural equation models with reflective measurement models.

[5]Recall that in multiple regression, standardized coefficient of interaction effects are affected by changes in the means of the variables or the correlations between predictor and moderator variable (Allison, 1977).

9.4.3 Setting True Population Parameters

To set the true population parameters for the PLS models, we took into account different combinations of values so as to show whether they are recovered by the TsPLs procedure. In addition, we are especially interested in analyzing the performance when there are few indicators per construct. In social sciences, particularly in literature regarding technological change and technology watch studies, relationships among variables are often small or moderate[6]. Thus, we wish to establish whether or not this method is able to recover this type of relationship. Hence, the values of the coefficients were also set up in an attempt to get closer to the obtained coefficients when estimating patent value models. Table 9.1 shows the true population values of weights, loadings, linear, nonlinear and interaction effects. We consider large values for all the true loadings, at least 0.7 in the case of two manifest variables per latent variable. This ensures the unidimensionality of the block of variables and it satisfies the condition imposed by the PLS Path Modelling algorithm.

9.5 Simulation Results

Figures 9.3, 9.5, 9.7 and 9.10 show the mean bias of the weights, path coefficients and loadings when sample sizes and the number of indicators increase. Figures 9.4, 9.8 and 9.9 show the mean relative bias of weights, loadings and path coefficients, depending on the sample size and the number of indicators. In Appendix 9.8, Tables 9.4 to 9.17 show the mean estimates, standard deviation, confidence interval, bias, variance, mean square error (MSE) and mean relative bias (MRB) for weights, path coefficients and loadings for the analyzed cases.

9.5.1 Estimating Weights in Formative Outer Models

The TsPLS procedure yields are similar to the PLS basic design (see Tables 9.4, 9.5, 9.7 and 9.9). Figure 9.3 reports the mean bias of weight estimates for models with two, four, six and eight indicators when the sample size varies from 50 to 500 observations. The procedure tends to overestimate the relationships in formative outer models and results are consistent with the theoretical PLS framework (Wold, 1982; Dijkstra, 2010). Increasing the sample size, the bias decreases, and for N=500, the procedure almost exactly recovers all the true values (small, moderate and large values). Biases also decrease with the increasing number of indicators per construct. It is interesting to note the behavior of the algorithm for outer models with two manifest variables. Data dispersion is clearly larger. For

[6] Cohen (1988) suggests that correlations of 0.1, 0.3, and 0.5 express small, medium and large effect sizes, respectively.

Table 9.1: True population values for weights, loadings, linear, nonlinear and interaction effects; a model with three formative exogenous constructs and one reflective endogenous variable; cases for two, four, six and eight indicators in each outer model.

Coefficient	**2 MVs**	**4 MVs**
Weights	(0.8,0.5)	(0.2,0.3,0.5,0.7)
	(0.4,0.8)	(0.2,0.4,0.6,0.5)
	(0.1,0.9)	(0.3,0.5,0.7,0.2)
Linear effects	(0.5,0.4,0.3)[a]	(0.5,0.4,0.3)[b]
Nonlinear effects	0.3	0.3
Interaction effects	0.3	0.3
Loadings	(0.7,0.8)	(0.6,0.7,0.8,0.9)

Coefficient	**6 MVs**	**8 MVs**
Weights	(0.5,0.3,0.4,0.3,0.5,0.1)	(0.3,0.3,0.4,0.3,0.4,0.3,0.2,0.3)
	(0.2,0.4,0.6,0.4,0.2,0.3)	(0.3,0.3,0.4,0.4,0.2,0.3,0.4,0.2)
	(0.3,0.6,0.2,0.3,0.4,0.2)	(0.4,0.5,0.4,0.3,0.2,0.1,0.3,0.2)
Linear effects	(0.5,0.4,0.3)[c]	(0.5,0.4,0.3)[d]
Nonlinear effects	0.3	0.3
Interaction effects	0.3	0.3
Loadings	(0.6,0.7,0.8,0.9,0.6,0.7)	(0.6,0.7,0.8,0.9,0.6,0.7,0.8,0.9)

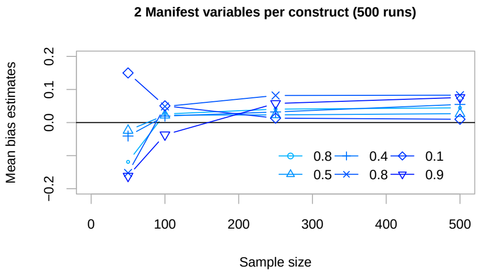
[a] For N=50 the true vector was (0.3,0.2,0.3).
[b] For N=50 and N=100 the true vectors were (0.3,0.3,0.2) and (0.3,0.4,0.2), respectively.
[c] For N=50 and N=100 the true vectors were (0.3,0.3,0.2) and (0.3,0.4,0.2), respectively.
[d] For N=50 and N=100 the true vectors was (0.3,0.4,0.2).

models with four, six and eight indicators, the biases markedly decrease for sample sizes larger than 250.

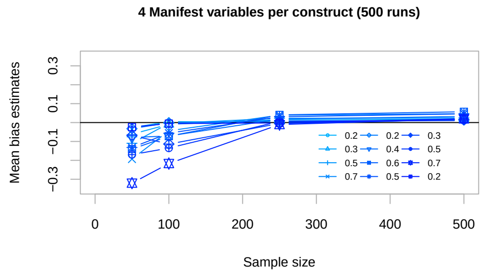
Figure 9.4(a) shows the mean relative bias for a weight of 0.5 (true value) when increasing the sample size and the number of indicators. MRB substantially decreases with increasing sample size. As in the simulated cases in Chapter 6, estimates improve by increasing the sample size more than by increasing the number of observable variables. Tables 9.4 to 9.9 also show that the largest MRBs are exhibited for models with the smallest sample sizes (N=50). For N=250 and two, four, six and eight indicators per construct, the average MRBs are 8%, 3%, 3% and 1%, respectively. In addition, variability and mean square errors decrease by increasing the sample size or by increasing the number of manifest variables in all the simulated cases.

9.5.2 Estimating Linear, Nonlinear and Interaction Effects

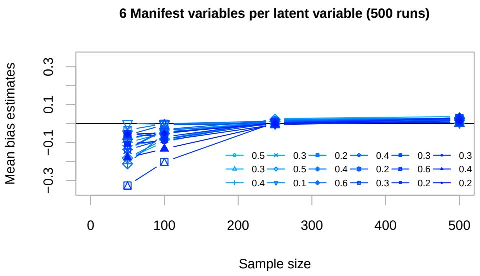
Figures 9.5, 9.6 and 9.7 show the mean bias of linear, nonlinear and interaction effects for models with two, four, six and eight manifest variables per construct



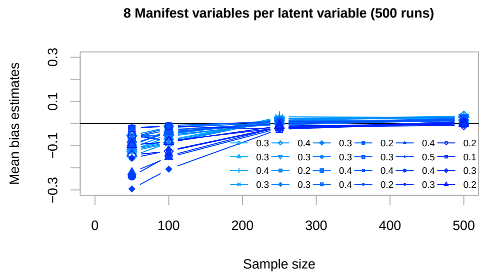
(a) Two indicators



(b) Four indicators

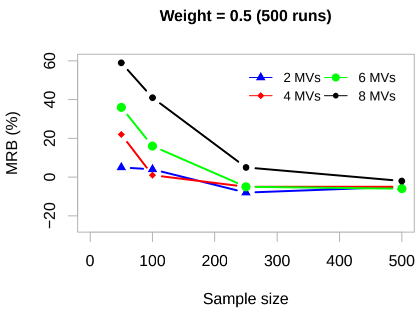


(c) Six indicators

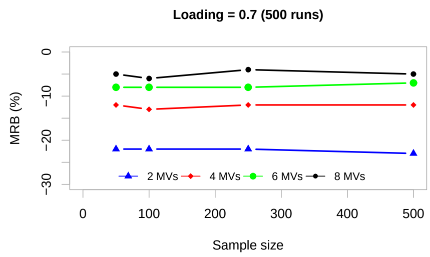


(d) Eight indicators

Figure 9.3: Mean bias of weight estimates. Highlighting the influence of the sample size and the number of indicators.



(a) MRB for a weight of 0.5

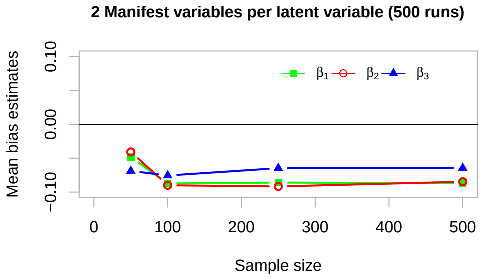


(b) MRB for a loading of 0.7

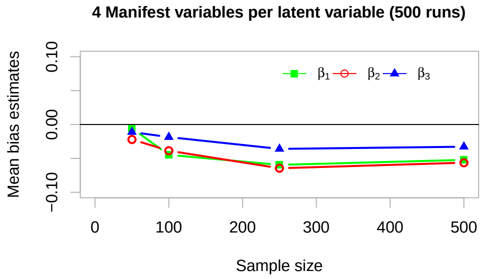
Figure 9.4: Mean relative bias of a weight and a loading. Highlighting the influence of the sample size and the number of indicators per construct.

when the sample size varies from 50 to 500 observations. Tables 9.10 to 9.13 in section 9.8 report the full list of results for inner relationships.

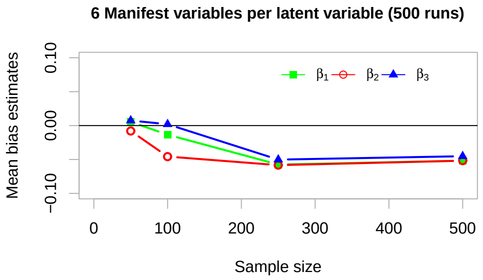
The TsPLS procedure clearly underestimates the linear, nonlinear and interaction effects. This happens for all sample sizes and number of indicators considered. Accuracy of estimates improves with increasing sample size and number of indicators. It is worth mentioning that when formative measurement models with two



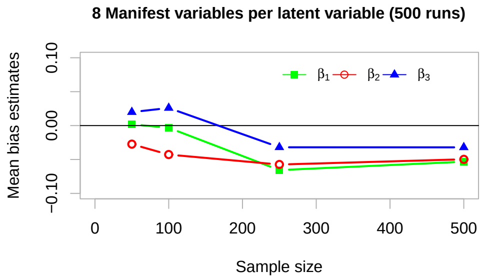
(a) Two indicators



(b) Four indicators



(c) Six indicators

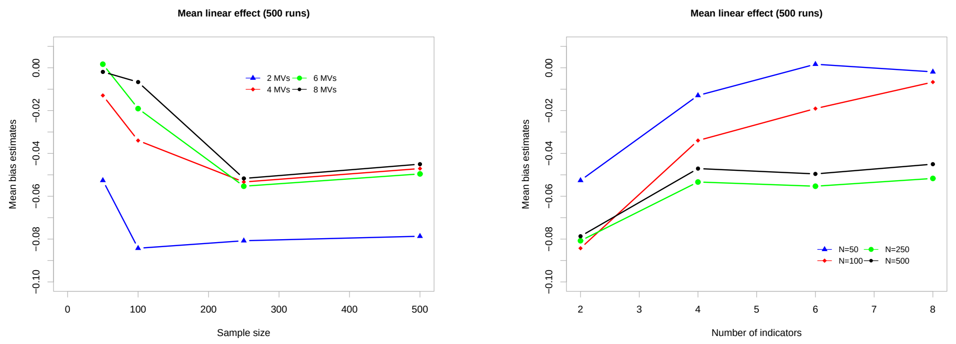


(d) Eight indicators

Figure 9.5: Mean bias of linear effects. Highlighting the influence of the sample size and the number of indicators.

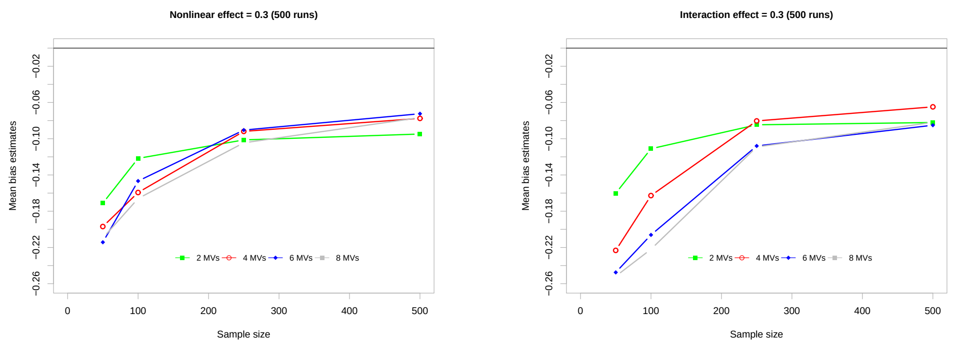
indicators per construct are considered, the bias of estimates of linear effects is less than 10% for all sample sizes. Figure 9.5 shows that biases decrease when increasing the number of indicators. Figure 9.7 clearly shows that estimates of nonlinear and interaction effects are more accurate when the sample size and the number of indicators are increased. For small sample sizes, however, estimates are more biased, regardless of the number of indicators considered in outer models.

Figures 9.8 and 9.9 report the mean relative bias for linear, nonlinear and interaction effects. Figure 9.8 shows the remarkable influence of the number of indicators per construct on the mean relative bias of linear effects, especially for sample sizes 50 and 100. For instance, for N=100 the MRBs of linear effects decrease an average 16.7% when the number of indicators varies from two to six in formative outer models. As expected, accuracy is higher for linear effects than for nonlinear and interaction effects (see Figure 9.9 and Tables 9.10 to 9.13). However, MRBs decrease when increasing the sample size and the number of indicators per construct. As for weights and linear effects, from N=250 on, the MRBs decrease remarkably, reaching values close to 20% when the sample size is 500. Both factors, sample size and number of indicators, improve accuracy and precision of linear, nonlinear and interaction effect estimates.



(a) Sample size (b) Number of indicators

Figure 9.6: Average mean bias of linear effects. Highlighting the influence of the sample size and the number of indicators per construct.



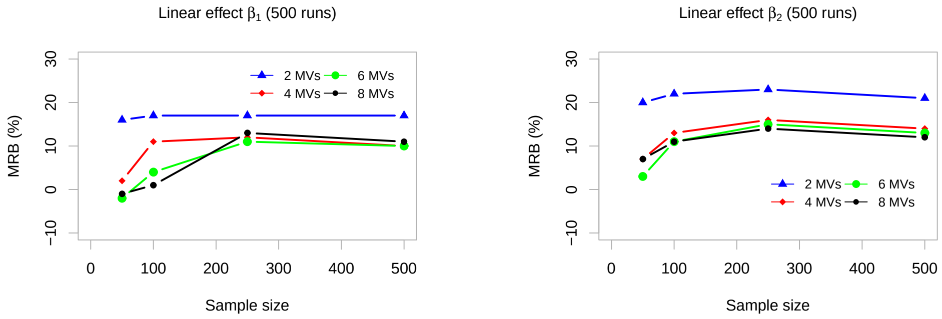
(a) Nonlinear effect of 0.3 (b) Interaction effect of 0.3

Figure 9.7: Mean bias of nonlinear and interaction effects. Highlighting the influence of the sample size and the number of indicators per construct.

9.5.3 Estimating Loadings in Reflective Outer Models

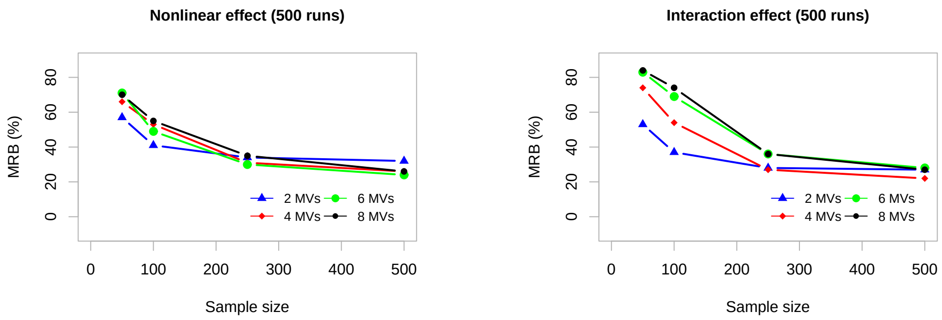
As can be seen in Figure 9.10, the estimates of loadings in all cases are very close to the true values, regardless of the sample size and number of manifest variables per construct. Mean biases are less than 20%. TsPLS procedure overestimates population values. Moreover, as in the case of the PLS Path Modelling Mode C procedure, a higher number of manifest variables seems to be more important than a higher sample size for decreasing the bias of the estimates in reflective outer models. This is clearly seen in Figure 9.4(b) where the mean relative bias for a loading of 0.7 strongly decreases when the number of indicators increases. So, this suggests that TsPLS estimates are "inconsistent," that they are only consistent at large.

Summing up, for the proposed interactive models, the estimates of the effects are quite consistent with previous results. In general, findings suggest that for-



(a) Linear effect β_1 (b) Linear effect β_2

Figure 9.8: Mean relative bias of linear effects. Highlighting the influence of the sample size and the number of indicators per construct.



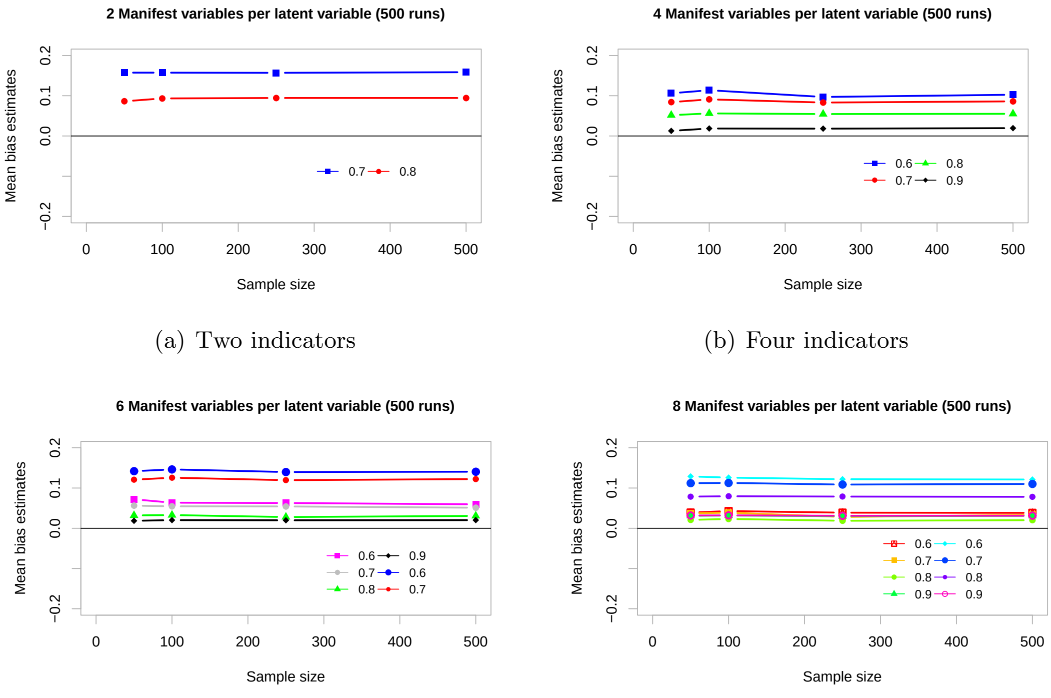
(a) Nonlinear effect of 0.3 (b) Interaction effect of 0.3

Figure 9.9: Mean relative bias of nonlinear and interaction effects. Highlighting the influence of the sample size and the number of indicators per construct.

mative and reflective relationships tend to be overestimated; the estimates are more reliable when outer models include four or more indicators; and precision increases remarkably from a sample size of 250. In addition, linear, nonlinear and interaction effects tend to be underestimated.

9.6 A Case Study: Patent Value Models with Nonlinearities

A second-order model of patent value has been proposed (Martínez-Ruiz & Aluja-Banet, 2009). The structural model includes three formative exogenous constructs –the knowledge stock (KS) used by the applicant to create an invention, the technological scope (TS) of the invention, and the international scope (IS) of protection– and two reflective endogenous latent variables –the technological use-



(a) Two indicators

(b) Four indicators

(c) Six indicators

(d) Eight indicators

Figure 9.10: Mean bias of loadings. Highlighting the influence of the sample size and the number of indicators.

fulness and the patent value[7]. To the best of our knowledge, there has been no previous study of interaction or nonlinear effects between variables that determine the patent value. Hence, we are interested in finding evidence for the existence of nonlinearities in an exploratory way. Since PLS is more suited for analyzing exploratory and causal-predictive models (Fornell & Bookstein, 1982), TsPLS is used to investigate these issues. To do this, three models were examined: a linear additive model (Figure 9.11(a)), an interactive model (Figure 9.11(c)) and a non-linear model (Figure 9.11(d)). The linear additive model is the second-order model of patent value. The interactive model –besides describing the linear relationships between the constructs– draws the moderating effect of the knowledge stock on the relationship between international scope and both patent value and technological usefulness (see Figure 9.11(b)). In Figure 9.11(c), the interaction term is displayed as $KS \times IS$. The nonlinear model depicts the linear relationships and the nonlinear effect of the international scope on both patent value and technological usefulness. The nonlinear term is represented as IS^2 in Figure 9.11(d).

Models are estimated using a patent sample as described in Chapter 7. In

[7]See Chapter 7 for patent value model specification.

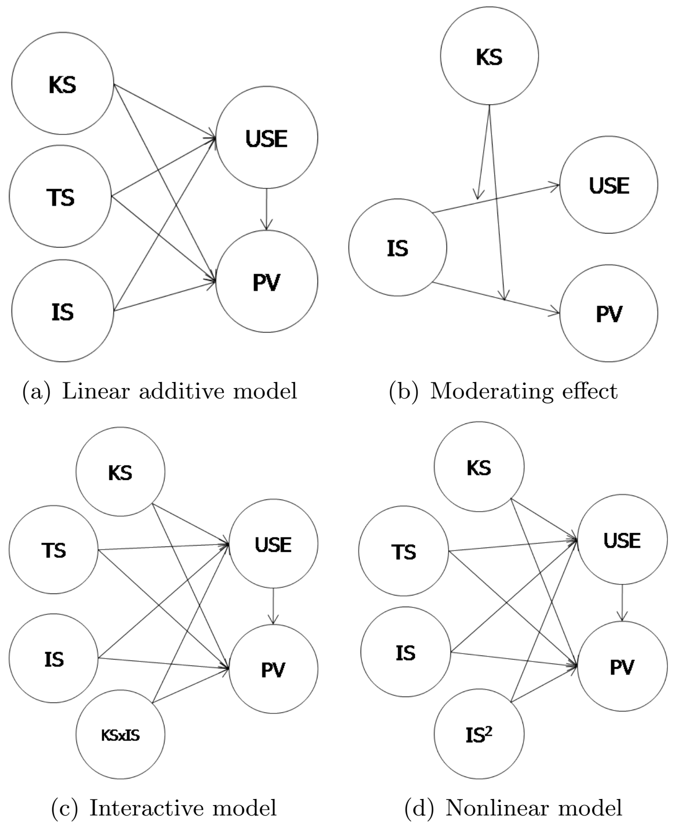


Figure 9.11: Linear additive model, moderating effects, interactive and nonlinear models of patent value; patent value (PV) and technological usefulness (USE) are endogenous latent variables; knowledge stock (KS), technological scope (TS), and international scope (IS) are formative exogenous constructs.

order to analyze whether there is a pattern in the magnitude of the effects and significance level of structural relationships, the models were estimated with three data sets: patents applied for in 1990-1991 (N=129), in 1995-1996 (N=128) and in 1999-2000 (N=536).

Table 9.2 shows the path coefficients obtained with the TsPLS procedure for the linear, interactive and nonlinear models of patent value with the three data sets. Table 9.3 shows the pattern of significance of the structural relationships for the second-order models of patent value with and without linearities. Recall that the TsPLS procedure yields standardized constructs and –to ensure the interpretability of the results– unstandardized interaction and nonlinear scores. Table 9.2 also shows the coefficient of multiple determination R^2 for endogenous constructs as well as the effect size f^2, F-values and p-values of the interaction and nonlinear terms on endogenous constructs. R^2 represents the amount of variance in the unobservable variables explained by the predictors. The effect size f^2 contrasts the difference of R^2 for the interactive/nonlinear models and the linear additive model. The effect size f^2 is computed as $(R^2_{\text{interactive model}} - R^2_{\text{linear additive model}})/(1 - R^2_{\text{interactive model}})$ where $R^2_{\text{interactive model}}$ considers all predictors and $R^2_{\text{linear additive model}}$ excludes the interaction term. Cohen (1988) suggests that f^2 values of 0.02, 0.15 and 0.35 are indicative of small, medium and large interaction effect sizes.

Table 9.2: Path coefficients for the linear, interactive and nonlinear models of patent value; coefficient of multiple determination R^2 for endogenous constructs, effect size f^2, F-test and p-values of interaction and nonlinear terms on endogenous constructs

Model	**Endogenous Variable**	**Exogenous Variable**	**1990-1991**	**1995-1996**	**1999-2000**
Linear additive model	Patent value	Knowledge stock	0.226	0.229	0.293
		Technological scope	0.227	0.227	0.272
		International scope	0.233	0.167	0.237
		Technological usefulness	0.669	0.698	0.699
	Technological usefulness	Knowledge stock	0.180	0.299	0.073
		Technological scope	0.316	0.334	0.207
		International scope	0.143	0.236	0.201
		R^2 of patent value	0.998	0.999	0.998
		R^2 of usefulness	0.220	0.338	0.104
Interactive model	Patent value	Knowledge stock	0.223	0.227	0.289
		Technological scope	0.227	0.227	0.273
		International scope	0.228	0.165	0.234
		$Know.stock \times Int.scope$	0.008	0.007	0.013
		Technological usefulness	0.669	0.698	0.701
	Technological usefulness	Knowledge stock	0.185	0.309	0.106
		Technological scope	0.315	0.331	0.191
		International scope	0.148	0.247	0.223
		$Know.stock \times Int.scope$	-0.010	-0.045	-0.120
		R^2 of patent value	0.998	0.999	0.998
		R^2 of usefulness	0.220	0.341	0.120
		Effect size f^2 of interaction term on patent value	0.062	0.000	0.091
		F of interaction term on patent value	7.743	0.000	48.215
		p-value	(0.006)	(0.992)	(0.000)
		Effect size f^2 of interaction term on usefulness	0.000	0.002	0.004
		F of interaction term on usefulness	0.006	0.247	2.009
		p-value	(0.938)	(0.620)	(0.157)
Nonlinear model	Patent value	Knowledge stock	0.225	0.226	0.287
		Technological scope	0.227	0.225	0.272
		International scope	0.122	-0.175	-0.135
		$Int.Scope^2$	0.068	0.210	0.346
		Technological usefulness	0.669	0.701	0.701
	Technological usefulness	Knowledge stock	0.181	0.307	0.082
		Technological scope	0.315	0.336	0.205
		International scope	0.281	1.381	0.787
		$Int.Scope^2$	-0.086	-0.707	-0.547
		R^2 of patent value	0.999	1.000	0.999
		R^2 of usefulness	0.220	0.345	0.107
		Effect size f^2 of nonlinear term on patent value	0.307	2.999	1.180
		F of nonlinear term on patent value	38.121	368.852	626.792
		p-value	(0.000)	(0.000)	(0.000)
		Effect size f^2 of nonlinear term on usefulness	0.000	0.005	0.001
		F of nonlinear term on usefulness	0.029	0.634	0.375
		p-value	(0.865)	(0.427)	(0.540)

Path coefficients and pattern of significance of the linear additive models are the same as those presented in Chapter 7. Findings shows that the linear effects

Table 9.3: Pattern of significance of the structural relationships for the second-order models of patent value; ***, **, * and "." at the 0, 0.001, 0.01 and 0.05 significance level, respectively

Model	Endogenous Variable	Exogenous Variable	1990-1991	1995-1996	1999-2000
Linear additive model	Patent value	Knowledge stock	***	***	***
		Technological scope	***	***	***
		International scope	***	***	***
		Technological usefulness	***	***	***
	Technological usefulness	Knowledge stock	*	***	.
		Technological scope	***	***	***
		International scope	-	**	***
Interactive model	Patent value	Knowledge stock	***	***	***
		Technological scope	***	***	***
		International scope	***	***	***
		$Know.stock \times Int.scope$	**	**	**
		Technological usefulness	***	***	***
	Technological usefulness	Knowledge stock	*	***	*
		Technological scope	***	***	***
		International scope	-	**	***
		$Know.stock \times Int.scope$	-	-	**
Nonlinear model	Patent value	Knowledge stock	***	***	***
		Technological scope	***	***	***
		International scope	***	***	***
		$Int.Scope^2$	***	***	***
		Technological usefulness	***	***	***
	Technological usefulness	Knowledge stock	*	***	.
		Technological scope	***	***	***
		International scope	-	-	.
		$Int.Scope^2$	-	-	-

obtained with the different data sets are fairly stable, providing empirical evidence for the consistency of the linear additive model. The value of R^2 is 0.99 for patent value in the three analyzed time periods, and 0.22, 0.33 and 0.10 for technological usefulness in 1990-1991, 1995-1996 and 1999-2000, respectively. These values are very similar to those obtained in the interactive and nonlinear model. This would indicate at first glance that the moderate and nonlinear terms do not help much to explain the dependent variables, even though there are significant relationships between variables (Table 9.3). An extended discussion on the results for the linear additive models can be found in Chapter 7.

As expected, the linear effects remain in the interactive model. The relationships are also significant. The moderating effects of knowledge stock on the relationship between international scope and patent value are small, but significant, for the three analyzed time-periods. For patents applied for in 1990-2000, for instance, one standard deviation increase in knowledge stock will both impact patent value by 0.234 and increase the impact of international scope to patent value from 0.289 to 0.302. The interaction term (knowledge stock $\times$ international scope) has a small and significant effect size f^2 on patent value in 1990-1991 (f^2=0.06, F=7.74, p-value=0.006) and in 2000-2001 (f^2 = 0.09, F=48.21, p-value=0.00). There is no evidence of an effect size f^2 when estimating the model with data from 1995-1996. There is no evidence of a significant effect size of the interaction term on technological usefulness. Chin *et al.* (2003, p. 211) pointed out that "even a small interaction effect can be meaningful under extreme moderating conditions, if the resulting beta changes are meaningful, then it is important to take these conditions into account." So, findings would indicate that the effect of the international scope on the patent value is conditional on the value of the knowledge stock. Recall that the knowledge stock represents the base of knowledge that was used by the applicant to create an invention. It is formed by two manifest variables: number of inventors and number of backwards citations. It has been shown that both variables are significantly related to the knowledge stock and that the number of inventors contributes to the formation of the construct more than the backward citations do (Martínez-Ruiz & Aluja-Banet, 2009). On the other hand, the international scope refers to the geographic zones where protection is sought in the priority period. Recall that all patents have been granted in the U.S. So, in the proposed models, international scope is formed by two dummy variables indicating whether the invention has been protected in the priority period in Germany and Japan, two major technology-producing countries in the renewable energy field. When a company has invested a large stock of knowledge in the creation of a new technology, one can hypothesize that the invention will have a significant impact –and hence value– on the technological development of renewable energies. Therefore, the company would tend to quickly protect the invention in the major renewable energy producers. This may explain the moderating effect of the knowledge stock on the relationship between international scope and patent value. Another way to explain the effect is possible, if the variable number of inventors is closely observed. For patents in renewable energy, the mean, median and maximum value of the number of inventors is 2.22, 2 and 14, respectively. This means that approximately 50% of patents have more than two inventors (with a maximum of 14). Such inventors may work in different countries, which could encourage the search for wider geographical protection.

Significant linear effects also tend to remain in the nonlinear models; but, as expected, the relationship between international scope and patent value changes in the three time periods. Surprisingly, in the nonlinear models, the nonlinear term of the international scope has a large and significant effect on patent value in 1990-1991 (f^2 = 0.30, F=30.12, p-value=0.00), 1995-1996 (f^2 = 2.99, F=368.85, p-value=0.00) and 2000-2001 (f^2 = 1.18, F=626.79, p-value=0.00). There is no evidence of a significant effect of the nonlinear term on technological usefulness.

These empirical findings would confirm the importance of international scope as a determinant variable of patent value. These results are in line with those of Guellec & van Pottelsberghe (2000) who found a nonlinear relationship between the grant rate of patent applications and the number of designated states (countries where the invention is protected). It is worth noting that the samples used in this investigation are much smaller than those used by Guellec & van Pottelsberghe (2000) (N=22,911). This speaks well of the capabilities of the TsPLS procedure for estimating these types of relationships with small samples (N from 100 to 500).

9.7 Final Remarks

For the studied models, findings suggest that the TsPLS procedure offers a way to build "proper indices" for linear, nonlinear and interaction terms and to estimate the relationships between them. The estimates are always biased. The procedure shows a tendency to overestimate outer relationships and underestimate inner relationships. With respect to the accuracy and precision of the estimates according to the number of indicators and sample size, the results can be summarized as follows: the empirical evidence suggests that (1) for reflective relationships, the number of indicators (for instance, four or more) is a more important factor for accuracy than the sample size; estimates with small sample sizes prove to have an MRB that is less than 10%; (2) for formative relationships, the sample size (for instance, N=250) is a more important factor for accuracy than the number of indicators; estimates with a small number of indicators prove to have an MRB that is less than 10%; for structural relationships –that is to say linear, nonlinear and interaction effects– the number of indicators proved to be a more important factor for accuracy and precision of the estimates. So, even though accurate estimates can be obtained for formative exogenous relationships with few indicators per construct, a higher number of observed variables is desirable for obtaining accurate estimates of structural relationships.

On the other hand, the behavior of the Two-Step PLS Path Modelling Mode C procedure is found to be similar to the approaches studied by Henseler & Chin (2010) for structural models with reflective measurement models (PLS Path Modelling Mode A), at least in regards to the estimates of inner relationships. See, for instance, Figure 4 in the article by Henseler & Chin (2010, p. 98-100) and compare it with Figure 9.6 in this chapter. Henseler & Chin (2010, p. 107) said that "in our analysis, we focused on the interaction term alone, and did not include quadratic terms. Although quadratic terms have already been included in PLS path models (cf. Pavlou & Gefen, 2005), it remains unclear for researchers how this should be done." In this research, we have included nonlinear effects in the same way that interaction effects have been considered. That is, they have been calculated after obtaining linear scores and they are not standardized. Other alternatives may be explored in future research, however. In addition, the Two-Step PLS Path Modelling Mode C (TsPLS) procedure is a limited-information approach since nonlinear and interaction effects are not taken into account when calculating the linear terms. Therefore, comparing the results of this procedure with those obtained using a Wold's or hybrid approach should also be the foci of future research.

In terms of patent value models, there are a number of model variations that can be studied. For instance, the nonlinear effect of technological scope on the patent value may also be addressed. For this reason, rather than trying to find a significant effect, it seems advisable to look for a pattern of significant relationships using several samples, especially given the longitudinal nature of the patent value problem. A significant pattern provides strong evidence for the existence of relationships between variables.

9.8 Appendix: Tables

Table 9.4: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and two indicators per latent variable, 500 runs

N	Block	True Weights	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.8	0.681	0.681	0.646	0.712	-0.119	0.464	0.478	15
		0.5	0.476	0.476	0.438	0.508	-0.024	0.227	0.227	5
	2	0.4	0.359	0.359	0.310	0.409	-0.041	0.129	0.131	10
		0.8	0.648	0.648	0.602	0.692	-0.152	0.419	0.443	19
	3	0.1	0.250	0.250	0.206	0.298	0.150	0.062	0.085	-150
		0.9	0.738	0.738	0.688	0.779	-0.162	0.545	0.571	18
100	1	0.8	0.826	0.131	0.815	0.839	0.026	0.017	0.018	-3
		0.5	0.522	0.193	0.500	0.540	0.022	0.037	0.038	-4
	2	0.4	0.420	0.290	0.392	0.441	0.020	0.084	0.084	-5
		0.8	0.849	0.181	0.831	0.867	0.049	0.033	0.035	-6
	3	0.1	0.151	0.379	0.117	0.187	0.051	0.144	0.146	-51
		0.9	0.863	0.307	0.832	0.884	-0.037	0.095	0.096	4
250	1	0.8	0.841	0.079	0.831	0.850	0.041	0.006	0.008	-5
		0.5	0.523	0.120	0.510	0.533	0.023	0.014	0.015	-5
	2	0.4	0.432	0.176	0.418	0.447	0.032	0.031	0.032	-8
		0.8	0.882	0.091	0.875	0.890	0.082	0.008	0.015	-10
	3	0.1	0.113	0.257	0.089	0.140	0.013	0.066	0.066	-13
		0.9	0.958	0.060	0.952	0.963	0.058	0.004	0.007	-6
500	1	0.8	0.844	0.058	0.840	0.850	0.044	0.003	0.005	-6
		0.5	0.527	0.087	0.522	0.532	0.027	0.008	0.008	-5
	2	0.4	0.455	0.116	0.446	0.464	0.055	0.014	0.017	-14
		0.8	0.883	0.060	0.875	0.891	0.083	0.004	0.011	-10
	3	0.1	0.110	0.189	0.097	0.122	0.010	0.036	0.036	-10
		0.9	0.976	0.033	0.972	0.978	0.076	0.001	0.007	-8

Table 9.5: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and four indicators per latent variable, 500 runs

N	Block	True Weights	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.2	0.173	0.390	0.130	0.215	-0.027	0.152	0.153	13
		0.3	0.239	0.372	0.197	0.282	-0.061	0.139	0.142	20
		0.5	0.390	0.366	0.355	0.420	-0.110	0.134	0.146	22
		0.7	0.507	0.362	0.478	0.533	-0.193	0.131	0.168	28
	2	0.2	0.168	0.382	0.144	0.196	-0.032	0.146	0.147	16
		0.4	0.321	0.378	0.293	0.348	-0.079	0.143	0.149	20
		0.6	0.460	0.363	0.424	0.490	-0.140	0.132	0.151	23
		0.5	0.368	0.390	0.345	0.393	-0.132	0.152	0.170	26
	3	0.3	0.232	0.403	0.198	0.258	-0.068	0.163	0.167	23
		0.5	0.330	0.413	0.296	0.369	-0.170	0.171	0.200	34
		0.7	0.379	0.453	0.341	0.417	-0.321	0.205	0.308	46
		0.2	0.176	0.422	0.147	0.208	-0.024	0.178	0.179	12
100	1	0.2	0.205	0.244	0.181	0.232	0.005	0.060	0.060	-3
		0.3	0.291	0.234	0.270	0.314	-0.009	0.055	0.055	3
		0.5	0.496	0.225	0.471	0.520	-0.004	0.050	0.050	1
		0.7	0.656	0.203	0.639	0.674	-0.044	0.041	0.043	6
	2	0.2	0.199	0.318	0.183	0.221	-0.001	0.101	0.101	0
		0.4	0.329	0.320	0.288	0.358	-0.071	0.102	0.107	18
		0.6	0.528	0.308	0.508	0.546	-0.072	0.095	0.100	12
		0.5	0.444	0.306	0.422	0.462	-0.056	0.094	0.097	11
	3	0.3	0.186	0.370	0.150	0.225	-0.114	0.137	0.150	38
		0.5	0.365	0.380	0.331	0.406	-0.135	0.144	0.163	27
		0.7	0.483	0.386	0.457	0.511	-0.217	0.149	0.196	31
		0.2	0.195	0.387	0.161	0.234	-0.005	0.150	0.150	3
250	1	0.2	0.201	0.125	0.189	0.214	0.001	0.016	0.016	-1
		0.3	0.315	0.116	0.305	0.327	0.015	0.013	0.014	-5
		0.5	0.523	0.100	0.516	0.531	0.023	0.010	0.010	-5
		0.7	0.739	0.085	0.731	0.749	0.039	0.007	0.009	-6
	2	0.2	0.208	0.169	0.190	0.227	0.008	0.029	0.029	-4
		0.4	0.418	0.160	0.401	0.436	0.018	0.025	0.026	-5
		0.6	0.641	0.128	0.631	0.651	0.041	0.016	0.018	-7
		0.5	0.531	0.145	0.516	0.547	0.031	0.021	0.022	-6
	3	0.3	0.295	0.205	0.279	0.314	-0.005	0.042	0.042	2
		0.5	0.506	0.182	0.487	0.528	0.006	0.033	0.033	-1
		0.7	0.689	0.157	0.675	0.703	-0.011	0.025	0.025	2
		0.2	0.198	0.207	0.182	0.218	-0.002	0.043	0.043	1
500	1	0.2	0.214	0.087	0.206	0.225	0.014	0.007	0.008	-7
		0.3	0.319	0.078	0.311	0.327	0.019	0.006	0.006	-6
		0.5	0.525	0.077	0.516	0.533	0.025	0.006	0.006	-5
		0.7	0.747	0.060	0.742	0.751	0.047	0.004	0.006	-7
	2	0.2	0.213	0.112	0.202	0.222	0.013	0.013	0.013	-7
		0.4	0.432	0.106	0.424	0.440	0.032	0.011	0.012	-8
		0.6	0.657	0.092	0.650	0.669	0.057	0.008	0.012	-10
		0.5	0.546	0.101	0.537	0.554	0.046	0.010	0.012	-9
	3	0.3	0.316	0.151	0.302	0.329	0.016	0.023	0.023	-5
		0.5	0.511	0.135	0.500	0.521	0.011	0.018	0.018	-2
		0.7	0.719	0.108	0.707	0.733	0.019	0.012	0.012	-3
		0.2	0.220	0.157	0.207	0.231	0.020	0.025	0.025	-10

Table 9.6: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and six indicators per latent variable, 500 runs

N	Block	True Weights	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.5	0.316	0.343	0.279	0.363	-0.184	0.118	0.152	37
		0.3	0.214	0.337	0.174	0.261	-0.086	0.114	0.121	29
		0.4	0.283	0.344	0.234	0.321	-0.117	0.118	0.132	29
		0.3	0.200	0.336	0.168	0.237	-0.100	0.113	0.123	33
		0.5	0.318	0.347	0.285	0.355	-0.182	0.120	0.153	36
		0.1	0.101	0.339	0.069	0.133	0.001	0.115	0.115	-1
	2	0.2	0.166	0.352	0.141	0.191	-0.034	0.124	0.125	17
		0.4	0.262	0.344	0.219	0.297	-0.138	0.118	0.137	35
		0.6	0.388	0.342	0.341	0.432	-0.212	0.117	0.162	35
		0.4	0.266	0.332	0.235	0.293	-0.134	0.110	0.128	33
		0.2	0.139	0.325	0.109	0.170	-0.061	0.106	0.109	30
		0.3	0.223	0.341	0.199	0.252	-0.077	0.116	0.122	26
	3	0.3	0.161	0.365	0.127	0.194	-0.139	0.133	0.153	46
		0.6	0.273	0.406	0.234	0.314	-0.327	0.165	0.272	55
		0.2	0.146	0.365	0.108	0.186	-0.054	0.133	0.136	27
		0.3	0.193	0.374	0.147	0.227	-0.107	0.140	0.151	36
		0.4	0.220	0.392	0.172	0.257	-0.180	0.154	0.186	45
		0.2	0.144	0.364	0.110	0.188	-0.056	0.132	0.135	28
100	1	0.5	0.403	0.260	0.377	0.429	-0.097	0.068	0.077	19
		0.3	0.264	0.287	0.240	0.287	-0.036	0.083	0.084	12
		0.4	0.314	0.278	0.293	0.340	-0.086	0.077	0.085	22
		0.3	0.251	0.278	0.229	0.270	-0.049	0.077	0.080	16
		0.5	0.418	0.255	0.399	0.433	-0.082	0.065	0.072	16
		0.1	0.096	0.298	0.072	0.125	-0.004	0.089	0.089	4
	2	0.2	0.200	0.236	0.179	0.223	0.000	0.056	0.056	0
		0.4	0.366	0.230	0.350	0.388	-0.034	0.053	0.054	8
		0.6	0.547	0.202	0.526	0.563	-0.053	0.041	0.044	9
		0.4	0.369	0.220	0.347	0.394	-0.031	0.048	0.049	8
		0.2	0.194	0.240	0.180	0.213	-0.006	0.058	0.058	3
		0.3	0.257	0.236	0.239	0.276	-0.043	0.055	0.057	14
	3	0.3	0.218	0.326	0.185	0.242	-0.082	0.106	0.113	27
		0.6	0.398	0.338	0.363	0.426	-0.202	0.114	0.155	34
		0.2	0.130	0.338	0.107	0.155	-0.070	0.114	0.119	35
		0.3	0.210	0.325	0.180	0.234	-0.090	0.106	0.114	30
		0.4	0.267	0.350	0.235	0.300	-0.133	0.123	0.141	33
		0.2	0.151	0.347	0.110	0.196	-0.049	0.120	0.123	25

Table 9.7: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and six indicators per latent variable, 500 runs

N	Block	True Weights	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.5	0.527	0.100	0.519	0.536	0.027	0.010	0.011	-5
		0.3	0.307	0.115	0.298	0.318	0.007	0.013	0.013	-2
		0.4	0.419	0.106	0.408	0.430	0.019	0.011	0.012	-5
		0.3	0.314	0.116	0.306	0.324	0.014	0.014	0.014	-5
		0.5	0.523	0.107	0.514	0.534	0.023	0.012	0.012	-5
		0.1	0.107	0.128	0.095	0.121	0.007	0.016	0.016	-7
	2	0.2	0.213	0.152	0.200	0.226	0.013	0.023	0.023	-7
		0.4	0.398	0.141	0.390	0.406	-0.002	0.020	0.020	1
		0.6	0.612	0.125	0.602	0.626	0.012	0.016	0.016	-2
		0.4	0.418	0.146	0.407	0.430	0.018	0.021	0.022	-5
		0.2	0.200	0.159	0.182	0.218	0.000	0.025	0.025	0
		0.3	0.297	0.153	0.282	0.314	-0.003	0.023	0.023	1
	3	0.3	0.305	0.198	0.282	0.329	0.005	0.039	0.039	-2
		0.6	0.606	0.161	0.591	0.620	0.006	0.026	0.026	-1
		0.2	0.208	0.205	0.189	0.227	0.008	0.042	0.042	-4
		0.3	0.294	0.186	0.275	0.309	-0.006	0.035	0.035	2
		0.4	0.397	0.195	0.384	0.407	-0.003	0.038	0.038	1
		0.2	0.203	0.202	0.185	0.222	0.003	0.041	0.041	-1
500	1	0.5	0.537	0.075	0.532	0.543	0.037	0.006	0.007	-7
		0.3	0.316	0.081	0.309	0.325	0.016	0.007	0.007	-5
		0.4	0.429	0.078	0.422	0.435	0.029	0.006	0.007	-7
		0.3	0.322	0.082	0.314	0.330	0.022	0.007	0.007	-7
		0.5	0.529	0.075	0.522	0.535	0.029	0.006	0.007	-6
		0.1	0.106	0.081	0.100	0.113	0.006	0.007	0.007	-6
	2	0.2	0.207	0.115	0.198	0.214	0.007	0.013	0.013	-3
		0.4	0.416	0.106	0.407	0.425	0.016	0.011	0.012	-4
		0.6	0.630	0.092	0.621	0.637	0.030	0.009	0.009	-5
		0.4	0.422	0.102	0.412	0.430	0.022	0.010	0.011	-5
		0.2	0.208	0.112	0.199	0.215	0.008	0.013	0.013	-4
		0.3	0.321	0.112	0.310	0.335	0.021	0.013	0.013	-7
	3	0.3	0.326	0.152	0.314	0.337	0.026	0.023	0.024	-9
		0.6	0.631	0.127	0.618	0.643	0.031	0.016	0.017	-5
		0.2	0.207	0.152	0.195	0.224	0.007	0.023	0.023	-4
		0.3	0.316	0.150	0.298	0.333	0.016	0.023	0.023	-5
		0.4	0.429	0.143	0.413	0.443	0.029	0.020	0.021	-7
		0.2	0.209	0.154	0.193	0.224	0.009	0.024	0.024	-5

Table 9.8: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and eight indicators per latent variable, 500 runs

N	Block	True Weights	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	1	0.3	0.187	0.308	0.164	0.212	-0.113	0.095	0.108	38
		0.3	0.180	0.315	0.147	0.209	-0.120	0.099	0.114	40
		0.4	0.274	0.317	0.251	0.298	-0.126	0.101	0.117	31
		0.3	0.185	0.323	0.149	0.213	-0.115	0.104	0.117	38
		0.4	0.252	0.320	0.227	0.279	-0.148	0.102	0.124	37
		0.3	0.191	0.313	0.164	0.218	-0.109	0.098	0.110	36
		0.2	0.147	0.317	0.112	0.175	-0.053	0.101	0.103	27
		0.3	0.181	0.314	0.156	0.211	-0.119	0.098	0.113	40
	2	0.3	0.245	0.280	0.226	0.271	-0.055	0.078	0.081	18
		0.3	0.233	0.271	0.208	0.256	-0.067	0.074	0.078	22
		0.4	0.299	0.274	0.274	0.322	-0.101	0.075	0.085	25
		0.4	0.305	0.284	0.277	0.331	-0.095	0.081	0.090	24
		0.2	0.137	0.271	0.107	0.165	-0.063	0.073	0.077	31
		0.3	0.235	0.281	0.211	0.262	-0.065	0.079	0.083	22
		0.4	0.307	0.270	0.271	0.339	-0.093	0.073	0.082	23
		0.2	0.181	0.268	0.161	0.210	-0.019	0.072	0.072	10
	3	0.4	0.180	0.358	0.131	0.221	-0.220	0.128	0.177	55
		0.5	0.205	0.347	0.180	0.239	-0.295	0.120	0.207	59
		0.4	0.161	0.344	0.134	0.189	-0.239	0.119	0.176	60
		0.3	0.144	0.358	0.103	0.180	-0.156	0.128	0.152	52
		0.2	0.147	0.351	0.111	0.187	-0.053	0.123	0.126	26
		0.1	0.079	0.350	0.041	0.120	-0.021	0.123	0.123	21
		0.3	0.158	0.352	0.131	0.182	-0.142	0.124	0.144	47
		0.2	0.104	0.342	0.082	0.126	-0.096	0.117	0.126	48
100	1	0.3	0.226	0.262	0.203	0.252	-0.074	0.069	0.074	25
		0.3	0.242	0.262	0.224	0.261	-0.058	0.068	0.072	19
		0.4	0.312	0.277	0.285	0.335	-0.088	0.077	0.085	22
		0.3	0.222	0.263	0.203	0.243	-0.078	0.069	0.075	26
		0.4	0.314	0.241	0.290	0.340	-0.086	0.058	0.066	21
		0.3	0.251	0.270	0.230	0.276	-0.049	0.073	0.075	16
		0.2	0.184	0.264	0.160	0.210	-0.016	0.069	0.070	8
		0.3	0.215	0.266	0.189	0.240	-0.085	0.071	0.078	28
	2	0.3	0.247	0.219	0.231	0.272	-0.053	0.048	0.051	18
		0.3	0.254	0.211	0.237	0.266	-0.046	0.045	0.047	15
		0.4	0.371	0.213	0.354	0.388	-0.029	0.045	0.046	7
		0.4	0.359	0.213	0.341	0.377	-0.041	0.045	0.047	10
		0.2	0.182	0.213	0.158	0.207	-0.018	0.046	0.046	9
		0.3	0.272	0.222	0.250	0.300	-0.028	0.049	0.050	9
		0.4	0.364	0.214	0.341	0.388	-0.036	0.046	0.047	9
		0.2	0.191	0.229	0.169	0.213	-0.009	0.052	0.052	4
	3	0.4	0.248	0.300	0.227	0.272	-0.152	0.090	0.113	38
		0.5	0.294	0.300	0.274	0.321	-0.206	0.090	0.133	41
		0.4	0.254	0.304	0.221	0.296	-0.146	0.092	0.114	37
		0.3	0.178	0.301	0.151	0.209	-0.122	0.090	0.105	41
		0.2	0.122	0.292	0.098	0.139	-0.078	0.085	0.091	39
		0.1	0.089	0.313	0.065	0.115	-0.011	0.098	0.098	11
		0.3	0.175	0.306	0.155	0.201	-0.125	0.094	0.109	42
		0.2	0.117	0.321	0.090	0.145	-0.083	0.103	0.110	42

Table 9.9: True weights, mean weight estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and eight indicators per latent variable, 500 runs

N	Block	True Weights	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
250	1	0.3	0.324	0.118	0.315	0.335	0.024	0.014	0.015	-8
		0.3	0.307	0.113	0.297	0.316	0.007	0.013	0.013	-2
		0.4	0.431	0.111	0.416	0.442	0.031	0.012	0.013	-8
		0.3	0.323	0.114	0.312	0.334	0.023	0.013	0.014	-8
		0.4	0.416	0.115	0.406	0.427	0.016	0.013	0.014	-4
		0.3	0.314	0.118	0.304	0.325	0.014	0.014	0.014	-5
		0.2	0.214	0.123	0.205	0.224	0.014	0.015	0.015	-7
		0.3	0.324	0.113	0.316	0.335	0.024	0.013	0.013	-8
	2	0.3	0.302	0.152	0.288	0.316	0.002	0.023	0.023	-1
		0.3	0.313	0.150	0.291	0.333	0.013	0.022	0.023	-4
		0.4	0.395	0.136	0.385	0.407	-0.005	0.018	0.018	1
		0.4	0.400	0.134	0.388	0.413	0.000	0.018	0.018	0
		0.2	0.205	0.149	0.196	0.214	0.005	0.022	0.022	-3
		0.3	0.300	0.156	0.286	0.313	0.000	0.024	0.024	0
		0.4	0.400	0.143	0.386	0.416	0.000	0.020	0.020	0
		0.2	0.210	0.149	0.198	0.222	0.010	0.022	0.022	-5
	3	0.4	0.379	0.170	0.364	0.393	-0.021	0.029	0.029	5
		0.5	0.475	0.166	0.461	0.490	-0.025	0.027	0.028	5
		0.4	0.386	0.171	0.373	0.401	-0.014	0.029	0.030	4
		0.3	0.284	0.189	0.269	0.299	-0.016	0.036	0.036	5
		0.2	0.180	0.183	0.160	0.201	-0.020	0.033	0.034	10
		0.1	0.074	0.192	0.059	0.093	-0.026	0.037	0.038	26
		0.3	0.283	0.181	0.268	0.297	-0.017	0.033	0.033	6
		0.2	0.189	0.182	0.172	0.204	-0.011	0.033	0.033	5
500	1	0.3	0.332	0.082	0.322	0.338	0.032	0.007	0.008	-11
		0.3	0.319	0.084	0.314	0.325	0.019	0.007	0.007	-6
		0.4	0.428	0.075	0.423	0.434	0.028	0.006	0.006	-7
		0.3	0.325	0.077	0.318	0.330	0.025	0.006	0.006	-8
		0.4	0.434	0.076	0.428	0.441	0.034	0.006	0.007	-8
		0.3	0.330	0.082	0.322	0.338	0.030	0.007	0.008	-10
		0.2	0.217	0.085	0.210	0.226	0.017	0.007	0.007	-9
		0.3	0.327	0.086	0.318	0.337	0.027	0.007	0.008	-9
	2	0.3	0.316	0.108	0.307	0.325	0.016	0.012	0.012	-5
		0.3	0.313	0.108	0.303	0.327	0.013	0.012	0.012	-4
		0.4	0.415	0.106	0.408	0.423	0.015	0.011	0.012	-4
		0.4	0.425	0.097	0.416	0.433	0.025	0.009	0.010	-6
		0.2	0.213	0.104	0.204	0.222	0.013	0.011	0.011	-7
		0.3	0.314	0.105	0.301	0.328	0.014	0.011	0.011	-5
		0.4	0.421	0.102	0.410	0.433	0.021	0.011	0.011	-5
		0.2	0.212	0.101	0.203	0.222	0.012	0.010	0.010	-6
	3	0.4	0.409	0.132	0.394	0.424	0.009	0.017	0.017	-2
		0.5	0.510	0.131	0.502	0.519	0.010	0.017	0.017	-2
		0.4	0.399	0.131	0.383	0.414	-0.001	0.017	0.017	0
		0.3	0.298	0.137	0.282	0.309	-0.002	0.019	0.019	1
		0.2	0.206	0.145	0.194	0.218	0.006	0.021	0.021	-3
		0.1	0.103	0.146	0.084	0.122	0.003	0.021	0.021	-3
		0.3	0.293	0.140	0.283	0.301	-0.007	0.020	0.020	2
		0.2	0.199	0.139	0.186	0.213	-0.001	0.019	0.019	0

Table 9.10: True path coefficients, mean estimates of linear, nonlinear and interaction effects, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with two indicators per latent variable, 500 runs

N	Effect	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	Linear (1)	0.3	0.252	0.252	0.233	0.269	-0.048	0.063	0.066	16
	Linear (2)	0.2	0.159	0.159	0.145	0.175	-0.041	0.025	0.027	20
	Linear (3)	0.3	0.231	0.231	0.216	0.247	-0.069	0.053	0.058	23
	Nonlinear	0.3	0.129	0.129	0.114	0.141	-0.171	0.017	0.046	57
	Interaction	0.3	0.140	0.140	0.128	0.154	-0.160	0.020	0.045	53
100	Linear (1)	0.5	0.413	0.071	0.406	0.421	-0.087	0.005	0.013	17
	Linear (2)	0.4	0.310	0.081	0.303	0.320	-0.090	0.006	0.015	22
	Linear (3)	0.3	0.224	0.102	0.218	0.231	-0.076	0.010	0.016	25
	Nonlinear	0.3	0.178	0.065	0.172	0.185	-0.122	0.004	0.019	41
	Interaction	0.3	0.189	0.101	0.180	0.195	-0.111	0.010	0.022	37
250	Linear (1)	0.5	0.414	0.044	0.410	0.419	-0.086	0.002	0.009	17
	Linear (2)	0.4	0.308	0.047	0.305	0.312	-0.092	0.002	0.011	23
	Linear (3)	0.3	0.235	0.044	0.231	0.240	-0.065	0.002	0.006	22
	Nonlinear	0.3	0.199	0.037	0.195	0.203	-0.101	0.001	0.012	34
	Interaction	0.3	0.215	0.046	0.213	0.218	-0.085	0.002	0.009	28
500	Linear (1)	0.5	0.413	0.030	0.410	0.416	-0.087	0.001	0.008	17
	Linear (2)	0.4	0.315	0.033	0.313	0.318	-0.085	0.001	0.008	21
	Linear (3)	0.3	0.236	0.032	0.233	0.238	-0.064	0.001	0.005	21
	Nonlinear	0.3	0.205	0.024	0.203	0.208	-0.095	0.001	0.010	32
	Interaction	0.3	0.218	0.033	0.216	0.221	-0.082	0.001	0.008	27

Table 9.11: True path coefficients, mean estimates of linear, nonlinear and interaction effects, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with four indicators per latent variable, 500 runs

N	Effect	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	Linear (1)	0.3	0.294	0.156	0.280	0.309	-0.006	0.024	0.024	2
	Linear (2)	0.3	0.278	0.161	0.261	0.293	-0.022	0.026	0.026	7
	Linear (3)	0.2	0.189	0.194	0.171	0.206	-0.011	0.038	0.038	6
	Nonlinear	0.3	0.103	0.111	0.093	0.114	-0.197	0.012	0.051	66
	Interaction	0.3	0.077	0.147	0.066	0.086	-0.223	0.022	0.071	74
100	Linear (1)	0.4	0.355	0.081	0.349	0.362	-0.045	0.007	0.009	11
	Linear (2)	0.3	0.261	0.096	0.252	0.269	-0.039	0.009	0.011	13
	Linear (3)	0.2	0.181	0.117	0.169	0.194	-0.019	0.014	0.014	9
	Nonlinear	0.3	0.141	0.083	0.131	0.149	-0.159	0.007	0.032	53
	Interaction	0.3	0.137	0.117	0.124	0.151	-0.163	0.014	0.040	54
250	Linear (1)	0.5	0.440	0.042	0.436	0.444	-0.060	0.002	0.005	12
	Linear (2)	0.4	0.335	0.044	0.332	0.339	-0.065	0.002	0.006	16
	Linear (3)	0.3	0.264	0.042	0.260	0.269	-0.036	0.002	0.003	12
	Nonlinear	0.3	0.208	0.039	0.205	0.212	-0.092	0.001	0.010	31
	Interaction	0.3	0.220	0.048	0.214	0.224	-0.080	0.002	0.009	27
500	Linear (1)	0.5	0.448	0.027	0.445	0.450	-0.052	0.001	0.003	10
	Linear (2)	0.4	0.344	0.029	0.341	0.346	-0.056	0.001	0.004	14
	Linear (3)	0.3	0.267	0.030	0.265	0.270	-0.033	0.001	0.002	11
	Nonlinear	0.3	0.222	0.025	0.221	0.224	-0.078	0.001	0.007	26
	Interaction	0.3	0.235	0.033	0.233	0.239	-0.065	0.001	0.005	22

Table 9.12: True path coefficients, mean estimates of linear, nonlinear and interaction effects, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with six indicators per latent variable, 500 runs

N	Effect	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	Linear (1)	0.3	0.305	0.153	0.291	0.317	0.005	0.023	0.023	-2
	Linear (2)	0.3	0.292	0.169	0.270	0.311	-0.008	0.029	0.029	3
	Linear (3)	0.2	0.207	0.204	0.188	0.226	0.007	0.041	0.042	-4
	Nonlinear	0.3	0.086	0.108	0.079	0.095	-0.214	0.012	0.058	71
	Interaction	0.3	0.052	0.136	0.038	0.066	-0.248	0.018	0.080	83
100	Linear (1)	0.3	0.287	0.090	0.278	0.296	-0.013	0.008	0.008	4
	Linear (2)	0.4	0.354	0.084	0.346	0.364	-0.046	0.007	0.009	11
	Linear (3)	0.2	0.202	0.109	0.191	0.214	0.002	0.012	0.012	-1
	Nonlinear	0.3	0.153	0.075	0.147	0.159	-0.147	0.006	0.027	49
	Interaction	0.3	0.094	0.095	0.084	0.103	-0.206	0.009	0.051	69
250	Linear (1)	0.5	0.443	0.041	0.438	0.446	-0.057	0.002	0.005	11
	Linear (2)	0.4	0.342	0.043	0.337	0.346	-0.058	0.002	0.005	15
	Linear (3)	0.3	0.250	0.045	0.247	0.253	-0.050	0.002	0.005	17
	Nonlinear	0.3	0.210	0.038	0.206	0.214	-0.090	0.001	0.010	30
	Interaction	0.3	0.192	0.048	0.187	0.197	-0.108	0.002	0.014	36
500	Linear (1)	0.5	0.448	0.029	0.446	0.451	-0.052	0.001	0.004	10
	Linear (2)	0.4	0.348	0.028	0.345	0.351	-0.052	0.001	0.003	13
	Linear (3)	0.3	0.255	0.030	0.253	0.257	-0.045	0.001	0.003	15
	Nonlinear	0.3	0.228	0.028	0.225	0.230	-0.072	0.001	0.006	24
	Interaction	0.3	0.215	0.032	0.212	0.218	-0.085	0.001	0.008	28

Table 9.13: True path coefficients, mean estimates of linear, nonlinear and interaction effects, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with eight indicators per latent variable, 500 runs

N	Effect	True Path Coefficients	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	Linear (1)	0.3	0.302	0.151	0.286	0.318	0.002	0.023	0.023	-1
	Linear (2)	0.4	0.373	0.120	0.362	0.383	-0.027	0.014	0.015	7
	Linear (3)	0.2	0.220	0.191	0.202	0.242	0.020	0.036	0.037	-10
	Nonlinear	0.3	0.090	0.099	0.080	0.101	-0.210	0.010	0.054	70
	Interaction	0.3	0.048	0.120	0.036	0.062	-0.252	0.014	0.078	84
100	Linear (1)	0.3	0.297	0.077	0.290	0.303	-0.003	0.006	0.006	1
	Linear (2)	0.4	0.357	0.080	0.350	0.364	-0.043	0.006	0.008	11
	Linear (3)	0.2	0.226	0.110	0.215	0.238	0.026	0.012	0.013	-13
	Nonlinear	0.3	0.134	0.074	0.125	0.142	-0.166	0.005	0.033	55
	Interaction	0.3	0.078	0.092	0.069	0.089	-0.222	0.009	0.058	74
250	Linear (1)	0.5	0.434	0.042	0.431	0.439	-0.066	0.002	0.006	13
	Linear (2)	0.4	0.343	0.046	0.338	0.349	-0.057	0.002	0.005	14
	Linear (3)	0.3	0.268	0.045	0.264	0.272	-0.032	0.002	0.003	11
	Nonlinear	0.3	0.195	0.042	0.193	0.199	-0.105	0.002	0.013	35
	Interaction	0.3	0.191	0.051	0.187	0.196	-0.109	0.003	0.015	36
500	Linear (1)	0.5	0.447	0.029	0.444	0.450	-0.053	0.001	0.004	11
	Linear (2)	0.4	0.350	0.031	0.348	0.353	-0.050	0.001	0.003	12
	Linear (3)	0.3	0.268	0.029	0.266	0.271	-0.032	0.001	0.002	11
	Nonlinear	0.3	0.223	0.027	0.221	0.226	-0.077	0.001	0.007	26
	Interaction	0.3	0.218	0.031	0.215	0.221	-0.082	0.001	0.008	27

Table 9.14: True loadings, mean loadings estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and two indicators per latent variable, 500 runs

N	True Loadings	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.7	0.857	0.857	0.850	0.865	0.157	0.735	0.760	-22
	0.8	0.886	0.886	0.880	0.891	0.086	0.785	0.793	-11
100	0.7	0.857	0.039	0.855	0.860	0.157	0.002	0.026	-22
	0.8	0.893	0.026	0.890	0.896	0.093	0.001	0.009	-12
250	0.7	0.857	0.023	0.854	0.859	0.157	0.001	0.025	-22
	0.8	0.894	0.016	0.893	0.896	0.094	0.000	0.009	-12
500	0.7	0.859	0.016	0.857	0.860	0.159	0.000	0.025	-23
	0.8	0.894	0.012	0.893	0.896	0.094	0.000	0.009	-12

Table 9.15: True loadings, mean loadings estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and four indicators per latent variable, 500 runs

N	True Loadings	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.706	0.107	0.692	0.719	0.106	0.011	0.023	-18
	0.7	0.784	0.075	0.777	0.791	0.084	0.006	0.013	-12
	0.8	0.852	0.049	0.849	0.855	0.052	0.002	0.005	-6
	0.9	0.913	0.026	0.910	0.915	0.013	0.001	0.001	-1
100	0.6	0.714	0.064	0.707	0.719	0.114	0.004	0.017	-19
	0.7	0.791	0.046	0.787	0.794	0.091	0.002	0.010	-13
	0.8	0.856	0.032	0.853	0.859	0.056	0.001	0.004	-7
	0.9	0.919	0.016	0.917	0.920	0.019	0.000	0.001	-2
250	0.6	0.697	0.040	0.693	0.700	0.097	0.002	0.011	-16
	0.7	0.783	0.026	0.781	0.785	0.083	0.001	0.008	-12
	0.8	0.854	0.018	0.853	0.856	0.054	0.000	0.003	-7
	0.9	0.918	0.008	0.917	0.919	0.018	0.000	0.000	-2
500	0.6	0.702	0.027	0.700	0.705	0.102	0.001	0.011	-17
	0.7	0.786	0.019	0.784	0.788	0.086	0.000	0.008	-12
	0.8	0.855	0.012	0.854	0.856	0.055	0.000	0.003	-7
	0.9	0.919	0.006	0.919	0.920	0.019	0.000	0.000	-2

Table 9.16: True loadings, mean loadings estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and six indicators per latent variable, 500 runs

N	True Loadings	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.672	0.094	0.663	0.682	0.072	0.009	0.014	-12
	0.7	0.756	0.066	0.752	0.761	0.056	0.004	0.007	-8
	0.8	0.832	0.051	0.829	0.837	0.032	0.003	0.004	-4
	0.9	0.919	0.022	0.916	0.921	0.019	0.001	0.001	-2
	0.6	0.742	0.073	0.736	0.747	0.142	0.005	0.026	-24
	0.7	0.821	0.050	0.816	0.825	0.121	0.002	0.017	-17
100	0.6	0.664	0.066	0.656	0.671	0.064	0.004	0.008	-11
	0.7	0.755	0.051	0.750	0.759	0.055	0.003	0.006	-8
	0.8	0.833	0.033	0.830	0.836	0.033	0.001	0.002	-4
	0.9	0.920	0.016	0.919	0.922	0.020	0.000	0.001	-2
	0.6	0.746	0.052	0.740	0.752	0.146	0.003	0.024	-24
	0.7	0.826	0.033	0.823	0.828	0.126	0.001	0.017	-18
250	0.6	0.663	0.041	0.660	0.666	0.063	0.002	0.006	-10
	0.7	0.754	0.028	0.752	0.757	0.054	0.001	0.004	-8
	0.8	0.828	0.020	0.826	0.830	0.028	0.000	0.001	-3
	0.9	0.920	0.009	0.919	0.921	0.020	0.000	0.000	-2
	0.6	0.740	0.033	0.738	0.742	0.140	0.001	0.021	-23
	0.7	0.820	0.020	0.818	0.821	0.120	0.000	0.015	-17
500	0.6	0.660	0.028	0.657	0.662	0.060	0.001	0.004	-10
	0.7	0.751	0.021	0.749	0.753	0.051	0.000	0.003	-7
	0.8	0.831	0.014	0.829	0.832	0.031	0.000	0.001	-4
	0.9	0.920	0.006	0.920	0.921	0.020	0.000	0.000	-2
	0.6	0.741	0.021	0.739	0.743	0.141	0.000	0.020	-23
	0.7	0.822	0.015	0.821	0.823	0.122	0.000	0.015	-17

Table 9.17: True loadings, mean loadings estimates, standard deviations, confidence intervals, biases, variances, mean square errors and mean relative bias (%) for models with nonlinearities and eight indicators per latent variable, 500 runs

N	True Loadings	Mean	S.D.	L.Bound	U.Bound	Bias	Var	MSE	MRB (%)
50	0.6	0.639	0.103	0.628	0.649	0.039	0.011	0.012	-7
	0.7	0.737	0.071	0.731	0.745	0.037	0.005	0.006	-5
	0.8	0.821	0.049	0.816	0.825	0.021	0.002	0.003	-3
	0.9	0.931	0.019	0.929	0.933	0.031	0.000	0.001	-3
	0.6	0.729	0.069	0.724	0.733	0.129	0.005	0.021	-21
	0.7	0.812	0.051	0.809	0.816	0.112	0.003	0.015	-16
	0.8	0.879	0.033	0.876	0.881	0.079	0.001	0.007	-10
	0.9	0.932	0.018	0.930	0.934	0.032	0.000	0.001	-4
100	0.6	0.643	0.066	0.638	0.646	0.043	0.004	0.006	-7
	0.7	0.739	0.050	0.734	0.744	0.039	0.002	0.004	-6
	0.8	0.823	0.034	0.821	0.825	0.023	0.001	0.002	-3
	0.9	0.932	0.012	0.930	0.933	0.032	0.000	0.001	-4
	0.6	0.726	0.050	0.721	0.730	0.126	0.003	0.018	-21
	0.7	0.813	0.037	0.811	0.815	0.113	0.001	0.014	-16
	0.8	0.880	0.023	0.878	0.882	0.080	0.001	0.007	-10
	0.9	0.932	0.012	0.931	0.933	0.032	0.000	0.001	-4
250	0.6	0.639	0.043	0.634	0.642	0.039	0.002	0.003	-6
	0.7	0.728	0.033	0.726	0.731	0.028	0.001	0.002	-4
	0.8	0.819	0.020	0.817	0.820	0.019	0.000	0.001	-2
	0.9	0.930	0.007	0.930	0.931	0.030	0.000	0.001	-3
	0.6	0.722	0.030	0.719	0.725	0.122	0.001	0.016	-20
	0.7	0.809	0.021	0.807	0.810	0.109	0.000	0.012	-16
	0.8	0.879	0.013	0.878	0.880	0.079	0.000	0.006	-10
	0.9	0.931	0.007	0.931	0.932	0.031	0.000	0.001	-3
500	0.6	0.638	0.030	0.636	0.641	0.038	0.001	0.002	-6
	0.7	0.733	0.022	0.731	0.735	0.033	0.001	0.002	-5
	0.8	0.820	0.014	0.819	0.821	0.020	0.000	0.001	-3
	0.9	0.931	0.005	0.930	0.932	0.031	0.000	0.001	-3
	0.6	0.721	0.024	0.719	0.724	0.121	0.001	0.015	-20
	0.7	0.810	0.015	0.809	0.812	0.110	0.000	0.012	-16
	0.8	0.878	0.009	0.877	0.879	0.078	0.000	0.006	-10
	0.9	0.931	0.005	0.931	0.932	0.031	0.000	0.001	-3

Table 9.18: True values, mean estimates, mean absolute deviations, biases, mean square errors, and mean relative bias (%) for weights, path coefficients and loadings, Chin et al.s' results (2003)

MVs per LV	**Sample Size**	**Path Coefficient**	**True Value**	**Mean**	**S.D.**	**Bias**	**MSE**	**MRB (%)**[a]
2	20	$x - y$	0.3	0.186	0.276	-0.114	0.089	38.00
		$z - y$	0.5	0.330	0.286	-0.170	0.111	34.00
		$x * z - y$	0.3	0.162	0.352	-0.138	0.143	46.00
	50	$x - y$	0.3	0.195	0.187	-0.105	0.046	35.00
		$z - y$	0.5	0.326	0.218	-0.174	0.078	34.80
		$x * z - y$	0.3	0.172	0.236	-0.128	0.072	42.67
	100	$x - y$	0.3	0.208	0.130	-0.092	0.025	30.67
		$z - y$	0.5	0.326	0.195	-0.174	0.068	34.80
		$x * z - y$	0.3	0.169	0.181	-0.131	0.050	43.67
	150	$x - y$	0.3	0.256	0.091	-0.044	0.010	14.67
		$z - y$	0.5	0.382	0.143	-0.118	0.034	23.60
		$x * z - y$	0.3	0.256	0.12	-0.044	0.016	14.67
	200	$x - y$	0.3	0.199	0.120	-0.101	0.025	33.67
		$z - y$	0.5	0.328	0.183	-0.172	0.063	34.40
		$x * z - y$	0.3	0.176	0.143	-0.124	0.036	41.33
	500	$x - y$	0.3	0.198	0.109	-0.102	0.022	34.00
		$z - y$	0.5	0.328	0.176	-0.172	0.061	34.40
		$x * z - y$	0.3	0.165	0.142	-0.135	0.038	45.00
4	20	$x - y$	0.3	0.215	0.250	-0.085	0.063	28.33
		$z - y$	0.5	0.334	0.264	-0.166	0.070	33.20
		$x * z - y$	0.3	0.250	0.370	-0.050	0.137	16.67
	50	$x - y$	0.3	0.232	0.153	-0.068	0.023	22.67
		$z - y$	0.5	0.386	0.159	-0.114	0.025	22.80
		$x * z - y$	0.3	0.274	0.186	-0.026	0.035	8.67
	100	$x - y$	0.3	0.256	0.091	-0.044	0.008	14.67
		$z - y$	0.5	0.382	0.143	-0.118	0.020	23.60
		$x * z - y$	0.3	0.256	0.120	-0.044	0.014	14.67
	150	$x - y$	0.3	0.245	0.086	-0.055	0.007	18.33
		$z - y$	0.5	0.397	0.122	-0.103	0.015	20.60
		$x * z - y$	0.3	0.242	0.1	-0.058	0.010	19.33
	200	$x - y$	0.3	0.243	0.081	-0.057	0.007	19.00
		$z - y$	0.5	0.397	0.118	-0.103	0.014	20.60
		$x * z - y$	0.3	0.242	0.082	-0.058	0.007	19.33
	500	$x - y$	0.3	0.242	0.069	-0.058	0.005	19.33
		$z - y$	0.5	0.396	0.11	-0.104	0.012	20.80
		$x * z - y$	0.3	0.222	0.087	-0.078	0.008	26.00

[a] Data are reproduced from the Chin et al.s' paper (2003, p. 204); bias, MSE and MRB are computed based on reported results.

Chapter 10

Summary of Conclusions and Future Research

10.1 Summary of conclusions and author's contributions

Two general goals were raised in this thesis: First, to investigate causality relationships among variables that determine the value of patents, by establishing structural and measurement models for patent value; second, to investigate the performance of PLS Path Modelling with Mode C in the context of patent value models. In this thesis, each chapter is an independent study. While the first chapters focus on giving a theoretical background on the several topics involved in this research, Chapters 7, 8 and 9 focus on the first objective; and Chapters 6 and 9 on the second. Conclusions and author's contributions can be summarized as follows.

10.1.1 Patent Value

In an exploratory way, patent value models have been established and formulated in increasing complexity. They have attempted to incorporate theoretical complexities of technological change and patent data in a multidimensional approach toward the patent value problem. In what follows, we describe the chief contributions to the field of Patent Value.

1. The following structural equations models were proposed (Chapter 7, 8 and 9):

    - Patent value as a first-order model.
    - Patent value as a second-order model.
    - Patent value models with longitudinal latent variables.

- Patent value models with longitudinal manifest variables.
- Patent value models including nonlinearities.

2. Patent value models include constructs and latent variables, each measured by a set of indicators. Some of the variables were collected through an extensive review of the literature and others were proposed, as in the case of technological usefulness; relationships between unobserved variables and each block of manifest variables are presented for consideration. In the first-order model of patent value, constructs, indicators and relationships were defined as follows:

    - Knowledge stock represents the base of knowledge that was used by the applicant to create an invention. This existing knowledge encourages the inventive activity and may come from within or outside the company. The number of applicants, the number of inventors and the backward citations form this construct.
    - Technological scope is related to the potential utility of an invention in some technological fields. Manifest variables for this construct are the number of IPC classes where the patent is classified, along with the number of claims of the patents. A formative relationship is modeled between indicators and construct.
    - International scope refers to the geographic zones where the invention is protected during the priority period. We defined two dummy variables that consider whether the invention had been protected in Japan or in Germany, two of the major producers of renewable energies. Formative relationships are considered in this block of variables.
    - Patent value will be reflected in the number of times that the patent is cited and in the patenting strategy pursued by the company over time. Mainly, the number of forward citations and the family size reflect this construct.

    The first-order model considers patent value as an endogenous latent variable depending on the knowledge stock, technological scope and international scope. The exogenous constructs give an *a priori* value of patents. Thus, the intrinsic characteristics of the patents at the time of its application, along with the patenting strategy of the company in the priority period, may give a preliminary idea of patent value. In contrast, patent value estimated through forward citations and family size gives an *a posteriori* value for patents. This value is obtained over time and is given by others. Conceptualizing the patent value as a potential and a recognized value of intangible assets is also a contribution of this thesis.

3. The second-order model considers that patent value is jointly given by those variables that determine the *a priori* and *a posteriori* patent value. Hence, the indicators that were initially related to the patent value are associated with a fifth underlying latent variable related to the potential usefulness of patents. We called this latent variable "technological usefulness." In this case, patent value is considered as a second-order latent variable that is influenced by all of other constructs in a second-order model.

4. While exploring the first- and second-order models of patent value, a stable pattern of path coefficients was found across samples in different time periods (Chapter 7). This provides evidence on the role of the knowledge stock, technological scope and international scope as determinants of patent value and technological usefulness. There are two particularly important variables: the technological and the international scope. However, a strong relationship was found between technological usefulness and patent value across samples. Thus, isolating the *a priori* (potential) and *a posteriori* components can reveal the extent to which each contributes to the patent value. These results were confirmed when the longitudinal models of patent value were estimated. It was shown that the potential value of patents is small compared to the value that is given later –although maybe useful for detecting valuable patents at an early stage. The stability of results when estimating the models with different data sets provides evidence for the replicability of the models, contributing to the systematization of intangible assets measurement.

5. The relationships that are found are reliable because samples are controlled (Chapter 3). It is known that there are a number of factors that may affect patent value. Among others, considerations must be made in relation to:

    - The technological area where the patents are protected. We considered the renewable energy technologies.
    - The country where patents are applied for and granted. We considered patents applied for and granted in the U.S.
    - The type of patent applicant. We considered companies.
    - Grant and application years. We organized the data by application year.

    There is an element which introduces heterogeneity in the patent samples. Companies protecting renewable energy technologies belong to different industries. At least, this is true if we consider the Standard Industrial Classification (SIC) code of each company in the U.S. However, this seems to be a characteristic of this sector with which it has to deal. Other researchers have encountered this same problem (Dechezleprêtre *et al.*, 2009).

6. The estimation results of the patent value models with longitudinal data suggest that the contribution of the knowledge stock used by companies to create their inventions, the technological scope of the inventions and the international scope of protection are variables that contribute little to the patent value when compared to technological usefulness (Chapter 8). In the first patent value model, technological usefulness is measured by time-dependent manifest variables. Thus, we average the contribution of the longitudinal indicators, and the patent value is modeled as an endogenous latent variable formed by the weighted contribution of the predictors. Hence, this model gives an overall measure of the patent value. To the best of our knowledge, Lanjouw & Schankerman (2004) are the only researchers who have used a similar approach. They remarked that the use of a composite index for patent value reduces variability in the unobservable construct, and that the latter is most useful when "one averages –either the mean over time for a given firm or the mean over firms for a given year."

   In the second patent value model, technological usefulness is modeled as a time-dependent latent variable. Patent value is also formed by the weighted sum of all the constructs, but now the model allows for the analysis of changes in the technological usefulness over different time periods. The estimates obtained for loadings in the first model and path coefficients in the second allow us to observe how the patent value increases, stabilizes and then decreases over time. We provide empirical evidence for the importance of considering the longitudinal nature of the indicators in the patent value problem, especially for forward citations, which are the most widely used indicator of patent value.

7. We were successful in finding a pattern of interaction and nonlinear effects between studied variables across different patent data sets. For interaction effects, we have found a small and significant moderating effect of knowledge stock on the relationship between international scope and patent value. Exploratory analysis shows that international scope has a nonlinear effect on patent value. Thus, patents increase in value if applicants seek to protect the invention in Japan or Germany. Recall that all patents are applied for and granted in the U.S. However, we think that more research and evidence is needed to interpret the results appropriately.

8. To introduce a multidimensional perspective of the patent valuation problem. To the best of our knowledge, the use of structural models with unob-

served variables in the field of technological change has been scarce. This research proposes that a holistic and multidimensional model may offer a robust understanding of the different variables that determine patent value. The models are strongly based on the theory developed by the technological change scientific community and a thorough review of the literature on patent valuation studies.

10.1.2 PLS Path Modelling

In this thesis, the performance of PLS Path Modelling was deeply studied in the context of the proposed models. The procedure aims to compute latent variable or construct scores specified in PLS-models or component-based models. Herman Wold originally defined three modes for the PLS algorithm. So, three PLS procedures may be identified:

- PLS Path Modelling Mode A for SEMs with reflective outer models.
- PLS Path Modelling Mode B for SEMs with formative outer models.
- PLS Path Modelling Mode C for SEMs with formative and reflective outer models.

Our chief contributions to the field of Partial Least Squares Path Modelling are comprised of:

1. Empirical evidence on the performance of PLS Path Modelling with Mode C. Several empirical studies have been made previously regarding this issue. However, results have been inconclusive. If properly implemented, PLS Path Modelling can adequately capture some of the complex dynamic relationships involved in models. As seen in Chapter 6, our research through a careful simulation study shows that PLS Path Modelling with Mode C performs according to the theoretical framework established for PLS procedures and PLS-models (Wold, 1982; Krämer, 2006; Hanafi, 2007; Dijkstra, 2010). That is:
   - Inner relationships are underestimated.
   - Outer relationships are overestimated in reflective outer models.
   - Outer relationships are overestimated in formative outer models for N greater than 100 if manifest variables are correlated.
   - Outer relationships are underestimated in formative outer models if manifest variables are uncorrelated.

- Outer relationships are underestimated in formative outer models if no errors and disturbance terms are considered and manifest variables are uncorrelated.
- As seen in Figure 6.5(b), by increasing the sample size, there is no observation of a reduction in bias in estimates of outer relationships estimated using Mode A. This is in line with the lack of monotone convergence property of Mode A.
- Alternatively, the consistency of Mode B is shown. This is in line with the monotone convergence property of Mode B.

2. Empirical evidence for the consistency at large of PLS Path Modelling with Mode A (see Chapter 6, especially Figure 6.5(b)). Even though Mode A lacks the monotone convergence property, the bias of the estimates of outer relationships are closer and closer to zero when the number of indicators per latent variable increases. Herman Wold calls this property "consistency at large" (Wold, 1982) and suggests that PLS Path Modelling with Mode A may be a robust alternative when estimating structural equation models with reflective outer models.

3. Empirical evidence for formative outer models with few manifest variables. The estimates obtained when considering two indicators per construct were found to have a bias of less than 20%. A greater number of indicators, however, contributes to the estimates of inner relationships that are less biased.

4. In Chapter 6, a simulation study which considered three set-ups. Case A considered that covariances among manifest variables, disturbance terms in the inner relationships and errors in outer relationships are zero. Case B relaxes the assumption that disturbance terms and errors are zero, and case C also relaxes the assumption that manifest variables are uncorrelated. Case C is the most general case. All simulated cases considered that manifest variables in formative outer models are a census of the variables that form the constructs. Theory must support a proper definition of them. Therefore, more research is needed to study the effects of measurement errors in formative blocks of variables on minimization of residuals in the structural relationships.

5. Empirical evidence on the performance of a Two-Step PLS Path Modelling (TsPLS) with Mode C to estimate nonlinear and interaction effects among formative constructs. Findings suggest that the TsPLS procedure offers a way to build proper indices for linear, nonlinear and interaction terms and

to estimate the relationships between them. The yield found for PLS Path Modelling Mode C tends to remain in TsPLS. This procedure preserves the aforementioned monotone convergence properties for the PLS basic design. That is, TsPLS Mode A is consistent at large, TsPLS Mode B is consistent and consistent at large.

10.2 Limitations of the Study and Future Research

The valuation of patents, and hence technology, is a fascinating subject that continues to attract the attention of researchers, especially in these times of revolutionary emergence in new technologies and breakthroughs in many fields. This thesis is original in many ways. We have conducted an empirical study not carried out before, we have applied a novel technique in an area now of particular interest, we have provided knowledge in a novel way and, in particular, this has been a multidisciplinary study. This research fills a gap in the literature on patent value and PLS research, considering the former problem from a multidimensional perspective, and clarifying critical aspects of the PLS procedures. However, there are some issues that can be addressed in future research.

1. Patent value may be seen as a complex construct depending on a variety of elements. Our intuition tells us that probably there are a number of other factors that may be affecting patent value. General and specific market conditions, countries' legal frameworks, geographic proximity or accumulated scientific and technological knowledge are different dimensions that may be included in a structural equation model and further explored. Based on patent indicators, the proposed models are a first attempt to define a SEM for patent value.

2. The conclusions of this research are inferred from data sets. The sample comprises a total of 2,901 patents granted in the U.S in the field of renewable energy and published in 1990-1991, 1995-1996, 1999-2000 and 2005-2006. Our research in progress involves the estimation of the models with the population, that is with all patents granted in the field of renewable energy in the U.S.. These renewable energy patents include wind, solar, geothermal, wave/tide, biomass and waste energy.

3. The proposed models are based on the theory of technological change, data obtained from patent documents and patent databases. Some research in the field of technological change has involved validation of the model through a survey of technical experts on intellectual property, for instance. Our research suffers from this empirical validation and this is a compelling topic

for future research. However, we were looking to measure patent value in a way that is easily systematized, by taking advantage of information technologies available today.

4. In this research, attempts were made to relate patent value to the market value of companies. Even though we obtained promising results, they have not been included in this document. There are several reasons for this. One of them is the patent that was applied for in the U.S. Patent and Trademark Office. This is an issue for our research in progress.

PLS procedures are being increasingly used and –it is our impression that– the complexity of the applications are ahead of theoretical developments. This research generated useful preliminary findings. However, there are some questions that remain open and can be further explored in future research.

1. Our knowledge of the procedures for estimating second-order or higher-order models has to be expanded. The proposal made by Lohmöller to repeat the manifest variables for the higher-order construct has proved to be useful in estimating second-order models. However, this approach has an important drawback. As estimated by repeating the observed variables of the outer models, the coefficient of multiple determination R^2 obtained for the endogenous second-order construct is probably not a good indicator for assessing either the model fit or the quality of the model. To resolve this issue, more research is needed on:

    - The definition of a criterion for selecting variables to be repeated for the endogenous construct.
    - The determination of an alternative way to estimate higher-order models.
    - The development of an alternative index for assessing the overall fit of model to data.

2. The estimation of longitudinal models in Chapter 8 has followed a traditional approach and has been exploratory. In our view, PLS is a limited-information approach for studying longitudinal problems, since the autoregressive nature of longitudinal variables is not explicitly taken into account. Hence, the procedure does not consider all available information in the data. Thus, longitudinal PLS represents a wide field of research with many interesting results and many unsolved problems.

3. TsPLS has proven to be useful for estimating nonlinear and interaction effects. Although Henseler & Chin (2010) found that Wold's or hybrid ap-

proach "does not excel in any of the examined categories i.e. parameter accuracy, statistical power, and prediction accuracy" when estimating SEMs with reflective outer models, future research should consider comparing TsPLS with Wold's procedure, in view of the fact that both techniques can be used to estimate nonlinearities in SEMs with formative outer models.

4. In this research, we limited simulation studies to normal independent data. However, other conditions can be introduced and tested, such as: non-normal variables, multicollinearity among indicators of formative measurement models and multicollinearity between constructs. These tasks remain outstanding for **PLS-models**.

References

Ahlgren, P., Jarneving, B., & Rousseau, R. 2003. Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, **54**(6), 550–560.

Aigner, D., Hsiao, C., Kapteyn, A., & Wansbeek, T. 1987. Latent variable models in econometrics. *In:* Griliches, Z., & Intriligator, M. (Eds.), *Handbooks of econometrics* (pp. 1320–1393), vol. II. Los Angeles: Elsevier.

Aigner, D. J., & Goldberger, A. S. 1977. *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Albert, M. B., Avery, D., Narin, F., & McAllister, P. 1991. Direct validation of citation counts as indicators of industrially important patents. *Research Policy*, **20**(3), 251–259.

Allison, P. D. 1977. Testing for interaction in multiple regression. *The American Journal of Sociology*, **83**(1), 144–153.

Almeida, M., Coelho, P., & Vilares, M. 2007. Comparison of likelihood and PLS estimators for structural equation modeling: A simulation study with customer satisfaction data. *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and related methods: Proceedings of the PLS'07 international symposium* (pp. 44–47). As, Norway: Matforsk.

Aluja, T., & Sánchez, G. 2007. A Simulation Study of PATHMOX (PLS–PM Segmentation Tree) Sensitivity. *In: ESSEC-HEC Research Workshop on PLS Path Modeling in the Presence of Group Structure*. Cergy-Pontoise, 15 May.

Anderson, J. C., & Gerbing, D. W. 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, **103**(3), 411–423.

Anderson, T. W., & Rubin, H. 1956. Statistical inference in factor analysis. *In:* Neyman, J. (Ed.), *Proceedings of the Third Berkeley Symposium for Mathe-*

matical Statistics Problems (pp. 111–150), vol. 5. Berkeley, CA: University of California Press.

Areskoug, B. 1982. The first canonical correlation: Theoretical PLS analysis and simulation experiments. *In:* Jöreskog, K. G. & Wold, H. (Ed.), *Systems under indirect observation, part II* (pp. 96–117). Amsterdam: North–Holland.

Astebro, T. 2004. Key success factors for technological entrepreneurs' R&D projects. *IEEE Transactions on Engineering Management*, **51**(3), 314–321.

Baffi, G., Martin, E. B., & Morris, A. J. 1999a. Non–linear projection to latent structures revisited: The neural network PLS algorithm. *Computers and Chemical Engineering*, **23**, 1293–1307.

Baffi, G., Martin, E. B., & Morris, A. J. 1999b. Non–linear projection to latent structures revisited: The quadratic PLS algorithm. *Computers and Chemical Engineering*, **23**, 395–411.

Bagozzi, R. P. 1980. *Causal models in marketing.* New York: Wiley & Sons.

Bagozzi, R. P. 1984. A prospectus for theory construction in marketing. *Journal of Marketing*, **48**(1), 11.

Balachandra, R., & Friar, J. H. 1997. Factors for success in R&D projects and new product innovation: A contextual framework. *IEEE Transactions on Engineering Management*, **44**(3), 276–287.

Barclay, D., Higgins, C., & Thompson, R. 1995. The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies*, **2**(2), 285–324.

Barker, M., & Rayens, W. 2003. Partial least squares for discrimination. *Journal of Chemometrics*, **17**, 166–173.

Baron, R., & Kenny, D. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, **51**(6), 1173–1182.

Bastien, P., Esposito Vinzi, V., & Tenenhaus, M. 2005. PLS generalised linear regression. *Computational Statistics & Data Analysis*, **48**(1), 17–46.

Batista, J., & Coenders, G. 2000. *Modelos de ecuaciones estructurales* (Vol. 6). Madrid: La Muralla.

Bayer, A., Smart, J., & McLaughlin, G. 1990. Mapping intellectual structure of a scientific subfield through author citations. *Journal of the American Society for Information Science*, **41**(6), 444–452.

Bentler, P. M. 1980. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, **31**, 419–456.

Bentler, P. M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin/Quantitative Methods in Psychology*, **107**(2), 238–246.

Bentler, P. M., & Dudgeon, P. 1996. Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, **47**, 563–592.

Blalock, H. 1964. *Causal inferences in non-experimental research.* Chapel Hill: University of North Carolina Press.

Bohrnstedt, G. W., & Marwell, G. 1978. The reliability of products of two random variables. *Sociological Methodology*, **9**, 254–273.

Bollen, K. 1995. Structural equation models that are nonlinear in latent variables: A least-squares estimator. *Sociological Methodology*, **25**, 223–251.

Bollen, K., & Lennox, R. 1991. Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, **110**(2), 305–314.

Bollen, K., & Ting, K.F. 1993. Confirmatory tetrad analysis. *Sociological Methodology*, **23**, 147–175.

Bollen, K., & Ting, K.F. 1998. Bootstrapping a test statistic for vanishing tetrads. *Sociological Methods & Research*, **27**(1), 77–102.

Bollen, K., & Ting, K.F. 2000. A tetrad test for causal indicators. *Psychological Methods*, **5**(1), 3–22.

Bollen, K. A. 1989. *Structural equations with latent variables.* New York: John Wiley & Sons.

Bookstein, F. 1982. The geometric meaning of soft modeling with some generalizations. *In:* Jöreskog, G., & Wold, H. (Eds.), *Systems under indirect observation, Part II* (pp. 55–74). Amsterdam: North–Holland.

Browne, M. W. 1977. Generalized least-squares estimators in the analysis of covariance structures. *In:* Aigner, D. J., & Goldberger, A. (Eds.), *Latent variables in socio–economic models* (chapter 13). New York: North-Holland.

Bucic, T., & Gudergan, S. 2004. The impact of organizational setting on creativity and learning in alliances. *Management*, **7**(3), 257–273.

Busemeyer, J., & Jones, L. 1983. Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychollogical Bulletin*, **93**(3), 549–562.

Caballero, R. J., & Jaffe, A. B. 2002. How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. *In:* Jaffe, A. B., & Trajtenberg, M. (Eds.), *Patents, citations and innovations: A window on the knowledge economy*. London: Massachussetts Institute of Technology.

Carpenter, M. P., Narin, F., & Woolf, P. 1981. Citation rates to technologically important patents. *World Patent Information*, **3**(4), 160.

Cassel, C., Hackl, P., & Westlund, A. 1999. Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, **26**(4), 435–446.

Cassel, C., Hackl, P., & Westlund, A. 2000. On measurement of intangible assets: A study of robusteness of partial least squares. *Total Quality Management*, **11**(7), 897–907.

Chin, W. W. 1995. Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, **2**, 315–319.

Chin, W. W. 1998a. The partial least squares approach for structural equation modeling. *In:* Marcoulides, G. A. (Ed.), *Modern methods for business research* (pp. 295–336). Mahwah, NJ: Erlbaum.

Chin, W. W. 2003. A permutation procedure for multi–group comparison of PLS models. *In:* Vilares, M. J., Tenenhaus, M., Coelho, P. S., Esposito Vinzi, V., & Morineau, A. (Eds.), *PLS and Related Methods: Proceedings of the PLS'03 International Symposium* (pp. 33-43). Lisbon: Decisia.

Chin, W. W., & Dibbern, J. 2010. A permutation procedure for multi–group PLS analysis. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* (in print. Heidelberg: Springer.

Chin, W. W., & Gopal, A. 1995. Adoption intention in GSS: Relative importance of beliefs. *Data Base for Advances in Information Systems*, **26**(2&3), 42–64.

Chin, W. W., & Marcolin, B. L. 1995. The holistic approach to construct validation in IS research: Examples of the interplay between theory and measurement. *In:* Compeau, D. (Ed.), *Proceedings of the 23rd Conference of the Administrative Sciences Association of Canada* (Vol. 14, no 4). Ontario, Canada.

Chin, W. W., & Newsted, P. R. 1999. Structural equation modeling analysis with small samples using partial least squares. *In:* Hoyle, R. H. (Ed.), *Statistical* *strategies for small sample research* (pp. 307–341). California: Sage Publications.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. 2003. A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic–mail emotion/adoption study. *Information Systems Research*, **14**(2), 189–217.

Chin, W.W. 1998b. Issues and opinion on structural equation modeling. *MIS Quarterly*, **22**(1), vii–xvi.

Chu, M. T., & Watterson, J. L. 1993. On a multivariate eigenvalue problem. I: Algebraic theory and a power method. *SIAM Journal of Scientific Computing*, **41**(5), 1089–1106.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., & Cohen, P. 1983. *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Noemi, C. 1990. Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement*, **14**(2), 183–196.

Connolly, R. A., & Hirschey, M. 1988. Market value and patents: A Bayesian approach. *Economics Letters*, **27**(1), 83.

Cool, K., Dierickx, I., & Jemison, D. 1989. Business strategy, market structure and risk–return relationships: A structural approach. *Strategic Management Journal*, **10**(6), 507–522.

Cudeck, R., & Browne, M. W. 1983. Cross–validation of covariance–structures. *Multivariate Behavioral Research*, **18**(2), 147–168.

Cudeck, R., & MacCallum, R. C. 2007. *Factor analysis at 100: Historical developments and future directions.* Mahwah: Lawrence Erlbaum.

Cudeck, R., du Toit, S., & Sörbom, D. 2001. *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog.* Scientific Software International.

Curtis, R., & Jackson, E. 1962. Multiple indicators in survey research. *The American Journal of Sociology*, **68**(2), 195–204.

Dechezleprêtre, A., Glachant, M., Hascic, I., Johnstone, N., & Ménière, Y. 2009. *Invention and transfer of climate change mitigation technologies on a global scale: A study drawing on patent data.* Working Paper 82. Fondazione Eni Enrico Mattei, Milano.

Diamantopoulos, A. 2006. The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modeling in Management*, **1**(1), 7–17.

Diamantopoulos, A., & Winklhofer, H. 2001. Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, **38**, 269–277.

Diamantopoulos, A., P., Riefler, & K.P., Roth. 2008. Advancing formative measurement models. *Journal of Business Research*, **61**(12), 1203–1218.

Dijkstra, T. 1983. Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics*, **22**, 67–90.

Dijkstra, T. 1992. On statistical inference with parameter estimates on the boundary of the parameter space. *British Journal of Mathematical and Statistical Psychology*, **45**, 289–309.

Dijkstra, T. 2010. Latent variables and indices: Herman Wold's basic design and partial least squares. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* in print. Heidelberg: Springer.

Duncan, O. D. 1957. The measurement of population–distribution. *Population Studies: A Journal of Demography*, **11**(1), 27–45.

Efron, B., & Tibshirani, R. J. 1993. *An introduction to the bootstrap.* New York: Chapman & Hall.

Escofier, B, & Pagés, J. 1994. Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis*, **18**(1), 121–140.

Esposito Vinzi, V. 2007. The PLS approach to data exploration and modeling: An everlasting matter of dispute or a playground for integrating different cultures? *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'07 International Symposium* (pp. 8–11). As, Norway: Matforsk.

Esposito Vinzi, V., & Lauro, C. 2005. Partial least squares. *Computational Statistics & Data Analysis*, **48**(1), 1–4.

Esposito Vinzi, V., Lauro, C. N., & Amato, S. 2005. PLS typological regression: Algorithmic, classification and validation issues. *In:* Vichi, M., Monari, P., Mignani, S., & Montanari, A. (Eds.), *New developments in classification and data analysis* (pp. 133-140). New York: Springer.

Esposito Vinzi, V., Ringle, C. R., Squillacciotti, S., & Trinchera, L. 2007. *Capturing and treating heterogeneity by response based segmentation in PLS path Modeling: A comparison of alternative methods by computational experiments*. Working Paper DR-07019. ESSEC Research Center, available at www45.essec.edu/pwr/servlet/showRef?bibID?=?7153.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S., & Tenenhaus, M. 2008. REBUS-PLS: A response-based procedure for detecting unit segments in PLS path modeling. *Applied Stochastic Models in Business and Industry*, **24**(5), 439–458.

Falguerolles, A. 2008. L'analyse des données: before and around. *Electronic Journal for History of Probability and Statistics*, **4**(2), December.

Fornell, C. 1985. *A second generation of multivariate analysis: Classification of methods and implications for marketing research.* Working Paper 414. Division of Research, Graduate School of Business Administration, University of Michigan, available at http://hdl.handle.net/2027.42/35621.

Fornell, C. 1992. A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, **56**(1), 6–21.

Fornell, C., & Bookstein, F. L. 1982. Two structural equation models: LISREL and PLS applied to consumer exit–voice theory. *Journal of Marketing Research*, **19**(4), 440.

Fornell, C., & Cha, J. 1994. Partial least squares. *Advanced Methods of Marketing Research*, 52–78.

Fornell, C., & Larcker, D. F. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, **18**(1), 39–50.

Frank, I. E. 1990. A non–linear PLS model. *Chemometrics and Intelligent Laboratory Systems*, **8**, 109–119.

Gallini, N. T. 2002. Patent policy and costly imitation. *Rand Journal of Economics*, **16**(2), 131–154.

Gentle, J. 2003. *Random number generation and Monte Carlo methods*. New York: Springer.

Gerlach, R. W., Kowalski, B. R., & Wold, H. O. 1979. Partial least–squares path modeling with latent–variables. *Analytica Chimica Acta–Computer Techniques and Optimization*, **3**(4), 417–421.

Gilbert, R. J., & Shapiro, C. 1990. Optimal patent length and breadth. *Rand Journal of Economics*, **21**(1), 106–112.

Goldberger, A. S. 1973. Structural equation models: An overview. *In:* Goldberger, A. S., & Duncan, O. D. (Eds.), *Structural equation models in the social sciences* (pp. 1–18). New York: Seminar Press.

González-Albo, B., & Zulueta, M. 2007. Estudio comparativo de bases de datos de patentes en internet. *Anales de Documentación*, **10**, 145–162.

Goodhue, D., Lewis, W., & Thompson, R. 2006. PLS, small sample size, and statistical power in MIS research. *In: HICSS'06: Proceedings of the 39th Annual Hawaii International Conference on System Science* (pp. 202.2). Washington, DC, USA: IEEE Computer Society.

Goodhue, D., Lewis, W., & Thompson, R. 2007. Statistical power in analyzing interaction effects: Questioning the advantage of PLS with product indicators. *Information Systems Research*, **18**(2), 211–227.

Gower, J. 2008. The biological stimulus to multidimensional data analysis. *Electronic Journal for History of Probability and Statistics*, **4**(2), December.

Green, J., & Scotchmer, S. 1997. On the division of profit in sequential innovation. *Rand Journal of Economics*, **26**(1), 20–33.

Grefermann, K., Oppenländer, K. H., Peffgen, E., Röthlingshfer, K. Ch., & Scholz, L. 1974. Patentwesen und technischer fortschritt. *In: IFO-Institut für Wirtschaftsforschung* (p. 45 part I). Göttingen: O. Schwartz.

Griliches, Z. 1981. Market value, R–and–D, and patents. *Economics Letters*, **7**(2), 183–187.

Griliches, Z. 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, **28**(4), 1661–1707.

Griliches, Z., Hall, B., & Pakes, A. 1988. *R&D, patents and market value revisited: Is there a second (technological opportunity) factor?* Working Paper 2624. National Bureau of Economic Research (NBER), Cambridge, MA.

Gudergan, S.P. 2005. PLS and confirmatory tetrad testing for formative measurement scales in marketing. *In:* Aluja, T., Casanovas, J., Esposito Vinzi, V., Morineau, A., & Tenenhaus, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'05 International Symposium* (pp. 103–108). Paris: Decisia.

Gudergan, S.P., Ringle, C.M., Wende, S., & Will, A. 2008. Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, **61**(12), 1238–1249.

Guellec, D., & van Pottelsberghe, B. 2000. Applications, grants and the value of patent. *Economics Letters*, **69**(1), 109–114.

Haavelmo, T. 1947. Methods of measuring the marginal propensity to consume. *Journal of the American Statistical Association*, **42**(237), 105–122.

Hadaya, P., & Cassivi, L. 2007. The role of joint collaboration planning actions in a demand–driven supply chain. *Industrial Management & Data Analysis*, **107**(7), 954–978.

Hahn, C. H., Johnson, M. D., Herrman, A., & Huber, F. 2002. Capturing customer heterogeneity using a finite mixture PLS approach. *Schmalenbach Business Review*, **54**(3), 243–269.

Hall, B. H., & Ziedonis, R. H. 2001. The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979–1995. *Rand Journal of Economics*, **32**(1), 101.

Hall, B. H., Griliches, Z., & Hausman, J.A. 1986. Patents and R & D: Is there a lag? *International Economic Review*, **27**(2), 265–283.

Hall, B. H., Jaffe, A., & Trajtenberg, M. 2001. *The NBER patent citations data file: Lessons, insights and methodological tools.* Working Paper 8498. National Bureau of Economic Research (NBER), Cambridge, MA.

Hall, B. H., Jaffe, A., & Trajtenberg, M. 2005. Market value and patent citations. *Rand Journal of Economics*, **36**(1), 16–38.

Hanafi, M. 2007. PLS path modeling: Computation of latent variables with the estimation mode B. *Computational Statistics*, **22**, 275–292.

Hanafi, M., & Qannari, E. 2005. An alternative algorithm to the PLS B problem. *Computational Statistics & Data Analysis*, **48**, 63–67.

Harhoff, D., & Reitzig, M. 2004. Determinants of opposition against EPO patent grants: The case of biotechnology and pharmaceuticals. *International Journal of Industrial Organization*, **22**(4), 443–480.

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. 1999. Citation frequency and the value of patented inventions. *Review of Economics and Statistics*, **81**(3), 511–515.

Harhoff, D., Scherer, F. M., & Vopel, K. 2003. Citations, family size, opposition and the value of patent rights. *Research Policy*, **32**(8), 1343–1363.

Hausman, J.A., Hall, B. H., & Griliches, Z. 1984. Econometric models for count data with an application to the patents: R&D relationship. *Econometrica*, **52**(4), 909–938.

Heiser, W. J. 2008. Psychometric roots of multidimensional data analysis in the Netherlands: From Gerard Heymans to John van der Geer. *Electronic Journal for History of Probability and Statistics*, **4**(2), December.

Henseler, J. 2009. On the convergence of the partial least squares path modeling algorithm. *Computational Statistics*, August. doi:10.1007/s00180–009–0164–x.

Henseler, J., & Chin, W. W. 2010. A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, **17**(1), 82–109.

Henseler, J., & Fassott, G. 2005. Testing nonlinear effects in PLS path models: An illustration of available procedures. *In:* Aluja, T., Casanovas, J., Esposito Vinzi, V., Morineaur, A., & Tenenhaus, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'05 International Symposium* (pp. 371–377). Paris: Decisia.

Henseler, J., & Fassott, G. 2007. A new and simple approach to multi–group analysis in PLS path modeling. *In: ESSEC-HEC Research Workshop on PLS Path Modeling in the Presence of Group Structure*. Cergy-Pontoise, 15 May.

Henseler, J., & Fassott, G. 2010. Testing moderating effects in PLS path models: An illustration of available procedures. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* (in print). Heidelberg: Springer.

Henseler, J., Wilson, B., & Dijkstra, T. 2007. Testing nonlinear effects in PLS path models: A simulation–based comparison of available approaches. *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'07 International Symposium* (pp. 132–135). As, Norway: Matforsk.

Henseler, J., Ringle, C. M., & Sinkovics, R. 2009. The use of partial least square path modeling in international marketing. *Advances in International Marketing*, **20**, 277–319.

Horst, P. 1961. Relations among m sets of measures. *Psychometrika*, **26**(2), 129–149.

Hsu, H. H., Chen, W. H., & Hsieh, M. J. 2006. Robustness testing of PLS, LISREL, EQS and ANN–based SEM for measuring customer satisfaction. *Total Quality Management*, **17**(3), 355–371.

Hui, S., & Wold, H. 1982. Consistency and consistency at large of partial least squares estimates. *In:* Jöreskog, G., & Wold, H. (Eds.), *Systems under indirect observation, Part II* (pp. 119–130). Amsterdam: North–Holland.

Hulland, J. 1999. Use of partial least square in strategic management research: A review of four recent studies. *Strategic Management Journal*, **20**, 195–204.

Igami, M., & Okazaki, T. 2007. *Capturing nanotechnology's current state of development via analysis of patents*. Tech. rept. Organisation for Economic Co-operation and Development, OECD, Paris.

Jaffe, A. B., & Lerner, J. 2001. Reinventing public R&D: patent policy and the commercialization of national laboratory technologies. *Rand Journal of Economics*, **32**(1), 167–198.

Jaffe, A. B., & Trajtenberg, M. 1999. International knowledge flows: Evidence from patent citations. *Economics of Innovation & New Technology*, **8**(1/2), 105.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Proceedings of the National Academy of Sciences*, **108**(3), 577–598.

Jakobowicz, E. 2007. *Contributions aux modèles d'équations structurelles à variables latentes*. PhD Thesis, Conservatoire National des Arts et Métiers, Paris.

Jakobowicz, E., & Derquenne, C. 2007. A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy. *Computational Statistics & Data Analysis*, **51**(8), 3666–3678.

Jakobowicz, E., & Saporta, G. 2007. A nonlinear PLS path modeling based on monotic B-spline transformations. *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'07 International Symposium* (pp. 108–111). As, Norway: Matforsk.

Jarvis, C. B., Mackenzie, S. B., Podsakoff, P. M., Mick, D. G., & Bearden, W. O. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, **30**(2), 199.

Johnstone, N., Hascic, I., Clavel, L., & Marical, F. 2007. *Renewable energy policies and technological innovation: Empirical evidence based on patent counts.* Tech. rept. Organisation for Economic Co-operation and Development, OECD, Paris.

Jöreskog, K. G. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183–202.

Jöreskog, K. G. 1970. A general method for analysis of covariance structures. *Biometrika*, **57**(2), 239–251.

Jöreskog, K. G. 1973. A general method for estimating a linear structural equation system *(pp. 85–112). In:* Goldberger, A. S., & Duncan, O. D. (Eds.), *Structural equation models in the social sciences.* New York: Seminar Press.

Jöreskog, K. G., & Goldberger, A. S. 1972. Factor–analysis by generalized least–squares. *Psychometrika*, **37**(3), 243–260.

Jöreskog, K. G., & Goldberger, A. S. 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, **70**(351), 631–639.

Jöreskog, K. G., & Sörbom, D. 1982. Recent developments in structural equation modeling. *Journal of Marketing Research*, **19**(4), 404–416.

Jöreskog, K. G., & Wold, H. 1982a. The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects. *In:* Jöreskog, G., & Wold, H. (Eds.), *Systems under indirect observation, Part I* (pp. 263–270). Amsterdam: North–Holland.

Jöreskog, K. G., & Wold, H. 1982b. *Systems under indirect observation, Part I and II.* Amsterdam: North–Holland.

Keesling, J. W. 1972. *Maximum likelihood approaches to causal analysis.* Ph.D. thesis, Department of Education: University of Chicago, Chicago.

Kenny, D., & Judd, C. 1984. Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, **96**(1), 201–210.

Klemperer, P. 1990. How broad should the scope of patent protection be? *Rand Journal of Economics*, **21**(1), 113–130.

Kmenta, J. 1986. *Elements of econometrics.* New York: MacMillan Publishing Company.

Krämer, N. 2005. Nonlinear partial least squares path models, 3rd World Conference of the IASC. *In: Jakobowicz, E. (2007) Contributions aux modeèles* *d'équations structurelles à variables latentes.* PhD Thesis Conservatoire National des Arts et Métiers, Paris.

Krämer, N. 2006. *Two negative results on partial least squares path models.* Technical Report 09/2006. Berlin Institute of Technology, Berlin.

Lance, C. E. 1988. Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. *Applied Psychological Measurement*, **12**(2), 163–175.

Lanjouw, J. O. 1998. Patent protection in the shadow of infringement: Simulation estimations of patent value. *Review of Economic Studies*, **65**(4), 671–710.

Lanjouw, J. O., & Schankerman, M. 2001. Characteristics of patent litigation: A window on competition. *Rand Journal of Economics*, **32**(1), 129–151.

Lanjouw, J. O., & Schankerman, M. 2004. Patent quality and research productivity: Measuring innovation with multiple indicators. *Economic Journal*, **114**(495), 441–465.

Lanjouw, J. O., Pakes, A., & Putnam, J. 1998. How to count patents and value intellectual property: The uses of patent renewal and application data. *Journal of Industrial Economics*, **46**(4), 405–432.

Lebart, L. 1985. *Tratamiento estadístico de datos: Métodos y programas.* Barcelona: Marcombo Boixarem.

Lebart, L. 2008. Exploratory multivariate data analysis from its origins to 1980: Nine contributions. *Electronic Journal for History of Probability and Statistics*, **4**(2), December.

Lerner, J. 1994. The importance of patent scope: An empirical analysis. *Rand Journal of Economics*, **25**(2), 319.

Lindgren, F., Geladi, P., & Wold, S. 1993. The kernel algorithm for PLS. *Journal of Chemometrics*, **7**, 45–49.

Little, T., Bovair, J., & Widaman, K. 2006. On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling*, **13**(4), 497–519.

Lohmöller, J. B. 1989. *Latent variables path modeling with partial least squares.* Heildelberg: Physical–Verlag.

Lyttkens, E. 1973. The fix-point method for estimating interdependent systems with underlying model specification. *Journal of the Royal Statistical Society*, **136**(3), 353–394.

MacCallun, R., & Browne, M. 1993. The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, **114**(3), 533–541.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. 2005. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, **90**(4), 710–730.

Malthouse, E., Tamhane, A., & Mah, R. 1997. Nonlinear partial least square. *Computers & Chemical Engineering*, **21**(8), 875–890.

Marcoulides, G. 2003. PLS model specification searches using optimization algorithms. *In:* Vilares, M. J., Tenenhaus, M., Coelho, P. S., Esposito Vinzi, V., & Morineau, A. (Eds.), *PLS and Related Methods: Proceedings of the PLS'03 International Symposium* (pp. 75–86). Lisbon: Decisia.

Marcoulides, G., & Saunders, C. 2006. PLS: A silver bullet? *MIS Quarterly*, **30**(2), iii–ix.

Marcoulides, G., Chin, W.W., & Saunders, C. 2009. A critical look at partial least squares modeling. *MIS Quarterly*, **33**(1), 171–175.

Markus, G. B., & Converse, P. E. 1979. Dynamic simultaneous equation model of electoral choice. *American Political Science Review*, **73**(4), 1055–1070.

Martens, M., Martens, H., & Wold, S. 1983. Preference of cauliflower related to sensory descriptive variables by partial least–squares (PLS) regression. *Journal of the Science of Food and Agriculture*, **34**(7), 715–724.

Martínez-Ruiz, A., & Aluja-Banet, T. 2008. Second-order model of patent and market value. *In:* Brito, P. (Ed.), *Proceedings of the COMPSTAT'2008 International Conference on Computational Statistics* (pp. 1445–1452). Heidelberg: Physica Verlag.

Martínez-Ruiz, A., & Aluja-Banet, T. 2009. Toward the definition of a structural equation model of patent value: PLS path modeling with formative constructs. *REVSTAT Statistical Journal*, **7**(3), 265–290.

Mathes, H. 1993. Global optimization criteria of the PLS–algorithm in recursive path models with latent variables. *In:* Haagen, K., Bartholomew, D.J., & Deistler, M. (Eds.), *Statistical modeling and latent variables* (pp. 229–248). Elsevier Science Publishers.

McDonald, R.P. 1962. A general approach to non-linear factor analysis. *Psychometrika*, **27**, 397–415.

McDonald, R.P. 1996. Path analysis with composite variables. *Multivariate Behavioral Research*, **31**(2), 239–270.

Megna, P., & Klock, M. 1993. The impact of intangible capital on Tobins–Q in the semiconductor industry. *American Economic Review*, **83**(2), 265–269.

Morgan, M. S. 1990. *The history of econometrics ideas: Historical perspectives on modern economics.* New York: Cambridge University Press, Cambridge.

Narin, F., Hamilton, K. S., & Olivastro, D. 1997. The increasing linkage between U.S. technology and public science. *Research Policy*, **26**(3), 317–330.

Noonan, R., & Wold, H. 1982. PLS path modeling with indirectly observed variables: A comparison of alternative estimates for the latent variable. *In:* Jöreskog, K. G. & Wold, H. (Ed.), *Systems under indirect observation, part II* (pp. 75–94). Amsterdam: North–Holland.

Nordhaus, W. 1967. *The optimal life of a patent.* Discussion Paper 241. Yale University, New Haven.

Ogasawara, H. 2003. Oblique factors and components with independent clusters. *Psychometrika*, **68**(2), 299–321.

Paci, R., Sassu, A., & Usai, S. 2004. International patenting and national technological specialization. *Technovation*, **17**(1), 25–38.

Pagés, J., & Tenenhaus, M. 2001. Multiple factor analysis combined with PLS path modeling: Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. *Chemometrics and Intelligent Laboratory Systems*, **58**(2), 261–273.

Pakes, A. 1985. On patents, R–and–D, and the stock–market rate of return. *Journal of Political Economy*, **93**(2), 390–409.

Pakes, A. 1986. Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica*, **54**(4), 755–784.

Pakes, A., & Schankerman, M. 1984. The rate of obsolescence of patents, research gestation lags and the private rate of return to research resources. *In:* Griliches, Z. (Ed.), *R&D, Patents and Productivity* (pp. 73–88). Chicago: University Chicago Press.

Pakes, A., Simpson, M., Judd, K., & Mansfield, E. 1989. Patent renewal data. *Brookings Papers on Economic Activity. Microeconomics*, **1989**, 331–410.

Palumbo, F., Romano, R., & Esposito Vinzi, V. 2008. Fuzzy PLS path modeling: A new tool for handling sensory data. *In:* Preisach, C., Burkhardt, H., Schmidt-Thieme, L., & Decker, R. (Eds.), *Data Analysis, Machine Learning, and Applications: Proceedings of the 31st Annual Conference of the German Classification Society* (pp. 689–696). Berlin: Springer.

Paxton, P., Curran, P., Bollen, K., Kirby, J., & Chen, F. 2001. Monte Carlo experiments: Design and implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, **8**(2), 287–312.

Peña, D. 1998. *Estadística, modelos y métodos II: Modelos lineales y series temporales.* Madrid: Alianza Universidad Textos.

Pearl, J. 2000. *Causality: Models, reasoning and inference.* New York: Cambridge University Press.

Pérez-Enciso, M., & Tenenhaus, M. 2003. Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, **112**, 581–592.

Petter, S., Straub, D., & Rai, A. 2007. Specifying formative constructs in information systems research. *MIS Quarterly*, **31**(4), 623–656.

Ping, R. 1995. A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research*, **32**, 336–347.

Qin, S. J., & McAvoy, T. J. 1992. Non–linear PLS modeling using neural networks. *Computers and Chemical Engineering*, **16**, 379–391.

Qureshi, I., & Compeau, D. 2009. Assessing between–group differences in information systems research: A comparison of covariance- and component-based SEM. *MIS Quarterly*, **33**(1), 197–214.

R Development Core Team. 2007. *R: A Language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Reitzig, M. 2003. What determines patent value? Insights from the semiconductor industry. *Research Policy*, **32**(1), 13–26.

Reitzig, M. 2004a. Improving patent valuations for management purposes: Validating new indicators by analyzing application rationales. *Research Policy*, **33**(6–7), 939–957.

Reitzig, M. 2004b. Strategic management of intellectual property. *Mit Sloan Management Review*, **45**(3), 35–40.

Ringle, C. M., Wende, S., & Will, A. 2010. The finite mixture partial squares approach: Methodology and application. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* in print. Heidelberg: Springer.

Ringle, C.M., & Schlittgen, R. 2007. A genetic segmentation approach for uncovering and separating groups of data in PLS path modeling. *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'07 International Symposium* (pp. 44–47). As, Norway: Matforsk.

Ringle, C.M., Wende, S., & Will, A. 2005a. Customer segmentation with FIMIX-PLS. *In:* Aluja, T., Casanovas, J., Esposito Vinzi, V., Morrineau, A., & Tenenhaus, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'05 International Symposium* (pp. 507–514). Paris: Decisia.

Ringle, C.M., Wende, S., & Will, A. 2005b. *SmartPLS 2.0 (beta)*. University of Hamburg, Hamburg, Germany.

Ringle, C.M., Wilson, B., & Götz, O. 2007. A Monte Carlo robustness study on formative measurement model specification in CBSEM and PLS. *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'07 International Symposium* (pp. 108–111). As, Norway: Matforsk.

Ringle, C.M., Götz, O., Wetzels, M., & Wilson, B. 2009. *On the use of formative measurement specifications in structural equation modeling: A Monte Carlo simulation study to compare covariance–based and partial least squares model estimation methodologies*. METEOR Research Memoranda RM/09/014. Maastricht University, Maastricht.

Rivette, K., & Kline, D. 1999. *Rembrandts in the attic: Unlocking the hidden value of patents*. Boston: Harvard Business School Press.

Rizopoulos, D., & Moustaki, I. 2008. Generalized latent variable models with non-linear effects. *British Journal of Mathematical and Statistical Psychology*, **61**, 415–438.

Rosipal, R. 2003. Kernel partial least squares for nonlinear regression and discrimination. *Neural Network World*, **13**(3), 291–300.

Rosipal, R., & Trejo, L. 2001. Kernel partial least squares in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, **2**, 97–123.

Rossiter, J. R. 2002. The C–O–A–R–S–E procedure for scale development in marketing. *International Journal of Research in Marketing*, **19**, 305–335.

Sánchez, G., & Aluja, T. 2006. PATHMOX: A PLS-PM segmentation algorithm. *In:* Esposito Vinzi, V., Lauro, C., Braverman, A., Kiers, H. A. L., & Schimek, M. G. (Eds.), *Electronic Proceedings of the Workshop on Knowledge Extraction and Modeling (KNEMO).* Napoli: Tilapia.

Sánchez, G., & Aluja, T. 2007. A simulation study of PATHMOX (PLS path modeling segmentation tree) sensitivity. *In:* Martens, H., Naes, T., & Martens, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'07 International Symposium* (pp. 33–36). As, Norway: Matforsk.

Sarstedt, M. 2008. A review of recent approaches for capturing heterogeneity in partial least square path modeling. *Journal of Modeling in Management*, **3**(2), 140–161.

Scepi, G., & Esposito Vinzi, V. 2003. PLS path modeling and multiple tables analysis: An integrated approach for time dependent data. *In:* Vilares, M., Tenenhaus, M., Coelho, P., Esposito Vinzi, V., & Morineau, A. (Eds.), *PLS and Related Methods: Proceedings of the PLS'03 International Symposium* (pp. 453–463). Paris: Decisia.

Schankerman, M., & Pakes, A. 1986. Estimates of the value of patent rights in European–countries during the post–1950 period. *Economic Journal*, **96**(384), 1052–1076.

Scheiness, R., Spirtes, P., Glymour, C., & Meek, C. 1994. *Tetrad II: Tools for discovery.* Erlbaum.

Schmoch, U., Grupp, H., Mannsbart, W., & Schwitalla, B. 1988. *Technikprognosen mit patentindikatoren.* Köln: Verlag TÜV-Rheinland.

Schneeweiss, H. 1991. Models with latent variables: LISREL versus PLS. *Statistica Neerlandica*, **45**(2), 145–157.

Schneeweiss, H. 1993. Consistency at large in models with latent variables. *In:* Haagen, K., Bartholomew, D.J., & Deistler, M. (Eds.), *Statistical modeling and latent variables* (pp. 229–248). Elsevier.

Serrano-Cinca, C., Fuertes-Callén, Y., & Gutiérrez-Nieto, B. 2007. Online reporting by banks: A structural modeling approach. *Online Information Review*, **31**(3), 310–332.

Sherry, E. F., & Teece, D. J. 2004. Royalties, evolving patent rights, and the value of innovation. *Research Policy*, **33**(2), 179–191.

Sohn, S. Y., Kim, H. S., & Moon, T. H. 2007. Predicting the financial performance index of technology fund for SME using structural equation model. *Expert Systems with Applications*, **32**(3), 890–898.

Sosik, J. J., Kahai, S. S., & Piovoso, M. J. 2009. Silver bullet or voodoo statistics? A primer for using the partial least squares data analytic technique in group and organization research. *Group & Organization Management*, **34**(1), 5–36.

Spearman, C. 1904. General intelligence, objectively determined and measured. *American Journal of Psychology*, **15**, 201–293.

Spearman, C. 1927. *The abilities of man: Their nature and measurement.* London: MacMillan and Co.

Squillacciotti, S. 2005. Prediction-oriented classification in PLS path modeling. *In:* Aluja, T., Casanovas, J., Esposito Vinzi, V., Morrineau, A., & Tenenhaus, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'05 International Symposium* (pp. 499–506). Paris: Decisia.

Squillacciotti, S. 2007. PLS typological path modeling. *In: ESSEC-HEC Research Workshop on PLS Path Modeling in the Presence of Group Structure.* Cergy-Pontoise, 15 May.

Squillacciotti, S. 2010. Prediction-oriented classification in PLS path modeling. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* in print. Heidelberg: Springer.

Steiger, J. H. 1979. Factor indeterminacy in the 1939's and the 1970's: Some interesting parallels. *Psychometrika*, **44**(1), 157–167.

Steiger, J. H. 1996. Coming full circle in the history of factor indeterminacy. *Multivariate Behavioral Research*, **31**(4), 617–630.

Temme, D., & Kreis, H. 2005. The PLS approach to structural equation modeling with latent variables: A software review. *In:* Aluja, T., Casanovas, J., Esposito Vinzi, V., Morineau, A., & Tenenhaus, M. (Eds.), *PLS and Related Methods: Proceedings of the PLS'05 International Symposium* (pp. 87–94). Paris: Decisia.

Tenenhaus, A., Giron, A., Viennet, E., Béra, M., Saporta, G., & Fertil, B. 2007. Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification. *Computational Statistics & Data Analysis*, **51**, 4083–4100.

Tenenhaus, M. 1998. *La régression PLS: Théorie et pratique.* Paris: Technip.

Tenenhaus, M. 2009. A criterion based PLS approach to SEM. *In:* Esposito Vinzi, V., Tenenhaus, M., & R., Guan. (Eds.), *Programme and Abstract of the 6th International Conference in Partial Least Squares Methods*. Beijing: Publishing House of Electronics Industry.

Tenenhaus, M., & Esposito Vinzi, V. 2005. PLS regression, PLS path modeling and generalized procrustean analysis: A combined approach for multiblock analysis. *Journal of Chemometrics*, **19**(3), 145–153.

Tenenhaus, M., & Hanafi, M. 2010. A Bridge between PLS Path Modelling and Multi-Block Data Analysis. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* (in print). Heidelberg: Springer.

Tenenhaus, M., Amato, A., & Esposito Vinzi, V. 2004. A global goodness-of-fit index for PLS structural equation modelling. *In: Proceedings of the XLII SIS Scientific Meeting* (pp. 739-742). Padova: CLEUP.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. M., & Lauro, C. 2005. PLS path modeling. *Computational Statistics & Data Analysis*, **48**, 159–205.

Tong, X. S., & Frame, J. D. 1994. Measuring national technological performance with patent claims data. *Research Policy*, **23**(2), 133–141.

Trajtenberg, M. 1990a. *Economic analysis of product innovation: The case of CT scanners*. Cambridge, Mass.: Harvard University Press.

Trajtenberg, M. 1990b. A penny for your quotes: Patent citations and the value of innovations. *Rand Journal of Economics*, **21**(1), 172–187.

Trajtenberg, M., Henderson, R., & Jaffe, A. 1997. University versus corporate patents: A window on the basicness of invention. *Economics of Innovation & New Technology*, **5**(1), 19–50.

Trinchera, L. 2007. REBUS-PLS; Response-based Units Segmentation in PLS Path Modeling. *In: ESSEC-HEC Research Workshop on PLS Path Modeling in the Presence of Group Structure*. Cergy-Pontoise, 15 May.

Trinchera, L., Squillacciotti, S., & Esposito Vinzi, V. 2006. PLS typological path modeling: A model-based approach to classification. *In:* Esposito Vinzi, V., Lauro, C., Braverman, A., Kiers, H. A. L., & Schimek, M. G. (Eds.), *Electronic Proceedings of the Workshop on Knowledge Extraction and Modeling (KNEMO)*. Napoli: Tilapia.

Upton, G. J. 2006. *A dictionary of statistics*. Oxford: Oxford University Press.

Valencia, J., & Díaz-Llanos, J. 2003. *Regresión PLS en las ciencias experimentales.* Madrid: Complutense.

Vilares, M., Almeida, M., & Coelho, P. 2010. Comparison of likelihood and PLS estimators for structural equation modeling: A simulation with customer satisfaction data. *In:* Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), *Handbook of partial least squares: Concepts, methods and applications* (in print). Heidelberg: Springer.

Westlund, A. H., Cassel, C. M., Eklöf, J., & Hackl, P. 2001. Structural analysis and measurement of customer perceptions, assuming measurement and specifications errors. *Total Quality Management*, **12**(7&8), 873–881.

Westlund, A. H., Källström, M., & Parmler, J. 2008. SEM-based customer satisfaction measurement: On multicolinearity and robust PLS estimation. *Total Quality Management*, **19**(7&8), 855–869.

Wetzels, M., Odekerken-Schroder, G., & van Oppen, C. 2009. Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, **33**(1), 177–195.

Wheaton, B., Muthén, B., Alwin, D., & Summers, G. 1977. Assessing reliability and stability in panel models. *Sociological Methodology*, **8**, 84–136.

Wiley, D. E. 1973. The identification problem for structural equation models with unmeasured variables. *In:* Goldberger, A. S., & Duncan, O. D. (Eds.), *Structural equation models in the social sciences* (pp. 69–83). New York: Academic Press.

Wilson, D. J. H., Irwin, G. W., & Lightbody, G. 1997. Nonlinear PLS modeling using radial basis functions. *In: American Control Conference.*

WIPO. *Understanding industrial property.* Tech. rept. World Intellectual Property Organization, Publication Number 895(E), downloaded on February 12, 2007.

Wold, H. 1966. A fix-point theorem with econometric background. *Arkiv för Matematik*, **6**(3), 209–220.

Wold, H. 1973a. Non linear iterative partial least squares (NIPALS) modeling: Some current developments. *In:* Krishnaiah, P. R. (Ed.), *Multivariate analysis III* (pp. 383–407). New York: Academic Press.

Wold, H. 1973b. Operative aspects of econometric and sociological models current developments of FP (fix–point) estimation and NIPALS (nonlinear iterative partial least squares) modeling. *Economie Appliquee*, **26**(2–4), 385–421.