

Reexamining the Role of Heterogeneous Agents in Stock Markets, Labor Markets, and Tax Policy

Anna Katharina Greulich

Economics and Business Department
Universitat Pompeu Fabra
Barcelona

Advisor: Albert Marcet

June 29, 2007

Dipòsit legal: B.5231-2008
ISBN: 978-84-691-4464-0

Contents

1	Introduction	1
2	Asset Pricing with Heterogeneous Epstein-Zin Agents	4
2.1	Introduction	5
2.2	The model	9
2.3	Three auxiliary models	12
2.3.1	A two-period model with heterogeneous agents	13
2.3.2	The infinite horizon model with a representative agent: comparative statics	16
2.3.3	Uncertainty in a single period	19
2.4	Numerical analysis	21
2.4.1	Calibration	22
2.4.2	Results	25
2.5	Conclusion	35
2.6	Appendix	37
2.6.1	Choice of state space and computational strategy	37
2.6.2	Complete versus incomplete markets	37
3	Dynamic Effects of Unemployment Insurance Reform	40
3.1	Introduction	41
3.2	The Model	45
3.2.1	The Individual's Problem	46

3.2.2	Firms	47
3.2.3	Determination of Wages and Matching Probabilities	48
3.2.4	Competitive Equilibrium	50
3.3	Calibration	51
3.4	Steady-State Analysis	54
3.5	Dynamic Analysis	57
3.5.1	The Benchmark - Transition Dynamics and Welfare Effects	57
3.5.2	The Effects of Announcement and the Nature of Turnover Costs	60
3.5.3	The Optimal Path of Reform	63
3.6	Conclusion	65
3.7	Appendix	67
3.7.1	Job Acceptance	67
3.7.2	The Algorithms	67
4	Optimal Capital and Labor Taxes with Heterogeneous Agents: Making Everybody Happy (with Albert Marcet)	70
4.1	Introduction	71
4.2	Models	75
4.2.1	The Baseline Model	75
4.2.2	Modifications of the model	82
4.3	Characterization of equilibria	84
4.3.1	Steady state and zero capital taxes	84
4.3.2	Recursive Formulation	89
4.3.3	Frontier of Equilibrium Set	90
4.4	Results	91
4.4.1	Calibration	91
4.4.2	The set of Pareto Optimal-Pareto Improving Plans	93
4.4.3	The Time Path of the Economy under POPI Plans	95
4.4.4	Isolating the Effects: Modified Models 1 and 2	97

4.5	Conclusion	101
4.6	Appendix	102
4.6.1	The maximization problem and first order conditions	102
4.6.2	Computational strategy: Approximation of the time path	105
Tables and Figures		114

List of Tables

1	Benchmark Parameter Values	116
2	Unconditional statistics of simulated and historical data	117
3	Long-horizon return regressions	117
4	Volatilities and correlations of individual and aggregate consumption growth ($w_{L,0} = 0.1$)	117
5	Means, standard deviations, and long horizon return regressions for the calibration to dividend data	118
6	Parameters	120
7	Welfare effects of lowering the replacement rate (old level: 60%) with- out announcement	120
8	Welfare effects of reducing the replacement rate from 60% to 40% de- pending on the length of the announcement period	120
9	Parameter Values of the Baseline Economy.	122

List of Figures

1	The timing of events in the two-period model	124
2	Equity premium (in % p.a.) and price-dividend ratio as functions of the wealth distribution	125
3	The share of the stock in each agent type's portfolio as a function of the wealth distribution	126
4	Stock holdings of the L-type as a function of his wealth share. (The dotted line represents $s_L = w_L$, i.e. the counterfactual of no asset trade.)	126
5	Sharpe ratio, equity premium (in % p.a.), and volatility of excess returns as functions of the wealth distribution Solid lines: heterogeneous agent model Dashed lines: emulation in the representative agent model	127
6	The Role of the intertemporal elasticity of substitution	128
7	Impulse responses: Difference between simulation and without a dividend growth shock of one standard deviation.	129
8	Steady-state welfare by asset level and employment status across replacement rates	131
9	The transition paths of vacancy-unemployment ratio, median asset holdings, unemployment and net wages	132
10	The transition paths of vacancy-unemployment ratio and net wages with announcement	133
11	The set of Pareto improvements in terms of welfare gains (in %) in the baseline model	135

12	Properties of different POPI programs (baseline model)	136
13	A typical time path for capital taxes	137
14	The time paths of selected variables for three POPI plans in the baseline model. (Time is in years.)	138
15	The frontier of feasible equilibria in terms of welfare gains (in %) in the baseline model and in MM1	139
16	The frontier of Pareto improving feasible equilibria in terms of welfare gains (in %) in the model without labor choice (MM2)	140
17	Illustrative paths of capital in the model without labor choice (MM2)	141

Acknowledgements

First of all, I would like to thank my advisor Albert Marcet for his guidance and continuous support, which has always gone beyond strictly academic questions. This work has also benefitted from numerous discussions with Christian Haefke, Guido Lorenzoni, Michael Reiter, Sevi Rodríguez-Mora, and Jaume Ventura. Michael Reiter also deserves a special mention for generous help with numerical problems. Further, I have received helpful comments from Daron Acemoglu, Marios Angeletos, Almut Balleer, Olivier Blanchard, Chris Telmer, Judit Montoriol, Manuel Oechslin, Harald Uhlig, Paolo Vanin, and Fabrizio Zilibotti. Financial support from the German National Merit Foundation, the German Academic Exchange Service, and the European Recovery Program is gratefully acknowledged. Finally, special thanks go to my parents and my sister for their unremitting support and encouragement. The same is true for my friends Almut Balleer, Stefan Berger, Emilia García-Appendini, Philipp Grosse, Philipp Kircher, Judit Montoriol, Xavier Ordeñana, Thomas Rangel, Volker Reinthaler, and Philip Sauré.

Chapter 1

Introduction

Real-world agents are heterogeneous in all kinds of economically relevant aspects such as risk aversion, wealth, and productivity. Nevertheless, large parts of macroeconomic research have traditionally been conducted in the representative agent framework. This is so despite the well-known fact that aggregation holds under very restrictive conditions only. That is to say, typically the behavior of a representative agent does not accurately reflect that of a heterogeneous population. The representative agent framework is thus a simplification whose appropriateness depends on the specific issue investigated. Three cases in which explicit consideration of heterogeneity is crucial are analyzed in the chapters of this thesis. They are taken from three different areas - asset pricing, labor market policy, and optimal taxation - and in each of them heterogeneity has different effects: Clearly, it is an important factor in judging welfare implications and political feasibility of policy problems, as the chapters on labor market policy and optimal taxation will illustrate. However, as we will see in the chapter on asset pricing, heterogeneity may also importantly change the nature of equilibria and give rise to different dynamics. A brief overview of the three chapters is given in turn.

In chapter 2 we introduce *ex ante* heterogeneity of agents into a general equilibrium asset pricing framework with Epstein-Zin preferences. There are two types of agents who differ exclusively in their risk aversion. They trade in a stock, whose dividend is the only source of consumption, and in a short-term bond in zero net supply. In equilibrium the less risk averse agents are leveraged in the stock, and their share in the economy's wealth is positively correlated with the dividend shock. Loosely speaking, "average risk aversion" declines when dividend growth is strong, which implies lower expected excess returns. At the same time the price-dividend ratio rises. Thus, in line with the data, a high price dividend ratio predicts low future excess returns. Moreover, predictability of excess returns displays the empirically observed pattern of R^2 s rising with horizon. We manage to generate R^2 s of similar magnitudes as in the data at all horizons. Without heterogeneity, by contrast, the price-dividend ratio and the equity premium would be constant.

In chapter 3 heterogeneity takes a different form. Agents do not differ innately, but in the presence of incomplete markets idiosyncratic income shocks leave them with heterogeneous wealth levels and thus heterogeneous amounts of self-insurance. The setting is a matching model of the labor market in the Mortensen-Pissarides tradition, and income shocks are transitions between employment and unemployment. We study the transition dynamics and welfare effects of reducing unemployment benefits. The dynamic analysis reveals significant transition costs that comparative statics would miss. The main reason is that initially individuals have to increase savings to self-insure. Nevertheless moderate benefit reductions increase average welfare of workers. Gains are much larger when the reform is announced in advance or phased in optimally. Workers can then extract windfalls otherwise accruing to firms with filled jobs which stem from the jump in vacancy costs following an unexpected reform.

In chapter 4, which is joint work with Albert Marcet, we study the optimal path for capital and labor taxes in a dynamic economy with agents who are heterogeneous in their ratio of human to physical capital. In an otherwise standard model we concentrate on tax reforms that are both Pareto efficient and Pareto improving. Also, we assume the capital tax rate can never rise above its initial level and lump-sum transfers between agents are impossible. We study the whole path for taxes, including the transition, from a current status quo to the long run steady state. We find that introducing all these elements into the analysis changes considerably the nature of tax reforms. In particular, we find that capital taxes have to be maintained at their status quo level for at least about ten years in order not to harm poor agents. Labor tax rates, by contrast, are initially lowered greatly in order to boost capital accumulation while capital taxes are still high. We show that in the absence of a non-distortive means of redistribution heterogeneity imposes a severe constraint on the optimal policy that drives the solution much further away from the first best than in the standard case in which only non-distortive means of raising revenue are lacking.

Chapter 2

Asset Pricing with Heterogeneous Epstein-Zin Agents

2.1 Introduction

By now a host of stylized facts about asset prices and their dynamics have been established that are hard to match jointly in consumption based asset pricing models. To mention just the ones that will be of concern to this work: Excess returns are high despite the relative smoothness of consumption and at the same time the riskless real rate of interest is very low and stable. These are the famous equity premium and risk free rate puzzles.¹ Another well known property of stock prices is their 'excess volatility', i.e. the fact that prices are subject to swings much greater than what seems explicable from changing cash flow and interest rate forecasts.² While these puzzles refer to unconditional properties of asset prices and returns, a number of conditional properties of asset prices and returns has been established too. There is now strong evidence that equity premium and Sharpe ratio are high in recessions and low in booms. Moreover, variables such as the price-dividend ratio are subject to similar swings and can serve to predict excess returns. The predictive power of the price-dividend ratio as measured by the R^2 ranges from about 5% at a yearly horizon to more than 50% when excess returns are compounded over 10 or more years.³

There is an extensive literature addressing some or all of these puzzles. However, it seems fair to say that no fully satisfactory framework has been provided to date. Two lines of attack have been pursued in many contributions:⁴ One is to endow models with some feature that allows to separate agents' attitude towards intertemporal substitution from their attitude towards risk. To this end authors have either employed the recursive utility framework of Kreps-Porteus/Epstein-Zin⁵ or various forms of consumption habits.⁶ The other approach has been to introduce market

¹These terms were coined by Mehra and Prescott (1985*a*) and Weil (1989) respectively.

²See Campbell and Shiller (1988*a*) and Campbell and Shiller (1988*b*).

³See for example Fama and French (1988) and Fama and French (1989).

⁴Of course there are other less common approaches, like for example the recent departures from log-normal, random walk dividends as in Bansal and Yaron (2004) and Weitzman (2007).

⁵E.g. Epstein and Zin (1989) and Weil (1989). In this type of utility aggregator intertemporal substitutability and risk aversion are governed by separate parameters.

⁶Early contributions in this latter vein are Abel (1990), Abel (1999), Campbell and Cochrane (1999) and Constantinides (1990). Habits of the ratio type (c/h) make the interest rate less sensitive

incompleteness, acknowledging the fact that individual consumption is much riskier than aggregate consumption.⁷ Campbell and Cochrane (1999) in the habit tradition and Constantinides and Duffie (1996) with incomplete markets are the two most successful contributions to date in that they replicate a large variety of asset pricing facts, but - as stated frankly in Cochrane (2005) - both of them are deliberately reverse-engineered and should be regarded as clever proofs of existence of a solution rather than as the ultimate economic stories settling the issues.

In this paper we therefore take another step towards explaining the aforementioned puzzle through an economically plausible mechanism. We introduce heterogeneity in agents' risk aversion into a general equilibrium asset pricing framework with Epstein-Zin preferences. Our economy is endowed with one unit of a stock that produces a stochastic dividend in each period. Two types of agents trade in this stock and in a riskless short-term bond that is in zero net supply such as to adjust their exposure to return risk to their risk preferences. In equilibrium the less risk averse agents are leveraged in the stock, and their share in the economy's wealth is positively correlated with the dividend shock. Loosely speaking, "average risk aversion" declines when dividend growth is strong, which implies lower expected excess returns. At the same time the price-dividend ratio rises, provided the intertemporal elasticity of substitution is greater than one. Thus, in line with the data, a high price dividend ratio predicts low future excess returns. Moreover, predictability of excess returns extends over many periods and displays the empirically observed pattern of R^2 s rising with horizon. Quantitatively, we manage to generate R^2 s of similar magnitudes as in the data at all horizons.

The model just described can only be solved numerically. However, we can analytically solve a set of simpler models that isolate certain features of the full model. These models provide a lot of intuition and help to pin down precisely which aspects

to risk aversion. Habits of the difference type ($c - h$) allow to choose a low value of the risk aversion parameter that is compatible with a low interest rate by raising the curvature of the marginal utility for a given risk aversion parameter.

⁷See for example Constantinides and Duffie (1996), Heaton and Lucas (1996), Marcet and Singleton (1999), and Telmer (1993).

of the full model drive which features of the solution. We thus show that both heterogeneous risk aversion and Epstein-Zin utility are essential ingredients in generating these results. The first simplified model illustrates how heterogeneous risk aversion causes time variation in equity premium and price-dividend ratio. The second model shows why Epstein-Zin utility allows the price-dividend ratio to predict excess returns with the right (negative) sign, while under standard CRRA utility, which ties together risk aversion and intertemporal willingness to substitute, it would not be possible to get the sign of predictability right unless risk aversion is unreasonably low. Finally, a third simplified version of our model helps to explain why the R^2 s of our long-horizon regressions rise strongly with horizon, as in the data. For this to happen shocks need to have very persistent effects. We demonstrate that when there is uncertainty about dividends in only one period, the effects of this shock persist into the entire future. In a model with a habit utility function, for example, this would not be the case.

Nevertheless, in our full dynamic framework shocks do not fully persist in asset prices either. The reason is that all the action in the model takes place in the transition to the steady-state in which the less risk averse agents own the entire economy and heterogeneity disappears with all its effects on asset price dynamics. The wealth share of the less risk averse agent grows because he earns a higher expected return rate on his portfolio. However, the return differential between the two agents' portfolios declines as the less risk averse type's wealth share increases because his leverage and the equity premium decrease. Thus once the less risk averse agent is sufficiently rich, his wealth share increases ever more slowly and converges to one. As this convergence becomes strong the paths of the economy after a shock and after no shock become less and less distinguishable and it seems as if the shock had faded away.⁸

As mentioned before, the related literature is huge. We therefore content ourselves with relating our work to a few recent contributions that are similar either technically

⁸It would be easy to eliminate this feature of the model in favor of a steady state with heterogeneity, in which asset prices and expected returns would continue to vary. One option would be to regard agents as dynasties in which with a certain probability children have different risk preferences than their parents. However, for the sake of clarity of exposition, we prefer to stick to the bare-bones version of the model at this point and leave this extension to future work.

or in focus. The model of Chan and Kogan (2002) is the one closest to ours. It also features heterogeneity in agents' risk aversion but utility functions are of the CRRA-type, defined over consumption relative to a slow moving external habit stock. Qualitatively their results are similar, but the predictive power of the price-dividend ratio for excess returns is negligible in their model and hardly increases with horizon. Moreover, with the help of our simplified models we are able to be much more precise about the intuitions for our results.

To the best of our knowledge the only other contribution to asset pricing featuring heterogeneous risk-aversion in combination with Epstein-Zin utility is Coen-Pirani (2005). His interest, however, is not in explaining broad sets of asset pricing facts. Rather his point is to show that margin requirements do not necessarily increase the volatility of stock prices. The use of Epstein-Zin utility seems to be motivated mostly by technical convenience.

Another contribution employing Epstein-Zin utility but closer to ours in thrust is Bansal and Yaron (2004). The model features a homogeneous agent and separate, exogenous processes for consumption and dividend growth. One main insight is that if these processes contain a small persistent component, which cannot be rejected from the data, moderate degrees of risk aversion are sufficient to generate a high equity premium, and the price-dividend ratio becomes volatile. The other insight is that fluctuating volatility in the driving processes further increases the risk premium and the volatility of the price-dividend ratio because under Epstein-Zin utility volatility risk is priced. Moreover, the equity premium and the price of risk then become time-varying and predictable from the price-dividend ratio.

Finally, Campbell and Cochrane (1999) is a classic that cannot be left unmentioned. Featuring habit utility of the difference type ($c - h$), it manages to generate counter-cyclical excess returns, predictability, and other features by making effective risk aversion counter-cyclical. In terms of its quantitative success at reproducing stylized asset pricing facts it is still the benchmark. However, the structure imposed on the way the habit evolves is very particular. Moreover, it does not solve the equity

premium puzzle either in the sense that the risk aversion required to match the equity premium is extremely high.

The remainder of this paper is organized as follows: Section 2.2 presents our dynamic heterogeneous model. In section 2.3 we take a step back and analyze our three simplified models in order to gain insights into the forces at work in our full model. Section 2.4.2 contains our numerical results for the full model. We first discuss the calibration of the model (section 2.4.1), then its qualitative properties (section 2.4.2), and finally present our quantitative results with a particular focus on long horizon predictability (section 2.4.2). Section 3.6 concludes.

2.2 The model

We consider an infinite-horizon exchange economy with two types of agents who differ exclusively in their risk aversion. Time is discrete. The wealth of the economy consists in one unit of a stock that produces a stochastic dividend in the form of a perishable consumption good each period. The dividend process constitutes the only source of uncertainty in this model. To be more specific:

The market structure: Agents can trade in two assets. One is the stock, whose total supply is one unit. Its ex-dividend value is equal to its (ex-dividend) price P . The other one is a riskless bond in zero net supply. Dividends D grow stochastically. In logs, they follow a random walk with drift

$$\Delta d' = \log\left(\frac{D'}{D}\right) = \mu + \epsilon, \quad (2.1)$$

where $\epsilon \sim N(0, \sigma^2)$ and i.i.d. over time.

The individual's maximization problem: Agents of both types aggregate utility from consumption streams according to the recursive preference specification of

Epstein-Zin. (Epstein and Zin 1989) They thus solve the following problem:

$$\begin{aligned}
V_i(W_i, \Gamma) &= \max_{\{C_i, \phi_i\}} \left[(1 - \delta)C_i^{1-1/\psi} + \delta[E(V_i(W'_i, \Gamma')^{1-\gamma_i})]^{\frac{1-1/\psi}{1-\gamma_i}} \right]^{\frac{\psi}{1-\psi}} \\
\text{s.t. } W'_i &= \left[\phi_i \frac{P(\Gamma') + D'}{P(\Gamma)} + (1 - \phi_i)R'(\Gamma) \right] (W_i - C_i) \\
\Gamma' &= G(\Gamma) \\
W'_i &\geq \underline{W}
\end{aligned}$$

In words, agents optimally choose their consumption C_i and their portfolio share of the stock, ϕ_i , conditional on their own wealth W_i and the aggregate state Γ , respecting their intertemporal budget constraint. This budget constraint states that an agent's wealth at the beginning of a period W'_i is equal to his investment in the previous period, $W_i - C_i$, times the gross portfolio return of agent i given his portfolio choice ϕ_i . $\underline{W} \leq 0$ is some lower bound on wealth which serves to rule out Ponzi schemes. Agents take the price of the stock P and the gross riskless rate of interest R as given and understand how they are determined in equilibrium as functions of the aggregate state Γ , which in equilibrium evolves according to the law of motion G .

We look for solutions where the state vector of the economy only contains the natural states (D, S_L, B_L) , with $S'_L = \phi_L(W_L - C_L)/P$ and $B'_L = (1 - \phi_L)(W_L - C_L)$ being the stock and bond holdings of the agent type with low risk aversion that result from the previous period's portfolio choice.⁹

Note that only γ , the coefficient of relative risk aversion, is indexed by i . This is where the two types differ. The elasticity of intertemporal substitution, ψ , on the other hand, is common to the two groups. It is well known that portfolio choice is largely governed by the parameter γ , while ψ primarily determines the intertemporal consumption profile.¹⁰ Thus, our specification of heterogeneity will lead to heterogeneous portfolios. Savings decisions, by contrast, will differ between agents only to the

⁹For the numerical solution we will be able to compact this state vector into a single dimension, which is the wealth distribution. See appendix 2.6.1.

¹⁰See Campbell and Viceira (2002), ch. 2.

extent that portfolio heterogeneity leads to different portfolio returns.

The distinction between risk aversion and elasticity of intertemporal substitution, which Epstein-Zin utility allows us to make, is crucial for our results even at a qualitative level. It not only enables us to simultaneously match the equity premium and the risk free rate without resorting to extremely high discount factors.¹¹ More importantly, choosing risk aversion and intertemporal elasticity of substitution separately is essential for the price-dividend ratio to predict equity returns with the correct - negative - sign. We will see this in detail in section 2.3.2. There it will also be shown how things would go wrong with CRRA utility.¹²

Competitive equilibrium: Due to the homotheticity of the utility function and the absence of non-tradable assets the consumption, stock holdings, and bond holdings of all agents of a certain type are proportional to their wealth level. We can therefore aggregate across all agents of a type such as to ignore individual wealth levels and focus on the wealth held by each group. Equivalently we can think of each group as consisting of a single agent who behaves competitively.

Equilibrium requires market clearing in the markets for consumption, stocks, and bonds. We index by L the variables referring to the type with the lower coefficient of risk aversion and by H those referring to the more risk averse type. The market clearing conditions at each state are then

$$C_L + C_H = D \quad \text{and} \quad (2.2)$$

$$\phi_L(W_L - C_L) + \phi_H(W_H - C_H) = P. \quad (2.3)$$

Equation (2.2) ensures clearing of the goods market. Equation (2.3) is the condition for stock market clearing. Bond market clearing follows by Walras' Law. Note that W_i denotes wealth at the beginning of the period, i.e. after dividends have been paid

¹¹We do need a high value of the risk aversion parameter, however.

¹²This is not to say that agents have to be homogeneous in their elasticity of intertemporal substitution. Within limits they may also differ in ψ . What we need is to be able to choose average risk aversion and average intertemporal elasticity of substitution separately. The role of these choices will be discussed in greater detail in sections 2.3.2 and 2.4.1.

and before consumption. Therefore, $W_L + W_H = P + D$. But investment takes place after consumption of the dividend, so total investment in the stock has to equal P , the ex-dividend value of the stock.

We are now ready to formally define a competitive equilibrium for this economy:

Definition 1 *A recursive competitive equilibrium in this economy is a set of consumption and portfolio rules $\{C_i(\Gamma), \phi_i(\Gamma)\}_{i=L,H}$, as well as a price function $P(\Gamma)$, an interest rate function $R'(\Gamma)$ and a law of motion for the state variables $G(\Gamma)$ such that*

1. *the allocation solves the individual optimization problem (2.2) for each agent type,*
2. *the market clearing conditions (2.2) and (2.3) are satisfied, and*
3. *$G(\Gamma)$ is consistent with individual choices.*

2.3 Three auxiliary models

The model we have just laid out requires numerical techniques for its solution. In order to gain some analytical insights and intuition for the forces at work, we turn to three simpler models as a first step in our analysis. The first of these auxiliary models is a two period model with heterogeneous agents, which will show how heterogeneous risk aversion leads to time-variation in the equity premium. The second model is the representative agent version of our infinite horizon model, which has closed form solutions for asset prices and returns that illustrate the relationship between equity premium and price-dividend ratio. Finally, as a third simplification, we analyze the model of section 2.2 with dividend risk in only one period. This will shed light on the persistence properties of our model.

2.3.1 A two-period model with heterogeneous agents

We consider a version of our model with only two investment periods and consumption taking place only after the second investment returns are realized. Even this little two-period model replicates several of the stylized asset pricing facts mentioned in the introduction: The equity premium varies 'counter-cyclically' in the sense that high growth rates are associated with low subsequent equity premia. Moreover, the stock price is 'excessively' volatile relative to a representative agent model, in which it would be constant. And finally, high prices 'predict' low future excess returns.¹³

The stock can now best be thought of as a tree that grows for two periods before it is chopped down and consumed. The tree's initial size is one. After the first period it has grown to X_1 , which is stochastic. Its final size is $X_1 \cdot X_2$, where X_2 is again stochastic and independent of X_1 . Initially, individuals choose portfolios of the tree and a riskless bond in zero net supply, given their initial wealth shares $w_{L,0}$ and $1 - w_{L,0}$. After the realization of X_1 they choose portfolios for the second period. The time line in figure 1 illustrates the sequence of events. Since there is no consumption at $t = 0, 1$, equilibrium only needs to determine the relative price between stocks and bonds, which will be denoted by P_t . The gross return to the bond is normalized to 1.

We do not specify any particular utility function but rather make a few plausible assumptions about the chosen portfolio shares $\phi_{i,t}$:

Assumption 2 *Agents' optimal portfolio shares of the stock are differentiable in their arguments risk aversion, γ_i , and equity premium, EP_t , with derivatives $\frac{d\phi_{i,t}}{d\gamma_i} < 0$ (for $EP_t > 0$) and $\frac{d\phi_{i,t}}{dEP_t} > 0$.*

The first assumption implies that for a given equity premium the less risk averse agent type has a greater share of stock in his portfolio. The second one will deliver the comparative statics with respect to the equity premium. Under these assumptions we can prove the following lemma, from which all the properties of asset prices mentioned above follow:

¹³This last point may be a bit of a stretch, since there is not much of a future.

Lemma 3 *If the initial wealth share of the less risk averse agent type, $w_{L,0}$, is sufficiently large or sufficiently small, a high growth rate in the first period, X_1 , raises the subsequent stock price P_1 and lowers the subsequent equity premium EP_1 .*

Proof. We derive how the equity premium during the second investment period depends on the realization of X_1 . The equity premium at this point is equal to the expected return to the stock, $E_1\left(\frac{X_2}{P_1}\right) - 1$, due to the normalization of the riskless rate. Hence, trivially, the stock price P_1 and the equity premium move in opposite directions. We proceed backwards in two steps. We first turn to the equilibrium at $t = 1$, i.e. when agents make their second portfolio choice, and establish that the equity premium for the second period depends negatively on the L-type's wealth share at this point, $w_{L,1}$. Then we show that, at least under certain circumstances, $w_{L,1}$ rises with the first period growth rate X_1 .

Market clearing in the stock market at $t = 1$ requires

$$\phi_1(\gamma_L, EP_1)w_{L,1} + \phi_1(\gamma_H, EP_1)(1 - w_{L,1}) = 1, \quad (2.4)$$

where $w_{L,1} = \frac{W_{L,1}}{P_1 X_1}$ is the wealth share of type L . The crucial question at this stage is, what happens as we vary $w_{L,1}$. By assumption 2 we have $\phi_{L,1} > \phi_{H,1}$ for any given equity premium. Thus a rise in $w_{L,1}$ goes along with an excess demand for stocks, which is eliminated by a rise in P_1 that reduces the equity premium such as to lower both types' portfolio share of stocks. Intuitively, the shift of wealth towards the L -type lowers 'average' risk aversion in the economy and therefore the equity premium.

Next we show how $w_{L,1}$ depends on the realization of X_1 , the growth rate of the stock in the first period. Notice that

$$w_{L,1} = \left[\phi_{L,0} + (1 - \phi_{L,0}) \frac{P_0}{P_1 X_1} \right] w_{L,0} \quad (2.5)$$

Moreover, market clearing at $t = 0$ and assumption 2 imply $\phi_{L,0} > 1 > \phi_{H,0}$. Thus,

the second term in brackets in equation (2.5) is negative and $w_{L,1}$ rises with X_1 , unless P_1 falls with an elasticity greater than one in absolute value. Substituting for $w_{L,1}$ in equation (2.4) from equation (2.5) and differentiating, we can calculate this elasticity as

$$\varepsilon(P_1, X_1) = \frac{(\phi_{L,1} - \phi_{H,1})(\phi_{L,0} - 1)w_{L,0}\frac{P_0}{P_1X_1}}{(w_{L,1}\frac{d\phi_{L,1}}{dEP_1} + (1 - w_{L,1})\frac{d\phi_{H,1}}{dEP_1})EP_1 - (\phi_{L,1} - \phi_{H,1})(\phi_{L,0} - 1)w_{L,0}\frac{P_0}{P_1X_1}} \quad (2.6)$$

The numerator and the second term in the denominator are equal and positive, and the first term in the denominator is positive as well. Thus we will find $\varepsilon(P_1, X_1) > 0$ if and only if $(w_{L,1}\frac{d\phi_{L,1}}{dEP_1} + (1 - w_{L,1})\frac{d\phi_{H,1}}{dEP_1})EP_1 > (\phi_{L,1} - \phi_{H,1})(\phi_{L,0} - 1)w_{L,0}\frac{P_0}{P_1X_1}$. This will for sure be the case for $w_{L,0}$ sufficiently close to zero or one, for in either case the right hand side of this condition goes to zero, while the left hand side is strictly positive as long as the L -type is not risk neutral. (Note that as $w_{L,0} \rightarrow 1$, $\phi_{L,0} \rightarrow 1$.) Thus, to recap, for $w_{L,0}$ small enough or large enough the first period growth rate X_1 affects the stock price P_1 positively and the equity premium EP_1 negatively. ■

The proof has shown clearly that at the heart of the variability of the equity premium and the (excess) volatility of the stock price there is a relative wealth effect between the agent types. It exists because they choose different portfolios. For comparison, if agents were not heterogenous, i.e. if we had $\phi_{L,t} = \phi_{H,t}$, we would have $\varepsilon(P_1, X_1) = 0$, and consequently the equity premium EP_1 would be constant.

Alternatively, we can interpret what is going on in this model as consumption insurance. The less risk averse agent type insures the more risk averse type. His share of final consumption will hence be higher the bigger total consumption $X_1 \cdot X_2$. Notice that the expectation of final consumption at $t = 1$, $E_1(X_1 \cdot X_2)$, rises with the first period growth rate X_1 . Consequently, L -types should expect a higher share of final consumption after high realizations of X_1 than after low ones.¹⁴ This is the case if their share of wealth, $w_{L,1}$, is higher after high X_1 .¹⁵

¹⁴If X_1 did not contain any information about the distribution of final consumption this would not be the case.

¹⁵At this level of generality we cannot rule out the alternative that $w_{L,1}$ is actually lower and

2.3.2 The infinite horizon model with a representative agent: comparative statics

In the two period model we saw how with heterogeneous agents shocks effectively shift the degree of risk aversion of the economy. In this section we will use this insight to make a short-cut that permits an analytical investigation into the nature of predictability in the infinite horizon model. While our dynamic heterogeneous agent model of section 2.2 is not amenable to analytical solution, its representative agent version features simple closed form solutions for asset prices and returns. We state these and do comparative statics with respect to the coefficient of relative risk aversion. Effectively, we ignore the non-linearities that prevent aggregation in the heterogeneous agent model and discuss the effect of a shock that once and for all shifts the wealth distribution such as to precipitate a certain change in average risk aversion. This analysis illustrates that working with Epstein-Zin utility rather than with the more standard CRRA form is crucial in order for the price-dividend ratio to predict excess returns with a negative sign, as in the data.

If we eliminate heterogeneity from the model of section 2.2 but maintain the assumption that dividend growth is i.i.d. we can derive analytical expressions for asset prices and returns. The equity premium equals

$$EP = \gamma\sigma^2. \tag{2.7}$$

The gross risk free rate is determined as

$$R = \delta^{-1} \exp \left[\frac{1}{\psi} \left(\mu + \frac{1}{2} \sigma^2 \right) - \frac{1}{2} \left(1 + \frac{1}{\psi} \right) \gamma \sigma^2 \right]. \tag{2.8}$$

subsequent portfolio returns are higher. But this case seems rather unintuitive and never arose in the numerical solutions to the dynamic model.

And finally, the price-dividend ratio can be expressed as

$$PD = \frac{\tilde{\delta}}{1 - \tilde{\delta}}, \quad (2.9)$$

$$\begin{aligned} \text{where } \tilde{\delta} &= R^{-1} \exp[-EP] \exp\left[\mu + \frac{1}{2}\sigma^2\right] \\ &= \delta \exp\left[\left(1 - \frac{1}{\psi}\right)\left(\mu + \frac{1}{2}\sigma^2\right)\right] \exp\left[\left(\frac{1}{\psi} - 1\right)\frac{1}{2}\gamma\sigma^2\right]. \end{aligned}$$

Equations (2.7) to (2.9) show that equity premium, risk free rate and price-dividend ratio are all constant in the representative agent model, ie. there is no predictability. Nevertheless, their comparative statics with respect to the coefficient of relative risk aversion, γ mimic how predictability and the like arise in the heterogeneous agent version.

Turning to these comparative statics, the equity premium clearly increases in risk aversion γ , while the risk free rate falls. The latter is a consequence of the increasing precautionary savings motive. Being the expected discounted sum of all future dividends, discounted at the risk adjusted rate and normalized by the current dividend, the price-dividend ratio depends negatively on both the risk free rate and the equity premium. An increase in γ lowers the risk free rate but raises the equity premium, as equations (2.7) and (2.8) reveal. The reaction of the price-dividend ratio is therefore ambiguous. Whether the fall in R or the rise in EP dominate depends on the intertemporal elasticity of substitution, ψ . The rise in the equity premium dominates and the price-dividend ratio falls if and only if $\psi > 1$. This is the case we have to focus on if we want to generate predictability with the right sign in our heterogeneous agent model. For, empirically, a high price-dividend ratio predicts low excess returns.¹⁶

Such a high value for the intertemporal elasticity of substitution seems to fly in the face of micro-evidence suggesting that ψ should be close to zero. (Hall 1988) However, the choice of $\psi > 1$ has other arguments in its favor beyond the pattern

¹⁶In a different model Bansal and Yaron (2004) work with an elasticity of intertemporal substitution greater one for the same reason.

of excess return predictability. For example, we know that the interest rate is very stable. But low values of the intertemporal elasticity of substitution make the interest rate very reactive to small variations in expected consumption growth. (Cochrane (2005), chapter 21) Taking the same short-cut that we have taken above for the case of changes in γ again for changes in the growth rate μ , this can also be seen from equation (2.8). Also, at $\psi < 1$ we would have that the price-dividend ratio is lower the stronger growth, which is counterfactual. (See equation 2.9.) In sum, in the case of the intertemporal elasticity of substitution, like in that of the coefficient of risk aversion, it seems that micro-evidence and the values implied by asset prices are plainly incompatible. In the asset pricing literature that focuses on conditional moments of asset prices rather than on the equity premium puzzle the micro-evidence is therefore typically ignored. High elasticities of intertemporal substitution (e.g. Bansal and Yaron (2004)) and high degrees of risk aversion (e.g. Campbell and Cochrane (1999)) are employed as need be to match features of asset prices.

What would happen with CRRA-utility? Imposing $\gamma = 1/\psi$ in equation (2.9) and differentiating with respect to the risk aversion coefficient shows that the price-dividend ratio predicts excess returns with the right sign only if $\gamma\sigma^2 < \mu + \sigma^2$, i.e. only if the equity premium is smaller than a number that is hardly bigger than the average growth rate of the economy.¹⁷ This would not come out of any reasonable calibration, for the right hand side of this condition will typically be around 1-3%, while estimates of the equity premium are always in excess of 4%. Thus, to recap, we need to employ Epstein-Zin utility with an intertemporal elasticity of substitution in excess of one in order to generate a negative relationship between the price-dividend ratio and future excess returns, as observed empirically.¹⁸

¹⁷The expected growth rate would be $\mu + \frac{1}{2}\sigma^2$, but σ^2 is an order of magnitude smaller than μ and hence does not matter much.

¹⁸Notice that our reasoning is true only for pure CRRA utility. CRRA with a habit can yield the right correlation between price-dividend ratio and equity premium because the habit introduces an element of mean reversion: a good shock raises the surplus of consumption over its habit level but creates the expectation of a reduction in the surplus as the habit adjusts. This lowers the interest rate and raises the price dividend ratio. If at the same time a force such as agent heterogeneity lowers average risk aversion the resulting correlation between price-dividend ratio and equity premium is negative as in the data. This is the - unstated - intuition behind the results of Chan and Kogan

2.3.3 Uncertainty in a single period

We now return to our dynamic heterogeneous agent model of section 2.2, however, there will be uncertainty about dividends in one period only. The purpose of this simplification is to show analytically that shocks have permanent effects on the allocation of consumption between agents and thus on the wealth distribution. We first argue that after the resolution of uncertainty agents consume the same fraction of the dividend forever. Then we show that this fraction depends on the realization of the dividend in the period of uncertainty. In this simplified setting we cannot make meaningful statements about the effect of the shock on asset prices because after the resolution of uncertainty the stock is a safe asset whose price is not influenced by agents' risk attitudes. However, we take the shift in the wealth distribution as an indication that in a model with repeated uncertainty we would observe permanent asset price effects through the shift in average risk aversion.

For the sake of precision and clarity we frame our discussion in the following lemma, which will be proved in turn:

Lemma 4 *Let dividends be log-normally distributed in period s and let them grow at a constant rate μ thereafter. Then the realization of the dividend shock, D_s , permanently affects the distribution of consumption and hence of wealth between agents.*

Proof. For all $t \geq s$ markets are trivially complete and stochastic discount factors equalize across agents. The stochastic discount factor of agent type i is

$$m_{i,t+1} = \delta \left(\frac{U_{i,t+1}}{E_t(U_{i,t+1}^{1-\gamma_i})^{\frac{1}{1-\gamma_i}}} \right)^{1/\psi - \gamma_i} \left(\frac{C_{i,t+1}}{C_{i,t}} \right)^{-1/\psi}. \quad (2.10)$$

For all $t \geq s$ the terms involving U cancel and the condition $m_{L,t+1} = m_{H,t+1}$ collapses to

$$\frac{C_{L,t+1}}{C_{L,t}} = \frac{C_{H,t+1}}{C_{H,t}}, \quad \forall t \geq s. \quad (2.11)$$

(2002).

Hence, agents consume the same fraction of the dividend forever from period s on. Call the fraction consumed by the less risk averse type $c_L = C_L/D$.

It remains to be shown that c_L depends on D_s . To do so we will use the two pricing conditions for the stock and the bond, which are

$$E_t(m_{i,t+1} \frac{P_{t+1} + D_{t+1}}{P_t}) = 1 \quad \text{and} \quad (2.12)$$

$$E_t(m_{i,t+1})R_{t+1} = 1, \quad (2.13)$$

and the fact that they have to be satisfied for both agents. For ease of exposition let us assume that dividends grow at a constant rate μ after period s , i.e. that the shock is constant. We can then write utility as $U_{i,s} = \left(\frac{1-\delta}{1-\delta \exp[\mu]}\right)^{\frac{\psi}{\psi-1}} C_{i,s}$, and the stochastic discount factor simplifies to $m_{i,s} = k_i C_{i,s}^{-\gamma_i}$, where k_i summarizes all terms that are determined before time s . Also, the stock price P_s will be proportional to D_s .

From now on we proceed by contradiction and show that the distribution of consumption cannot be independent of D_s . Suppose to the contrary that irrespective of the realization of D_s the less risk averse type has consumption share $c_L = \bar{c}_L$. Then setting the left hand sides of equations (2.12) and (2.13) equal for both agent types and canceling some terms we get

$$k_L \bar{c}_L^{-\gamma_L} E_{s-1}(D_s^{-\gamma_L}) = k_H (1 - \bar{c}_L)^{-\gamma_H} E_{s-1}(D_s^{-\gamma_H}) \quad \text{and}$$

$$k_L \bar{c}_L^{-\gamma_L} E_{s-1}(D_s^{1-\gamma_L}) = k_H (1 - \bar{c}_L)^{-\gamma_H} E_{s-1}(D_s^{1-\gamma_H}),$$

which implies

$$\frac{E_{s-1}(D_s^{-\gamma_L})}{E_{s-1}(D_s^{1-\gamma_L})} = \frac{E_{s-1}(D_s^{-\gamma_H})}{E_{s-1}(D_s^{1-\gamma_H})}. \quad (2.14)$$

Using the log-normality of D_s , this expression can be simplified to yield

$$\gamma_L = \gamma_H$$

which contradicts our assumption that agents are heterogeneous and $\gamma_L < \gamma_H$.

Thus c_L depends on D_s . Moreover, the wealth of agent i at time s is¹⁹

$$W_{i,s} = \frac{1}{1 - \delta} U_{i,s}^{1-1/\psi} C_{i,s}^{1/\psi} = \frac{C_{i,s}}{1 - \delta \exp[\mu]}$$

which implies that the wealth share of the less risk averse agent type is simply $w_L = c_L$.

This completes the proof. ■

At this level of generality it seems impossible to make stronger statements. Under the assumption that the dividend process can take only two realizations such as to make markets complete at all times²⁰ one can further show that the wealth and consumption shares of the L-type increase in D_s , which is in line with our findings for the two period model of section 2.3.1.

As mentioned before, the wealth shift that is precipitated by the shock in this simple model has no consequences for asset prices because after the shock there is no more uncertainty and hence risk aversion does not affect the pricing of the stock. Nevertheless, to the extent that the strong persistence property that we have found carries over to the full model, it is good news for the potential of the model to generate long horizon predictability. For it suggests that a shock will change expected excess returns not only today but for the entire future. For comparison, under CRRA with habits, as employed for example in Chan and Kogan (2002), a one-off shock to the level of dividends would have only transitory effects because everything mean-reverts as the habit catches up with consumption.

2.4 Numerical analysis

The three auxiliary models of section 2.3 have provided us with intuition and analytical insights into the mechanisms at work in our model. Against this backdrop we can now present the numerical results for the full dynamic model. These will be both

¹⁹See Cochrane (2006) for this and other useful derivations involving Epstein-Zin utility.

²⁰Appendix 2.6.2 discusses conjectures about the general relationship between our model and its complete markets version.

qualitative and quantitative in nature. We therefore first discuss the calibration we choose.

2.4.1 Calibration

We fix the parameters of our model on the basis of the annual US-data for the years 1890-1995 that is also used in Campbell (2003).²¹ The consumption data refers to real per-capita consumption of non-durables and services, stock market data is based on the S&P 500, and the real interest rate is derived by deflating six-months commercial paper, bought in January and rolled over in June.²²

The first important choice we have to make regards the parameters of the dividend growth process. While in reality dividends are only a small, but very volatile component of consumption, in our model dividends and aggregate consumption are identical. Unlike contributions like Campbell and Cochrane (1999), which work with a representative agent, we have to impose this equilibrium condition and cannot simply price the stock given separate processes for consumption and dividends.²³ In keeping with the related literature²⁴ we opt for using consumption data for the main calibration of the model. (See the first column of table 1.) As a check we also use dividend data. (See the second column of table 1.) Unfortunately, however, we cannot solve the model with dividend data for comparable levels of agent heterogeneity. The dilemma in the choice between consumption and dividend data is the following: If we calibrate the dividend process to consumption data, it will be very hard to get the volatilities of price-dividend ratio and returns even approximately right because consumption is much less volatile than dividends.²⁵ On the other hand, if we base our calibration on dividend data consumption becomes far too volatile, such as to

²¹This data can be downloaded from <http://kuznets.fas.harvard.edu/campbell/data.html>.

²²For greater detail consult the file `readmeus.txt` in the data set.

²³An - albeit complicated - solution would be to work with leveraged equity and bonds in endogenous positive supply. This avenue will be followed in future work.

²⁴E.g. Chan and Kogan (2002).

²⁵In fact, matching the empirical values of these volatilities would be of questionable desirability since it would mean to introduce too much heterogeneity. This criticism applies for example to Chan and Kogan (2002).

seemingly resolve the equity premium puzzles at low levels of risk aversion.

We next explain how we fix the preference parameters of our model. A summary of the resulting parameter values is contained in table 1. We choose average risk aversion,²⁶ time preference rate, and elasticity of intertemporal substitution such as to match average excess returns, the average safe rate, and the average price-dividend ratio in our data set. Instead of iteratively simulating our model to match the empirical values, we take the short-cut of inferring the parameters from the representative agent version of section 2.3.2, which yielded analytical solutions for the equity premium, the risk free rate, and the price-dividend ratio. Specifically, average risk aversion, call it $\bar{\gamma}$, is chosen to match the equity premium. Since in our data we measure the average log excess returns we have to adapt equation (2.7), which refers to levels. The resulting expression is $(\bar{\gamma} - 1/2)\sigma^2$. The risk free rate and the price-dividend ratio cannot be determined separately, as equations (2.8) and (2.9) reveal. We therefore fix δ at .96 per year, in keeping with a lot of macroeconomic literature, and choose ψ such as to approximately match both R and PD . The resulting values are reported in table 2 together with other unconditional moments. Admittedly, this method only permits an approximate matching of the asset pricing data in our full model, however, as we will see the deviations are an order of magnitude below the variation in parameter estimates from the different data sets that have been used in the literature.

The representative agent model only helps to choose average risk aversion. How does this concept relate to the individual risk aversions of the two agent types? It turns out that defining average risk aversion as the harmonic mean of type specific risk aversions, weighted by the types' wealth shares, yields a good approximation of the equity premium. To see why this is so think of the representative agent model as the limit of our heterogeneous agent model as γ_L and γ_H converge. In this limit consumption and returns are log-normal and i.i.d., such that the rules of myopic

²⁶How average risk aversion relates to γ_L and γ_H will be discussed below.

portfolio choice apply²⁷ and we have

$$\phi_i = \frac{EP}{\gamma_i \sigma^2}. \quad (2.15)$$

In the neighborhood of the homogeneous agent limit returns will not be too far from log-normal and i.i.d., so this portfolio rule will still be approximately true. Use it in the condition for market clearing in the stock market, $\phi_L w_L + \phi_H (1 - w_L) \equiv \bar{\phi} = 1$,²⁸ and cancel terms to get

$$\frac{1}{\gamma_L} w_L + \frac{1}{\gamma_H} (1 - w_L) = \frac{1}{\bar{\gamma}}. \quad (2.16)$$

Equation (2.16) suggests that, once we have fixed $\bar{\gamma}$, there are still two free parameters to pin down. However, the wealth distribution between agents is not really a choice. Rather, we can only fix the initial wealth distribution. As the model is fed with shocks, it evolves endogenously, and that in a non-stationary manner. The non-stationarity of the model will be discussed below in section 2.4.2. At this point it suffice to say that we will experiment with different combinations of initial wealth distribution and risk aversion levels that approximately reproduce the first moments of our data. In choosing the free risk aversion parameter we strive to generate enough agent heterogeneity to quantitatively match the long horizon predictability we find in our data. The resulting values, $\gamma_L = 36$ and $\gamma_H = 98.14$ for our benchmark calibration, may seem very far apart, suggesting that we impose a huge amount of heterogeneity. However, judging from the resulting portfolios this is not true, as we will see in section 2.4.2. The values for the calibration to dividend data are as far apart as numerically possible.

To generate the artificial asset pricing data necessary for calculating moments and running predictive regressions, we use repeated simulations of our model at quarterly frequency over a period of 100 years, roughly corresponding to the length of the data set. We alternatively tried simulating at annual or monthly frequencies, neither of

²⁷See Campbell and Viceira (2002) for details.

²⁸To be precise, w_L must be the post-consumption wealth distribution.

which changes our results in a significant way. The higher the frequency, the smaller is the departure from dynamically complete markets, but the higher the computational effort. This is why we decided in favor of quarters as a middle way.

2.4.2 Results

We present the results for our full model in two steps. Section 2.4.2 is of a mostly qualitative nature and serves to illustrate the wealth effect in our full model. The quantitative assessment of our model is deferred to section 2.4.2, where we present and discuss statistics obtained by simulating the model.

Heterogeneity and the wealth effect

In the following we elicit how and to what extent the insights from the three auxiliary models carry over to the dynamic heterogeneous agent case. To this end we graphically illustrate how various aspects of the solution such as the price-dividend ratio and the equity premium depend on the distribution of wealth between the two agent types. The wealth distribution, summarized by the wealth share of the less risk averse agent type w_L , is the only state variable in our numerical solution. Appendix 2.6.1 provides details on our choice of state space and the numerical solution.

Figure 2 traces out the equity premium and the price-dividend ratio as functions of the share of wealth held by the agent type with low risk aversion for our benchmark calibration. As predicted by the two-period model of section 2.3.1, the more wealth is held by the less risk averse agents, the lower is the risk premium. The price-dividend ratio, by contrast, rises with the wealth share of the L-types. This reflects our choice of an intertemporal elasticity of substitution greater than one, as discussed in the context of the representative agent model of section 2.3.2. Note that in the limit as w_L approaches zero or one the values of equity premium and price-dividend ratio equal their representative agent counterparts for risk aversions γ_H and γ_L respectively. This is not surprising because at these limits heterogeneity disappears and we are in

the respective representative agent worlds.

Correspondingly, at the extremes of the set of wealth distributions, the portfolio share of the stock of the agent type who holds all the wealth approaches one. This is illustrated in figure 3.²⁹ At the same time the other agent type's portfolio share approaches $\frac{\gamma_j}{\gamma_i}$, where j is the wealthy agent.³⁰ At interior wealth distributions, on the other hand, there is true heterogeneity, and the less risk averse agents hold leveraged positions in the stock, while the more risk averse ones hold a mixture of stock and bonds. Thus, the less risk averse agents take on a disproportionate amount of risk, in line with the insurance considerations that we discussed in section 2.3.1. Combining figures 2 and 3 we can now infer for the dynamic model the result that we proved in lemma 3 for the two period illustration: Since the less risk averse agents are leveraged in the stock (figures 3 and 4), their wealth share increases after good shocks, which lowers expected future excess returns and raises the price-dividend ratio (figure 2). Note that in speaking of wealth shares, we mean the wealth distribution before consumption. A good shock shifts pre-consumption wealth towards the low risk aversion types, and pre-consumption w_L is negatively (positively) related to the equity premium (the price-dividend ratio). This is necessary in order to be sure that consumption choices, which were absent in the two period model, do not pervert the effect.

In section 2.3.2 we invoked the simple solutions to the representative agent version of our model and did comparative statics with respect to the coefficient of risk aversion, arguing that the effects would be similar to those of varying the wealth

²⁹Interestingly, the L-type's leverage does not monotonously decrease in his wealth share even though the equity premium does so. This feature could be due to the deviations from i.i.d. log-normality of returns and consumption that are caused by heterogeneity. (See discussion below.) The amount of stock held by the L-type, on the other hand, increases smoothly with his wealth share. (See figure 4.) We checked very carefully that the non-monotonicity in the portfolio graph is not due to numerical inaccuracies. In particular, we tried approximating different objects. Also, we made sure that the risk of bankruptcy, which exists in discrete time unlike in the continuous time limit, does not cause the hump. To this end we solved the model at monthly and smaller trading intervals. The hump does not go away but if anything becomes slightly bigger as we get closer to continuous time.

³⁰This follows from equation (2.15), which implies $\phi_i = \frac{\gamma_i}{\gamma_i} \phi_j$ and the fact that $\phi_j = 1$.

distribution in the heterogeneous agent version. We now investigate how far this similarity goes. To this end we compare the equity premium, the Sharpe ratio, and the volatility of excess returns (all in dependence on w_L) to their counterparts in the representative agent model for different risk aversions. Figure 5 provides a graphical comparison. The solid lines represent the graphs for the heterogeneous agent model, the dashed ones are the result of specifying risk aversion according to equation (2.16) in the representative agent version of section 2.3.2.³¹ In the first panel the two graphs are virtually identical and hence overlap. Thus for the Sharpe ratio, i.e. for the price of risk, we practically have aggregation. That is to say, the price of risk in the heterogeneous agent economy with a wealth distribution described by w_L and type specific risk aversion levels γ_L and γ_H is identical to the price of risk in a representative agent economy with risk aversion $\gamma = \left[\frac{1}{\gamma_L} w_L + \frac{1}{\gamma_H} (1 - w_L) \right]^{-1}$. The equity premium, on the other hand, is always higher in the heterogeneous agent economy, as the second panel shows. This can be explained from the graph for the volatility of excess returns at the bottom of the figure. Heterogeneity introduces extra volatility in the economy, which implies a higher equity premium for a given price of risk. In the representative agent economy the risk free rate is constant and so is the price-dividend ratio, such that the volatility of excess returns is equal to the volatility of dividends. Heterogeneity introduces volatility into the price dividend ratio (and to a minor extent into the risk free rate), which makes excess returns more volatile than dividends.³²

Excess return volatility in the heterogeneous agent model is hump shaped, again with the limits at the two extremes of the wealth distribution being identical to the respective representative agent versions. It is a good indicator of the strength of the wealth shifts between the agents in different regions of the wealth distribution. For, as mentioned previously, return volatility in excess of the volatility of dividend

³¹To illustrate what we do: E.g. the dashed line for the equity premium represents equation (2.7) with equation (2.16) plugged in for γ , i.e. $EP = \left[\frac{1}{\gamma_L} w_L + \frac{1}{\gamma_H} (1 - w_L) \right]^{-1} \sigma^2$.

³²Still, at least for our calibration to consumption data and for the degree of heterogeneity we have imposed, the volatility of excess returns is far below its empirical counterpart of 18.5% per year. This is why the Sharpe ratio is far too high when the equity premium is in the right range.

growth is due to volatility in the price-dividend ratio, which in turn comes about by shifts in wealth between the two agent types. (See figure 2, panel 1.) Thus a given dividend shock has the strongest effect on the wealth distribution in its intermediate range. To see why this is so recall figure 3, which depicts the share of the stock in agents' portfolios across the range of wealth distributions. When the L-type owns almost all wealth his portfolio displays only little leverage. Hence shocks do not affect the wealth distribution very much. At the other extreme, when the L-type is very poor, he is very leveraged and thus very exposed to return risk, however, since he has so little wealth, the wealth distribution does not move very much in absolute terms either. Hence, in order to have a lot of effective heterogeneity for a given choice of risk aversion levels we should focus on intermediate, but lower rather than higher, values for w_L . This insight will guide us in the choice of initial wealth distribution for our simulations.

In section 2.3.2 we also argued that it was crucial to choose a value of the intertemporal elasticity of substitution in excess of one in order for the price-dividend ratio to predict excess returns with the right sign. Figure 6 confirms that this insight from the representative agent model carries over to the model with heterogeneity. The three panels graph the price dividend ratio, the equity premium and the volatility of excess returns as functions of the wealth distribution. Our benchmark calibration is depicted in dots. For comparison we plot the same functions for elasticities of substitution of one (in dashes) and .75 (solid lines). For the latter the price-dividend ratio falls as the wealth share of the less risk averse agents rises, while the curve for the equity premium retains its negative slope across all values of ψ . Thus a positive shock that increases w_L will lower both price-dividend ratio and future excess returns, contrary to the evidence on predictability. The fact that shocks and price dividend ratio are negatively correlated when the elasticity of intertemporal substitution is less than one also explains why in this case the volatility of excess returns is actually lower than without heterogeneity. When ψ is exactly one the price-dividend ratio is constant. Correspondingly, the volatility of excess returns is constant too at the

level of the volatility of dividend growth and the behavior of the equity premium is indistinguishable from its counterpart in our 'mock-heterogeneity' model in which we vary risk aversion in the representative agent model.³³

A fundamental difference between the dynamic heterogeneous agent model and its simplified versions of sections 2.3.2 and 2.3.3 regards stationarity. As mentioned previously, our model is non-stationary in the sense that the wealth distribution drifts over time. For realistic calibrations the wealth share of the less risk averse agents approaches one in the long run, and the only stable stationary steady state is at $w_L = 1$, i.e. in the representative agent limit with $\gamma = \gamma_L$.³⁴ Why is this so? Since the L-types have a higher share of stock in their portfolios and stocks command a return premium, they earn a higher return on their invested wealth. Their propensity to consume out of wealth, on the other hand, is lower precisely because their portfolio returns are higher and the elasticity of intertemporal substitution is greater than one.³⁵ Both forces lead the less risk averse agents to accumulate wealth faster than the more risk averse ones. However, the average rate of change of the wealth share of the less risk averse agent type is not constant. Rather, the pattern is hump shaped, similarly to that of the volatility of excess returns. The reasons are also similar: While the return advantage of the less risk averse agents tends to decline as their wealth share grows, this effect is more than compensated at low levels of w_L by the increase in wealth on which this return advantage acts.

The drift in the wealth distribution also has consequences for the persistence

³³Nevertheless, even in this special case there is no obvious aggregation result for the true heterogeneous agent model. The correspondence between the 'mock' and 'true' versions is exact only for the price dividend ratio and stock returns. To see this first note that with a unit elasticity of substitution income and substitution effects exactly cancel and each agent's consumption is a constant fraction $1 - \delta$ of his wealth. Market clearing for the consumption good then implies a constant price-dividend ratio, which in turn implies that stock returns are log-normal and i.i.d. Nevertheless, due to the time variation in the interest rate portfolio choice is not myopic.

³⁴For certain combinations of (very low) risk aversion coefficients it would be possible for the more risk averse agents to dominate in the long run. In those cases, even though the less risk averse agents will always earn higher expected simple returns on wealth, their log returns will be lower. See Coen-Pirani (2004) for details.

³⁵With CRRA utility this effect would be even more pronounced because the less risk averse agents would also be more willing to substitute intertemporally.

properties of our model. Recall that in our simplified model of section 2.3.3, where dividends were shocked only once, those shocks had fully persistent effects on the economy. The drift in the model with recurrent dividend uncertainty takes away some of that persistence. To see this, note that, like the wealth share of the L-type, all asset price moments converge to their counterparts in the representative agent model with $\gamma = \gamma_L$ in the long run. Like the wealth distribution, they move strongly as long as effective heterogeneity is big, i.e. in that intermediate range of w_L referred to previously, and then change ever more slowly to approach their steady state values asymptotically. Now consider a positive shock. All this shock does is to take the economy a little closer to its long run steady-state. This changes asset prices and their moments noticeably at first, but the effect decreases over time and vanishes asymptotically as the economy approaches $w_L = 1$. Figure 7 illustrates this behavior through a set of impulse responses. To obtain them we simulated the economy, starting at $w_L = 0.1$, setting all shocks to zero except for one after 10, 50 or 90 quarters, which we set to one standard deviation. The responses are graphed as the difference between time series with and without the respective impulse. This representation shows very clearly how the effect vanishes over time. Moreover, shocks that occur early on are more persistent than later ones and can even build up. This is because they occur at a time when the moderating effects of convergence are not dominant yet.

It is clearly an extreme implication of our model that predictability is an entirely transitory phenomenon. It could be avoided by viewing agents as dynasties whose future generations will have different risk preferences with a certain probability. This modification would introduce an element of mean reversion and hence allow for an interior stochastic steady state in which the wealth distribution and hence the equity premium and the price-dividend ratio still respond to dividend shocks. It would come at the price of taking away some persistence and thus some long horizon predictability. For the sake of clarity of exposition and in the light of a recent trend to view the past decades of asset pricing data featuring a high but declining equity premium as

a transitory period rather than as a steady state,³⁶ we prefer to evaluate the model during the transition and regard the extension to dynasties as an extension left for future work.

Simulation results

In order to compare our model to the data we simulate it for 1000 series of 100 years (400 quarterly draws of the shock) each.³⁷ Table 2 reports basic unconditional statistics. The columns below 'Model' refer to our benchmark calibration, simulated from the indicated initial wealth distributions. As explained in section 2.4.1, we do not strive to match prices and returns to a high precision because data statistics vary so widely between data samples. Also, as table 2 clearly shows, due to the drift inherent in our model the exact values depend on the choice of starting value. So the way to interpret table 2 is as showing that we are rather close to the data for the means and the autocorrelation of the price-dividend ratio, but far off in terms of the volatilities. We consider problematic only the low volatilities of excess returns and price-dividend ratio. The historical standard deviation of the risk free rate seems very high anyway and is likely due to ex post inflation. Campbell (2003) argues that the volatility of the *ex ante* real risk free rate should be very low. The low volatilities of excess returns and price-dividend ratio are at least partly due to our choice of data for the driving process. We work with consumption data, even though dividends are much more volatile than consumption.³⁸ The fact that excess returns are more volatile than consumption growth and that the price dividend ratio covaries positively with consumption growth (otherwise excess returns would be less volatile than Δc) indicates that our model goes in the right direction of creating "excess volatility". This reiterates what we saw graphically in section 2.4.2.

Table 3 presents long-horizon regressions of log excess stock returns on the log

³⁶See for example Cogley and Sargent (2005).

³⁷We use these short run simulations instead of a long run simulation for the same reasons as Adam, Marcet and Nicolini (2006): Heterogeneity and hence predictability is a transitory phenomenon in the model and vanishes in the long run. See Adam et al. (2006) for details on the method.

³⁸We discuss this choice with its pros and cons in section 2.4.1.

price-dividend ratio in historical and simulated data. We report the slope coefficients and the R^2 s up to horizons of 10 years. As mentioned in section 2.4.1, we choose the spread between the two types' risk aversion coefficients such as to generate predictability, as measured by the R^2 , of magnitudes similar to those we find in the data.³⁹ The estimation coefficients on the artificial data tend to be higher than in the historical data.⁴⁰ Unsurprisingly, in the light of the foregoing discussion of the properties of our model, there is less predictability at higher initial wealth shares of the less risk averse agents. This is because in these simulations we miss part of the range of wealth distributions for which effective heterogeneity is largest. In order to generate higher R^2 s also for higher starting values we would have to increase the spread between the two types' risk aversion coefficients further.

How 'big' is heterogeneity in our calibration anyway? The degree of leverage of the less risk averse agents may provide an intuitive measure. We find that on average the L-types hold stocks amounting to 128% of their wealth in the simulation for $w_{L,0} = 0.1$ and less in the other two simulations. The H-types' portfolio share of stocks is on average 41% in the simulation for $w_{L,0} = 0.1$. This difference may seem rather big, but one has to keep in mind that in this economy the stock is the only asset in positive net supply. Data presented in Vissing-Jorgensen (2002) show that even the 44% of stockholders in the U.S. hold only about half their financial wealth in stocks, with a standard deviation of 30%. This number translates into a coefficient of variation of the portfolio share of about 0.6, almost equal to the one in our model for the simulation for $w_{L,0} = 0.1$ and bigger than the ones for the other simulations. Taking into account that households owning stock likely own non-financial wealth as well and/or adding the households who do not participate in the stock market, the coefficient of variation calculated from the data would actually be higher than in our

³⁹In line with intuition more heterogeneity leads to more predictability because portfolios become more extreme and the wealth effect thus becomes stronger.

⁴⁰This may be because the volatility of the price dividend ratio in the simulated data relative to the real data is even lower than the relative volatility of excess returns in simulated and real data. All coefficients are significantly different from zero, both those estimated from artificial and from real data.

calibrated model. It therefore seems fair to say that the amount of heterogeneity we impose in our calibration is not excessive.

Another way of looking at heterogeneity in our model is by comparing the volatilities and correlations of individual and aggregate consumption. Table 4 contains the corresponding information. In our benchmark simulation the annual volatility of consumption growth of the less risk averse agents is about twice as big as that of aggregate consumption while the consumption growth volatility of the more risk averse agents is only little more than a third of its aggregate value. Thus quite a bit of consumption insurance is taking place between agents, and from this perspective heterogeneity does seem sizable. For completeness we also report the correlations of aggregate and individual consumption growth rates. They are positive but far from one, in particular where c_L is involved.

Table 5 presents means and standard deviations as well as predictive regressions for our alternative calibration using dividend data. Unfortunately the results are only partially comparable to those from the consumption calibration because we could not solve the model for comparable degrees of heterogeneity. For the dividend calibration we hence have to use a ratio of risk aversion coefficients of $\frac{\gamma_H}{\gamma_L} = \frac{5.683}{2.335} \approx 2.43$, compared to ≈ 2.71 in our benchmark calibration to consumption data. Nevertheless, this alternative calibration is instructive to analyze. Unsurprisingly, we achieve much less predictability than in our benchmark calibration. Nevertheless, the R^2 rises noticeably with horizon and reaches 5% at a 10 year horizon, which is still more than what Chan and Kogan (2002) achieve in their heterogeneous agent model. The volatilities of excess returns and price-dividend ratio are much higher now than in our benchmark calibration. The big increase in excess return volatility can clearly be traced back to the much higher volatility of dividend growth of 12.8%. Again, the positive correlation of price-dividend ratio and dividend growth add some more to the volatility of excess returns.

Nevertheless, the volatility of excess returns and price-dividend ratio is still much lower than in the data. It seems that in this model as it stands it is very hard

to generate sufficient volatility. One reason is that the price-dividend ratio here is bounded above and below by its respective values in the two polar representative agent models. At 18.86 and 23.97 in our benchmark calibration, these bounds are clearly too tight compared to the data. We could extend them somewhat by choosing a higher value for the elasticity of intertemporal substitution, but this would not help very much.⁴¹

This lack of volatility is a bit of a concern because it could bias upward our results regarding predictability. For in a way strong predictability and low volatility of price-dividend ratio and excess returns are two sides of the same coin: Strong predictability means that the variability of expected excess returns, i.e. of the equity premium, is high relative to that of ex post excess returns. Ex post excess returns vary for three reasons. The first is variation in expected excess returns as a consequence of wealth shifts between agents. Secondly, dividend shocks directly affect realized excess returns. And so do, finally, movements in the price-dividend ratio. Thus if the volatility of excess returns is too low because the price-dividend ratio is too stable, excess returns may display too little ex post volatility relative to their ex ante volatility, implying too much predictability. An answer to this concern is that the level of heterogeneity in our model is sufficiently moderate as to allow further increases. If we could increase heterogeneity so much as to generate enough volatility we would likely obtain far too much predictability.⁴² In the light of the difficulties of previous contributions to generate sufficient predictability this might actually be a welcome rather than a problematic result. Also, realistic extensions such as modeling equity as a leveraged claim on consumption would help to reduce predictability again by raising volatility for a given amount of heterogeneity.

The upshot of the foregoing discussion is that strong predictability of excess returns is indeed a feature of our model. This is true not only in the sense of generally sizable R^2 s but also in terms of significant rises in the R^2 with horizon. To understand

⁴¹On the downside, the drift in the wealth distribution would become much stronger, which we regard as undesirable.

⁴²Numerical limitations prevent us from doing so.

this latter property it is best to refer to the discussion of predictability in Cochrane (2005), chapter 20. The point made there is that for the R^2 s of predictive regressions to rise strongly with horizon the forecasting variable, i.e. the price-dividend ratio for our purposes, has to be very persistent. This is clearly the case in our model. Table 2 reveals that the first-order autocorrelation of the price-dividend ratio in our simulations is of the same order of magnitude (ca. 0.8) as in the data. This persistence of the price-dividend ratio can in turn be traced back to the highly persistent effect of dividend shocks on the wealth distribution. Recall from section 2.3.3 that if it were not for the drift in the wealth distribution persistence would actually be complete.

The overall magnitude of the R^2 s, on the other hand, has to do with our choice of Epstein-Zin utility. This choice allows us to impose heterogeneity only on the coefficient of risk aversion while leaving the elasticity of intertemporal substitution equal across agents. We can thus make the equity premium rather volatile while leaving the real rate fairly constant. It moves only to the extent that changing average risk aversion changes the strength of the precautionary savings motive. As a result in our model, unlike in Chan and Kogan (2002), return predictability is actually excess return predictability and not gross return predictability driven by predictable movements in the interest rate. In our benchmark calibration the volatility of the risk premium is more than 50% bigger than that of the risk free rate, while in Chan and Kogan (2002) it is almost exactly the other way round.

2.5 Conclusion

We have analyzed a model of heterogeneous Epstein-Zin agents who differ exclusively in their risk aversion. Our major finding is that in this world even at moderate degrees of heterogeneity the price-dividend ratio is a significant (negative) predictor of future excess returns. Regressions of excess returns, cumulated over different horizons, on the price-dividend ratio display the empirically observed pattern of R^2 s rising strongly with horizon. The magnitudes of the R^2 we find are close to their empirical values.

Heterogeneity in our model also creates some excess volatility of stock prices and excess returns, however, the magnitudes of these volatilities are still far below their empirical counterparts for our calibration. To raise them significantly, we would likely have to raise heterogeneity to absurd degrees. In order to improve this aspect of the model it would be interesting to introduce leveraged equity. I.e. bonds would be in positive net supply, and stocks would be a claim on consumption net of payments to bond holders. This would be a way to make dividends more volatile than consumption while retaining the general equilibrium character of the model, which is necessary with heterogeneous agents. We conjecture that this extension would be an alley towards achieving realistic amounts of volatility and predictability at the same time.

2.6 Appendix

2.6.1 Choice of state space and computational strategy

For all the results presented in section 2.4.2 we solve the model by approximating four functions on a grid of w_L , the wealth share of the less risk averse agent type. The functions describe c_L , the consumption of the L-type as a share of dividends, ϕ_L , the L-type's portfolio share of stocks, PD , the price-dividend ratio, and R , the gross risk free rate. We can use the wealth distribution as the single state variable because dividend growth is i.i.d.

For each wealth distribution we find next period's wealth distributions on a grid of dividend growth rates as a fixed point of

$$w'_L [PD(w'_L) + 1] \exp(\Delta d') = (w_L [PD(w_L) + 1] - c_L(w_L)) \left(\phi_L(w_L) \frac{PD(w'_L) + 1}{PD(w_L)} \exp(\Delta d') + (1 - \phi_L(w_L)) R(w_L) \right)$$

where we use our approximated functions. The resulting set of $w'_L(w_L, \exp(\Delta d'))$ allows to find the corresponding next period values of the approximated functions and ultimately to state the Euler equations for stocks and bonds for each agent. We iterate on the approximated functions to satisfy the Euler conditions using Broyden's algorithm.

For the grid of dividend growth rates we use 30 grid points, which are weighted to approximate a log-normal distribution. For the grid of wealth shares 25 Chebychev nodes or less are typically sufficient to achieve a precision of 10^{-5} or better off the grid.

2.6.2 Complete versus incomplete markets

In the model on which the results of our paper are based markets are incomplete because there are only two assets while dividend growth can have many realizations.

Also, with heterogeneous agents we are not aware of any spanning properties for this asset structure. Nevertheless, we conjecture that our results do not depend on this market incompleteness. Several arguments suggest so:

1. At a theoretical level, with one source of uncertainty and trade in a stock and a riskless bond markets are incomplete only due to discrete trading intervals. I.e. as we shrink the period length we get closer and closer to complete markets. Numerically, we have made use of this observation and solved the model for ever shorter frequencies in order to check to what extent the solution changes. It turns out that changes are hardly recognizable.
2. We have also programmed and solved the social planner's problem. (See below for the set-up and the computational approach.) It is computationally more involved, which makes it less practical for experimenting, and at the point of writing the program is not fully stable. However, for the calibrations tried results are hardly distinguishable from those for the incomplete markets version.
3. We have analyzed the simplified model of section 2.3.3 for the case in which the dividend shock can take only two values such that markets are complete. The results of lemma 4 uphold and can in fact be strengthened.

Social planner's problem

For completeness we explain how we solve the complete markets version of our model. The problem we solve is

$$\begin{aligned} \max_{\{C_{L,t}, C_{H,t}\}} \quad & \omega \log U_{L,1} + (1 - \omega) \log U_{H,1} \\ & C_{L,t} + C_{H,t} = D_t \forall t \geq 1 \end{aligned} \tag{2.17}$$

where $U_{i,t} = \left[(1-\delta)C_i^{1-1/\psi} + \delta[E(U_{i,t+1}^{1-\gamma_i})]^{\frac{1-1/\psi}{1-\gamma_i}} \right]^{\frac{\psi}{1-\psi}}$. We use the log-transformation of U in the planner's problem because in this formulation the welfare weight ω can be interpreted as the initial wealth share of the L-types, as will be shown below. This

property is useful for comparing the complete and incomplete markets versions of our model. Since the log is a positive monotone transformation $\log U$ represents the same preferences as U .

The first order conditions of this problem are equalization of stochastic discount factors (definition in equation (2.10)),

$$m_{L,t+1} = m_{H,t+1}, \quad (2.18)$$

for all periods and

$$\omega U_{L,1}^{1/\psi-1} C_{L,1}^{1/\psi} = (1 - \omega) U_{H,1}^{1/\psi-1} C_{H,1}^{1/\psi} \quad (2.19)$$

for period one.

The problem is thus not fully recursive, since we have an extra condition for the initial period. We handle this problem as follows: First we approximate c_L , the consumption share of the L-type, as well as $v_i \equiv V_i/D$, the value of each agent type, normalized by dividends, on a grid of dividend growth and last period's distribution of consumption $c_{L,-1}$. In this we make use of the stochastic discount factor condition (2.18) as well as the two recursive utility functions. Next we use equation 2.19 to find the initial consumption distribution that corresponds to the initial wealth distribution characterized by ω .

The interpretation of the welfare weight ω as the initial wealth share of the L-types is possible due to the linear homogeneity of U . This property implies $U_{i,1}^{1-1/\psi} C_{i,1}^{-1/\psi} = (1 - \delta) \frac{U_{i,1}}{\partial U_{i,1} / \partial C_{i,1}} = (1 - \delta) W_{i,1}$.⁴³ Dividing the corresponding conditions for each type by one another and noticing that $W_{L,1}/W_{H,1} = w_{L,1}/w_{H,1}$ we arrive at equation (2.19) with $w_{L,1} = \omega$.

⁴³See Cochrane (2006) for details on the properties of Epstein-Zin utility.

Chapter 3

Dynamic Effects of Unemployment Insurance Reform

3.1 Introduction

Recently considerable attention has been devoted to the analysis of the influence of unemployment insurance on labor market outcomes and welfare. Several main channels have been identified: Simple matching models like the one used in Pissarides (1990) show that, by increasing workers' outside option in bargaining, unemployment benefits raise wages and thus reduce vacancy creation, job finding rates and ultimately employment. The welfare consequences of this effect are ambiguous unless one imposes parameter restrictions.¹ A clearly negative consequence of unemployment insurance is above all the moral hazard that it induces among both unemployed and employed workers by distorting their search and work effort incentives respectively.² Furthermore, to the extent that unemployment insurance lowers employment it will increase welfare losses from distortive taxation because tax rates have to increase to keep revenue constant and the benefit itself has to be financed as well. On the positive side, Marimon and Zilibotti (1999), for example, have pointed to improvements in match quality and hence productivity resulting from the role of unemployment benefits as a subsidy to search. The most obvious benefit of unemployment insurance, however, is its insurance function for risk-averse individuals in an environment of incomplete asset markets. In this role it may not only increase utility but even enhance productive efficiency as Acemoglu and Shimer (1999) show. This is because, absent insurance, risk-aversion would lead to inefficiently high employment. On the other hand, it is well documented that even under incomplete markets individuals can self-insure quite well against temporary income shocks as long as they have access to a safe asset.³ Thus, given its adverse effects, rather low levels of unemployment insurance are likely to be optimal in steady-state.

Nevertheless, currently most developed countries exhibit fairly high levels of un-

¹E.g. the Hosios (1990)-condition.

²Contributions emphasizing search distortions are Shavell and Weiss (1979), Hopenhayn and Nicolini (1997), and more generally the optimal unemployment insurance literature. Moral hazard among employed workers has for example been discussed by Wang and Williamson (1996).

³See for example Krusell and Smith (1998) and the references cited therein.

employment benefits. - After tax replacement rates around 60% are by no means uncommon. And despite the above theoretical considerations it is not clear *a priori* whether reducing them would be beneficial, for self-insurance crucially depends on having appropriate asset levels. But accumulating assets takes time and has a cost of foregone consumption. Also temporary lack of insurance in the early phase of a reform will impose a welfare cost. Hence, looking at the transition from the initial conditions to the new steady-state is essential before making any statement regarding desirable levels of unemployment insurance.

Such a dynamic analysis of the consequences of reducing unemployment benefits is the main contribution of the present paper. To this end, we embed the standard Mortensen-Pissarides job matching model in an Aiyagari (1994)-type incomplete markets setting. As a by-product, insights are gained on the out-of-steady-state behavior under incomplete markets of this main workhorse of the recent macro-labor literature. In the model individuals are risk-averse and face idiosyncratic income uncertainty due to stochastic transitions between the states of employment and unemployment. In addition to unemployment insurance they have the possibility to self-insure by accumulating an asset that yields a safe exogenously fixed return, but they cannot borrow. Job loss occurs exogenously while the reemployment probability is endogenously determined by the vacancy-unemployment ratio and the matching function. Firms' hiring probability is correspondingly endogenous as well. A zero-profit condition determines vacancy creation in the presence of costly vacancy creation. Unions and employers, both without strategic motives, bargain over the wage, which consequently is unique despite the fact that heterogeneous individual assets imply heterogeneous outside options for the agents. For the sake of tractability, labor supply is inelastic and there is no search decision. Nevertheless, there are potentially strong effects of benefits on job finding rates and employment which work through the Nash bargaining and firms' vacancy creation. Appropriate calibration hence yields the right reduced form effects without the extra complications of further decision variables. The solutions for both the steady-state and the transition path are obtained numer-

ically for a calibration to Germany in the mid-1990s. The policy reforms considered are one-off changes in the level of unemployment benefits with and without previous announcement as well as optimal reform paths. We compare the effects for two variants of the model: First we maintain the standard assumption of firms facing a cost for each period that they open a vacancy. Then we alternatively assume that there is a fixed cost of hiring a worker independent of how long it takes to find him.

We find that, even though steady-state comparisons suggest significant welfare gains from reducing unemployment benefits, the dynamic analysis reveals important transition costs. For the standard vacancy cost specification even a one percentage point reduction of the replacement rate (without announcement) harms the unemployed workers. However, average worker welfare improves for moderate reductions. The transition costs are identified to stem on the one hand from the need to increase savings (and hence temporarily reduce consumption) to improve self-insurance and on the other hand from the drop in utility for those who become unemployed before they can adjust their asset holdings to the reduced unemployment income.

The welfare gains for workers can be increased greatly when the reform is announced some time in advance or if it is phased in gradually. The main reason, however, is not a reduction in transition costs that could be achieved this way. Rather announcement or phasing-in allow workers to extract gains from the reform that otherwise accrue to firms. These gains arise due to the upward jump in vacancies following an unexpected reform, which makes recruiting so much more costly that firms with filled jobs earn a large windfall. Bargaining allows workers to extract part of this windfall during an announcement period or during the phasing-in. Vacancies then no longer jump but increase steeply.

The results for our alternative assumption of a fixed hiring cost differ substantially. Despite the costs of transition, even big, unannounced reforms turn out favorably. But advance announcement does not further improve the welfare effects. Since with constant hiring costs there is no windfall to be appropriated by the workers, an announcement period only permits improvements in self-insurance before the reform

hits. However, these are quantitatively dominated by the losses from delaying the reform. In fact, an unannounced one-off change in the replacement rate is optimal when there is a fixed hiring cost.

The transition paths of unemployment and wages reflect the general equilibrium nature of the model. In the first periods following an unannounced reform gross wages drop sharply due to insufficient self-insurance, overshooting their new steady-state. This induces so much job creation that unemployment also overshoots with a lag before gradually converging to its new steady-state. Net wages, by contrast, at first drop but then increase above their old steady-state level because the increase in employment reduces the burden from unemployment insurance contributions and other taxes. With announcement dynamics differ in that wages first rise before the reform hits, which reflects the process of rent appropriation described.

The two papers that are most closely related to this work are Joseph and Weitzblum (2003) and Lentz (2003). To my knowledge they are the only contributions that take account of transition effects in the welfare analysis of unemployment insurance. Both, however, use partial equilibrium models in which the wage is exogenous and does not react to the policy change. Joseph and Weitzblum (2003) is a numerical analysis calibrated to the low-skilled segment of the French labor market. The authors find that even though steady-state comparisons suggest welfare gains from lowering unemployment insurance transition costs more than outweigh the gains. Lentz (2003) structurally estimates a search model with precautionary savings and variable search intensity. Using Danish data he finds that the search decision is only little distorted by unemployment insurance. Consequently optimal replacement rates turn out rather high, that is between 43% and 82%.

Static analyses of unemployment in general equilibrium search models with incomplete markets have been performed by Costain (1997) and Rebelo, Gomes and Greenwood (2003). In a life-cycle model with matching and search costs Costain (1997) finds mildly positive steady-state effects of unemployment insurance, which become significantly positive for higher coefficients of risk-aversion. Calibrating their

model to U.S. replacement rates and higher 'European' levels, Rebelo et al. (2003) find a negative impact of benefits on employment and welfare in a search model with endogenous labor supply.

The paper is structured as follows: Section 4.2 sets out the model and defines the concept of equilibrium used. The calibration is described in section 3.3. In section 3.4 the model's comparative statics for different benefit levels are analyzed. The transition dynamics and their welfare consequences are then investigated in section 3.5. Section 3.6 concludes.

3.2 The Model

In this section I set out the model I am using for my analysis. One of its building blocks is the Mortensen-Pissarides type matching framework with matching function and wage determination through Nash bargaining. The other important feature is market incompleteness in the sense that risk-averse, credit-constrained individuals, who face idiosyncratic income uncertainty due to the risk of job loss, can self-insure only via a safe asset a . There is no aggregate uncertainty since the focus is on unemployment insurance which can clearly cover idiosyncratic risk only. However, the labor market reform considered consists in an unanticipated shock to the level of unemployment benefits. Unemployment benefits are financed through a tax levied on the employed. The firm side is kept as simple as possible with one-worker firms producing a fixed output when they have a worker. Wages and vacancy-unemployment ratio (henceforth sometimes called market tightness) are endogenously determined in general equilibrium. The interest rate r is exogenously fixed at a level below the workers' discount rate β , that is to say the economy can be thought of as small and open.⁴

⁴The reason for not endogenizing the interest rate is that workers in reality only own a small fraction of the entire productive capital. With an endogenous interest rate this would either imply a very unrealistic capital-labor ratio or unrealistically high accumulation among workers, which would render an analysis of unemployment insurance useless.

3.2.1 The Individual's Problem

The economy is populated by a mass one of risk-averse individuals who are characterized by their asset holdings a_t and by their employment status $s_t \in \{u, e\}$. Their problem consists in optimally choosing consumption c_t and savings a_{t+1} in every period subject to their budget and borrowing constraints and taking into account the Markov transition probabilities $\pi_t(s_{t+1}|s_t)$ between unemployment and employment. The utility function satisfies standard conditions. Labor supply is fixed and there are no search costs. Hence the individual's value is

$$V_t(a_t, s_t) = \max_{c_t, a_{t+1}} \{u(c_t) + \beta(\pi_t(u|s_t)V_{t+1}(a_{t+1}, u) + \pi_t(e|s_t)V_{t+1}(a_{t+1}, e))\} \quad (3.1)$$

$$\begin{aligned} \text{s.t.} \quad c_t + a_{t+1} &\leq i_t(s_t) + (1+r)a_t \\ a_{t+1} &\geq 0 \end{aligned}$$

where $i_t(s_t)$ is his state-dependent non-asset income. For an employed worker this income equals the wage w_t minus taxes. For unemployed workers it consists in unemployment benefits b_t , which are a fraction $\rho_t \in [0, 1]$ of the net wage. This net replacement rate does not depend on the completed length of an individual unemployment spell but may vary over time during the implementation of a reform. Taxes serve both to finance unemployment benefits and to cover the state's other expenses which are assumed to be a fixed exogenous amount G . Hence they equal $\frac{u_t b_t + G}{1 - u_t}$ per employed worker, where u_t is the unemployment rate.⁵ This leaves after some manipulations:

$$i_t(s_t) = \begin{cases} \rho_t \frac{(1-u_t)w_t - G}{1-u_t + \rho_t u_t} & s_t = u \\ \frac{(1-u_t)w_t - G}{1-u_t + \rho_t u_t} & s_t = e \end{cases}$$

⁵This general purpose tax is introduced for calibration purposes only in order to achieve a realistic range for the wedge between gross and net wages. It can be thought of as the sum of a payroll tax and social security contributions aside from unemployment insurance. The individual's tax burden varies inversely with employment such as to capture the general equilibrium effect of employment on taxes.

The value function of the individual is time-indexed because out-of-steady-state wages, taxes, and transition probabilities are not constant. The individual is assumed to know the entire future path of these variables, which are determined in general equilibrium as set out below.

The Markov transition probabilities between unemployment and employment are determined as

$$\begin{bmatrix} \pi_t(u|u) & \pi_t(e|u) \\ \pi_t(u|e) & \pi_t(e|e) \end{bmatrix} = \begin{bmatrix} 1 - \theta_t q(\theta_t) & \theta_t q(\theta_t) \\ \lambda & 1 - \lambda \end{bmatrix} \quad (3.2)$$

Here λ is the exogenous rate of job destruction while $\theta_t q(\theta_t)$ is the job finding rate for unemployed workers. θ_t , the ratio of vacancies to unemployed, is the argument of the matching function $q(\cdot)$ (expressed as matches per vacancy), which is decreasing with an elasticity of less than one in absolute value. θ_t is exogenous to the worker and is determined in general equilibrium.

Note that, in the description of the individual's problem we have taken for granted that jobs are always accepted. This is in fact a safe thing to do because, with unemployment income being no greater than income from working ($\rho_t \leq 1$), individuals will accept any job they find provided $1 - \lambda \geq \theta_t q(\theta_t)$, i.e if their chances of having an employment opportunity next period are at least as big if they accept the offer as if they reject it. This will always hold if one chooses the time period short enough.

3.2.2 Firms

Firms can be in one of two states: They can have a worker and produce, or they can have a vacancy. Like the workers they know the future path of wages and matching probabilities. A filled job has value

$$J_t = p - w_t + \frac{1}{1+r} [\lambda O_{t+1} + (1-\lambda)J_{t+1}] \quad (3.3)$$

where p is the output of a firm that has a filled job. Since with probability λ the match is destroyed, the firm's value next period changes to O_{t+1} , the value of a vacancy, with probability λ . A firm that opens a vacancy has value

$$O_t = -\kappa + \frac{1}{1+r} [q(\theta_t)J_{t+1} + (1-q(\theta_t))O_{t+1}] \quad (3.4)$$

with κ being the cost of opening a vacancy for one period.⁶ Free entry implies that $O_t = 0$. Thus profits on average just cover hiring costs. This implies that *ex ante*, i.e. in the absence of unexpected parameter changes, the value of a filled job is given by

$$J_{t+1} = \frac{1+r}{q(\theta_t)}\kappa. \quad (3.5)$$

Combining equations (3.5) and (3.3) we can relate wages and market tightness according to

$$w_t = p + \frac{1-\lambda}{q(\theta_t)}\kappa - \frac{1+r}{q(\theta_{t-1})}\kappa \quad (3.6)$$

which again only holds absent policy shocks. In case policy shocks lead *ex post* to profits or losses, these are attributed to the rest of the world, which is assumed to be the residual claimant.

3.2.3 Determination of Wages and Matching Probabilities

Nash bargaining between employers' associations/firms and unions determines the wage in a given period conditional on current and future market tightness and taxes and future wages. That is to say, agents on both sides do not behave strategically such as to influence aggregate or future variables.⁷ The union's objective is the

⁶The alternative specification of the hiring cost referred to in the introduction will be introduced in section 3.5.2.

⁷Clearly, this assumption is strong in the context of centralized bargaining. But it simplifies the solution of the model greatly. Individual bargaining would make it unnecessary, however at the cost of wage differentials that are due not to productivity differences or job characteristics but asset holdings.

value of the median employed worker, that is of the worker whose assets correspond to the median of the asset distribution for employed workers. This form of wage determination achieves that there is a unique wage for all employees despite their heterogeneous outside options (due to heterogeneous wealth).⁸

Precisely, denoting by a_t^m the asset level of the median worker in period t , bargaining between firms and unions solves

$$\max_{w_t} (V_t(a_t^m, e) - V_t(a_t^m, u))^\sigma J_t^{(1-\sigma)} \quad (3.7)$$

The solution to this problem is

$$V_t(a_t^m, e) - V_t(a_t^m, u) = \frac{\sigma}{1-\sigma} V'_t(a_t^m, e) J_t. \quad (3.8)$$

The determination of the wage also pins down the number of vacancies created and hence market tightness and matching probabilities, even though with a lead of one period of vacancies. The mechanism is through free entry: Given future wages and matching probabilities the value of a firm one period ahead, J_{t+1} , is known. Entry and vacancy creation has to occur so long as to drive this period's value of a job opening to zero. This link from wage setting to matching probability (and consequently job finding rate, average duration of unemployment, unemployment rate, etc.) is crucial to keep in mind for the analysis of unemployment insurance below. For the level of unemployment benefits directly influences the bargaining strength of workers and thus wages and by the mechanism just explained employment dynamics. Thus, even though we do not model labor supply and search costs, which are usually held to provide the channels for effects of unemployment insurance on employment dynamics, the observed correlation between benefit levels and labor market variables

⁸This wage is nevertheless acceptable to all workers regardless of their asset position since jobs are never turned down in this setting (Cf. section 3.2.1).

⁹The presence of the $V'_t(a_t^m, e)$ -term on the right hand side is caused by the fact that workers have strictly concave utility unlike in most matching models. The derivative of the value function with respect to wages is therefore generally not equal one. Intuitively, an extra unit of wages translates into one unit loss of surplus for the firm but into $V'_t(a_t^m, e)$ units gain in surplus for the worker.

can be generated.

3.2.4 Competitive Equilibrium

To close the model the behavior of the aggregate state variables, that is unemployment and the asset distribution, has to be determined. The law of motion for the unemployment rate is

$$u_{t+1} = \lambda(1 - u_t) + (1 - \theta_t q(\theta_t))u_t. \quad (3.9)$$

To describe the evolution of the asset distributions for unemployed and employed workers it is necessary to first introduce some more notation. As explained in section 3.2.1, the individual's asset choice depends on his individual states plus the paths of wages, taxes, and tightness. Write it hence as $a_{t+1} = F(a_t, s_t; \Theta^t, \omega^t, u_t)$ where Θ^t and ω^t are the paths of tightness and wages, i.e. $\Theta^t = \{\theta_t, \theta_{t+1}, \dots\}$ and $\omega^t = \{w_t, w_{t+1}, \dots\}$. Today's unemployment u_t features as an argument in the policy function because the path for taxes is determined by the path for tightness and equation (3.9) together with u_t as an initial condition. The law of motion for the distribution functions of unemployed and employed workers' assets, $G_t^u(\tilde{a})$ and $G_t^e(\tilde{a})$, can now be described as

$$\begin{bmatrix} G_{t+1}^u(a_{t+1}) \\ G_{t+1}^e(a_{t+1}) \end{bmatrix} = \begin{bmatrix} 1 - \theta_t q(\theta_t) & \lambda \\ \theta_t q(\theta_t) & 1 - \lambda \end{bmatrix} \begin{bmatrix} G_t^u(\bar{a}_t^u(a_{t+1})) \\ G_t^e(\bar{a}_t^e(a_{t+1})) \end{bmatrix} \quad (3.10)$$

where $\bar{a}_t^{st}(a_{t+1}) = \max\{a_t | F(a_t, s_t; \Omega^t, \omega^t, u_t) = a_{t+1}\}$.¹⁰ Note that this law of motion is based on the agents' optimizing choices.

For convenience summarize the aggregate states as $\Gamma_t = (G_t^u(\tilde{a}), G_t^e(\tilde{a}), u_t)$ and denote their joint law of motion as $\Gamma_{t+1} = H_t(\Gamma_t)$. We are now set to define a (potentially non-steady-state) equilibrium for this economy:

¹⁰This formulation relies on continuity and monotonicity of the policy function F in a_t , which is warranted by standard nature of the household problem in this respect.

Definition 5 *A competitive equilibrium for the economy considered is a wage path ω , a tightness path Θ , a policy function F , and a law of motion H such that*

- (i) F solves the individual's problem,*
- (ii) ω solves the Nash bargaining,*
- (iii) H is generated by F and Θ , and*
- (iv) vacancies have zero value.*

In particular, a steady-state equilibrium is defined as follows:

Definition 6 *A competitive steady-state equilibrium for the economy considered is a competitive equilibrium in which θ and w are time-invariant and Γ is a fixed point of H .*

3.3 Calibration

The solution of the model is entirely numerical. In a first step the steady-state is solved for. Out of steady-state a time path between two steady-states with different levels of unemployment benefits is determined. The algorithms used to find the steady-state and the transition path are contained in appendix 3.7.2.

I make standard functional choices: Utility is taken to be of the CRRA type with coefficient of relative risk-aversion γ set to 2. This is within the acceptable range according to Mehra and Prescott (1985*b*). The matching function is Cobb-Douglas. Dividing it by vacancies and calling $\theta = \frac{v}{u}$, the ratio of vacancies to unemployment, we have $q(\theta) = \frac{m(u,v)}{v} = \chi \cdot \theta^\eta$. η is the elasticity of the matching function with respect to vacancies while χ is a scaling parameter. η is chosen to be -0.5 which is in the middle of the commonly used range of -0.4 to -0.6.¹¹ The weight of the worker in bargaining is 0.5. This is the value that has traditionally been assumed in Nash bargaining.¹² Also we thus have $\sigma = |\eta|$ as in most of the literature. The Hosios

¹¹See Petrongolo and Pissarides (2001).

¹²e.g. in Pissarides (1990).

(1990)-condition regarding absence of search externalities is nevertheless not readily applicable due to the concavity of the workers' utility function which breaks the link between σ and the share of the match surplus going to the workers. Output per match, p , is normalized to one. The time period is set to one month.

All remaining parameters are matched to observations on Germany for the mid-1990s. The parameters governing labor market flows can be inferred from data on unemployment, vacancies, and unemployment duration using the law of motion for unemployment. An unemployment rate of 8.2% (OECD standardized unemployment rate), vacancy-unemployment ratio of 0.10 (calculated from the vacancy data of Nickell and Nunziata (2001)), and average unemployment duration of 12.4 months (Machin and Manning (1999)) then imply a job destruction rate $\lambda = 0.72\%$ and matching efficiency $\chi = 0.254$.

The net replacement rate, that is the ratio of unemployment benefits to net wages ρ , is set to 60%. Obviously, as in most economies in Germany there exists in fact a multitude of replacement rates depending on previous wage, personal circumstances, employment history, and completed length of the unemployment spell. To summarize them into this one rate I construct a weighted average from OECD data (Martin (1996)) reporting net replacement rates for three different classes of workers and three unemployment durations. The weights are obtained from data in the Report on Poverty and Wealth of the Federal Government of Germany (Bundesregierung (2001)).¹³ It may be worth noting that a constant replacement ratio independent of the length of the unemployment spell is in fact not a bad approximation for many workers in Germany.

Finding one single rate of taxation is fraught with the same problems as finding a summary replacement rate. The rate According to the Federal Employment Agency (Bundesanstalt für Arbeit), total social security contributions (i.e employer and worker share together) amount to about 33% of labor costs, while wage taxes

¹³Martin (1996) also reports an 'overall average' of 54%. However, this measure is the *unweighted* average of the nine categories, in which for example long-term unemployed with spouse in work, who do not receive any benefits, are given far too much weight.

net of family benefits etc. amount to 7% for an average employee (German Institute for Economic Research (DIW), Bedau and Teichmann (1995)).¹⁴ We thus set taxes including unemployment insurance contributions (i.e. $\frac{u_t b_t + G}{1 - u_t}$) to 40% of gross wages. The magnitude of this tax wedge has direct consequences for the elasticity of the unemployment rate with respect to unemployment benefits because it determines the strength of the general equilibrium effect of benefits on unemployment: As unemployment increases with the level of benefits, the tax burden on each worker increases not only due to rising contributions to unemployment insurance but also due to the fact that G has to be financed by fewer employees. This tends to raise gross wages further and reduce employment even more. Hence, the resulting elasticity of unemployment with respect to the benefit level is a measure of success for the calibration of the tax burden. We compute this elasticity to be 0.75. This is reasonably close to the estimates of approximately one reviewed in Costain and Reiter (2003) given the absence of an endogenous search decision in our model. By assuming a higher tax burden the elasticity of unemployment to benefits could be further raised. But we choose to stick to the data in order to avoid the danger of overstating the welfare gains from reducing unemployment benefits because a fall in unemployment due to increased costly search is less welfare enhancing than one that is due to lower taxation.

The remaining parameters, the vacancy creation cost κ and the discount factor β , are calibrated such that the aggregate state variables, unemployment and asset distribution, match the data. Given that income variation in the model results from the risk of unemployment only it is clearly unrealistic to aim for a realistic asset distribution in all dimensions. Since the model is about dependent workers I focus on this segment of the population and in particular on the share of people with little or no assets. This seems reasonable since it is these people who are most affected by the level of unemployment benefits. Also by assets I understand liquid assets, i.e. those that can readily be used to smooth consumption in case of job loss. Of these 6.5% of

¹⁴More precisely: On so called 'gross wages' employers and employees each pay about 20% of social security contributions, and the average employee pays 8% in wage tax net of family benefits, subsistence allowance, etc. So in total we have a burden of 48/120 or 40% of labor costs.

German worker and unemployed households did not have any in 1998 (Bedau (1999), Mnnich (2001)). Since it is likely that even these households have some money on a current account, I calibrate the difference between discount rate and interest rate such as to achieve a share of 6.5% with less than half a monthly net wage in savings. This requires an annual discount rate of 5.5% while the real interest rate is set at 2% per year. This value of the interest rate is chosen despite the fact that the real return on (rather safe) public debt has tended to be higher. The reason is that people with little financial wealth tend to hold it in assets with extremely low yields such as savings accounts.

3.4 Steady-State Analysis

Before turning to the numerical results let us briefly recall the main economic forces at work in Mortensen-Pissarides type models. First of all, unemployment benefits have a positive effect on the (gross) wage because they improve the workers' position in the Nash bargaining. Secondly, higher wages go along with lower market tightness, which in steady-state translates into higher unemployment. Formally this relationship can be seen from equation (3.6) evaluated in steady-state. The economic logic is that at higher wages the flow income of firms is lower and hence, for given interest and job destruction rates, recruitment costs must be lower in order for them to break even in present value terms. But recruitment costs are lower only if market tightness is lower because vacancies are then matched to workers more rapidly. Thus, in choosing their preferred replacement rate workers face a trade-off between higher wage income when employed and the fraction of time they spend unemployed on average. In the presence of discounting, unemployed workers will tend to prefer a higher replacement rate than employed workers, whose potential unemployment spells are more distant. Risk-aversion introduces the aspect of consumption-smoothing and therefore implies higher optimal replacement rates for all workers.

Note that this analysis is not at odds with the constrained efficiency result of

Hosios (1990).¹⁵ First of all, the Hosios-result is applicable with risk neutral agents only. Secondly, in the above as well as in the analysis in the next sections we are concerned with the wellbeing of workers in their two employment states only and not with the value of the entire economy. The difference is the value of firms, which reacts to changes in the return to past investment in vacancies. The reason for this choice of welfare criterion is that not only in the model but arguably also in reality workers' preferences over different degrees of unemployment insurance do not depend on these firm value effects since they tend not to be the owners. Hence, with regard to the political economy aspects of unemployment insurance reform, it makes sense to focus on the value of workers under different policies only.

Let us now turn to the numerical results for the model with precautionary savings. Figure 8 summarizes the comparative static welfare analysis. The solid lines depict utility as a function of the replacement rate for different asset levels. In line with intuition both unemployed and employed workers have strictly positive optimal replacement rates for each asset level, welfare depends positively on asset holdings, and richer individuals prefer lower replacement rates because they are better self-insured. The dashed lines correspond to the average utility across the steady-state asset distribution for the respective worker group. Utility is measured in units of equivalent certain consumption per period. Even for the unemployed the optimal level of unemployment insurance is lower than the current one of 60% of net wages. For the employed workers it would be best to be in a steady-state with no unemployment insurance whatsoever. Since they are the great majority the same is true for total utility aggregating over all workers. This result is perhaps surprising given the above reasoning. The reason is that, as benefits are withdrawn, the steady-state asset distributions shift towards higher asset levels so strongly that, given the positive relationship between assets and welfare, individuals are better off on average when

¹⁵The Hosios-condition says that whenever the matching elasticity is equal to the workers' share in bargaining the decentralized equilibrium (without tax financed benefits) is the same as the social planner's solution, which suggests that any distortion caused by unemployment benefits would be detrimental to welfare.

benefits are lower. That is to say, the general equilibrium effect through the asset distribution dominates all more direct effects via the bargaining or risk-aversion. This result fits in well with the literature on self-insurance, which generally documents that a single safe asset provides a lot of insulation against temporary shocks (see for example Krusell and Smith (1998) and the references cited therein).

Our findings are robust to varying the calibration with regard to vacancy rate, risk aversion, interest rate, and discount rate. Changes in the assumed vacancy rate, which is likely to be measured rather imprecisely, are fully offset by changes in the resulting calibrated value of vacancy costs. Higher or lower risk aversion do not change the steady-state results significantly either. Higher risk-aversion primarily translates into higher savings. Interest rate and discount rate interact to largely determine both the steady-state distribution of asset holdings and its responsiveness to unemployment benefits. The former is what we calibrated to. The latter has turned out rather high.¹⁶ It could be reduced by increasing the wedge between time discount and interest rate. In the *status quo* savings would then be even lower than under the benchmark calibration (which features low savings anyway since we calibrate to the share of people with no savings) while the comparative statics would not change much. For example for a rather extreme calibration with an annual interest rate of 1%, time discount rate of 8% annually, and risk aversion set to three¹⁷ the qualitative picture sketched above still upholds but welfare gains are smaller.

For the calibration shown here the welfare effect of the shifting asset distribution is reinforced by the fact that net wages actually increase as benefits decrease. This is due to the high government spending G whose burden per worker sinks with increasing employment at lower benefit levels. However, lower tax burdens would not change

¹⁶For a reduction of the net replacement rate from 60% to 50%, for example, the ratio of average wealth to monthly income increases from about 1.9 to about 3.3. This is not surprising given the wedge between time discount and interest rate of only about 3% per year. Carroll and Samwick (1997) document the same phenomenon and find they have to increase the annual discount rate to 13% (for an interest rate of 2%) in order to match the empirical responsiveness of asset holdings to income uncertainty.

¹⁷This choice of parameters yields again roughly a 6.5% fraction of the population with assets less than half a monthly wage. Similar results are obtained if instead of higher risk-aversion a Stone-Geary utility function with a minimum consumption of about one sixth of the wage is used.

the general picture. Higher G on the other hand would even strengthen the results.¹⁸

3.5 Dynamic Analysis

The results presented thus far are mere steady-state comparisons and could possibly be used to judge the desirability of policy reforms. It was stressed that they hinge to a large extent on the response of the steady state asset distribution to the change in benefit level. However, the asset distribution can adjust only slowly. In the short run individuals are stuck with the assets they have at the point of regime change, which means low consumption during unemployment and initially also reduced consumption during spells of employment in order to build up higher savings. This section will therefore discuss the properties of the transition between steady-states. As a benchmark we will first study simple unannounced one-off changes in the replacement rate. Then we will allow for previous announcement of the reform and show that the results depend to some extent on the specification of recruitment cost in the Mortensen-Pissarides world. Both under the standard assumptions and under the alternative assumption that there is a fixed cost of hiring a worker we will finally discuss the optimal path of reform.

3.5.1 The Benchmark - Transition Dynamics and Welfare Effects

Figure 9 graphs the paths of market tightness, unemployment, median assets, and net wages following an unannounced reduction of the replacement rate from 60% to 50% of net wages. As can be seen from the behavior of the jump variables tightness and wage, the bargaining position of the workers first worsens strongly. The state of unemployment becomes so unattractive that low gross wages are negotiated. Free

¹⁸This should be the relevant direction to look at given the absence of endogenous search effort. Compare the argument made in section 3.3

entry of firms correspondingly drives up vacancies. The reduction in the gross wage is so large that net wages at first fall as well despite the reduction in the tax burden that is caused by the reduction in unemployment benefits. As workers accumulate assets and thus improve the degree of self-insurance, their outside option unemployment gains in relative value again. The negotiated wage hence increases as is mirrored by the gradual fall in the vacancy-unemployment ratio. Net wages soon rise above their old steady-state level even though gross wages always stay lower because the increase in employment lowers both unemployment insurance contributions and other taxes. The path of unemployment is a consequence of the path of the vacancy-unemployment ratio. It initially increases so strongly after the policy shock that after 19 months unemployment has already fallen below its new steady-state level. After 39 months it reaches its minimum.

This general pattern of the transition paths after a reduction in unemployment benefits is the same as the one described irrespective of calibration and size of reform. The only sensitive aspect is whether net wages converge to a new steady-state that is higher or lower than the old one. This depends on the level of government spending G in the economy. The higher it is, the stronger the positive general equilibrium effect of higher employment on net wages via lower per capita taxes. But for realistic ranges of taxation the net wage always increases.¹⁹

Welfare effects, on the other hand, depend strongly on the particular reform experiment. Table 7 summarizes the welfare effects of lowering the replacement rate from 60% to various lower levels. The column titled 'static' reports the comparative-static effects discussed in section 3.4 while the column titled 'dynamic' gives the true, dynamic effects that take the transition period into account. Clearly, the transition period matters for welfare, and that more so the bigger the reform: Not even a re-

¹⁹The general equilibrium effect from G on net wages hinges on the assumption that government expenditure is constant. The alternative assumption would be constant revenue per employee. Even though the net wage would then tend to be lower at lower replacement rates the welfare evaluation would not change much. For the extra revenue from higher employment would then have to be added at the point of calculating welfare. However, given that there is no compelling reason why exogenous government spending should be higher at low unemployment levels, we find our way of handling the issue more reasonable.

duction of one percentage-point is Pareto-improving. The 10%-decrease that looked beneficial even for the unemployed in the static perspective turns out to reduce the unemployed's welfare in the dynamic perspective. And the reduction of the replacement rate to 30% that statically seems to give to the employed the highest gain of all three reforms considered in fact reduces their welfare by more than half a percent. The dynamic perspective also permits to discriminate winners and losers of the various reforms by asset levels. It turns out that employment status is a far more important predictor than asset level. Only when the overall welfare effect for a group (employed/unemployed) is small, asset holdings separate winners and losers. This is the case for the employed when the replacement rate is reduced to 40%. Then the 17.5% poorest employees, i.e. those with savings of less than about one and a half net wages, lose while the wealthier employees gain. An explanation for this pattern of gains and losses is the persistence of income shocks. Given constant hazard rates, every unemployed worker will remain unemployed for another year on expectation while an employed worker can expect to hold on to his job for another $11\frac{1}{2}$ years. Thus, leaving aside general equilibrium effects via job finding rates, the costs of unemployment insurance are borne largely by today's employed while the benefits accrue primarily to today's unemployed.

As a measure of the transition cost we take the difference between the static and dynamic effects. This transition cost has two sources. One is the foregone consumption during the transition to the new steady-state asset distribution with more self-insurance. This cost is relevant during spells of employment, which is when individuals save. The other source is the utility loss incurred by those who experience unemployment during the early periods after the reform when their asset holdings are still inadequately low. Transition costs from both sources are higher per percentage point benefit reduction the lower benefits are. This is reflected in the more than linear increase of transition costs in the size of the reform (cf. the last column of table 7). Costs arising from accumulation increase because average asset holdings increase more than linearly as benefits decrease. And due to the concavity of the

utility function the drop in flow utility of those becoming unemployed early on after the reform is more than proportional to the reduction of benefits for a given asset level.

Unsurprisingly in the light of the above, the size of transition costs is sensitive to the calibration of the utility function and the interest rate. Higher risk-aversion and time preference rates and lower interest rates yield significantly less positive dynamic welfare effects of lowering unemployment benefits even though the steady-state results are only moderately less positive. For the alternative calibration mentioned in section 3.4 with a relative risk aversion of 3, time discount rate of 8% annually and interest rate of 1% annually, for example, a reduction of the net replacement rate to 50% has severely negative consequences for all workers.²⁰

3.5.2 The Effects of Announcement and the Nature of Turnover Costs

An obvious (and realistic) way to reduce the welfare cost of the transition is to announce the reduction of the replacement rate some time before it takes effect. This gives people time to partially adjust their asset holdings and be better self-insured once the reform hits. Also, since wage setting will be such as to give rise to job creation less people will be unemployed at the time of benefit reduction. But, on the other hand, announcement also means delaying the gains from the reform. We calculated the transition for several reforms with previous announcement of 12 or 24 months. It turns out that the cost of transition can be greatly reduced or even over-compensated such that a reduction of benefits to a replacement rate of 30% still improves average welfare of workers and only slightly harms the unemployed. In the

²⁰Our results are clearly at odds with the ones presented in Joseph and Weitzenblum (2003). The discrepancies arise from the calibration rather than from the model used. The principal differences are that Joseph and Weitzenblum do not calibrate the tax burden and that their elasticity of unemployment with respect to the replacement ratio is very low (despite the endogenous search choice!). Moreover, at the same replacement rate that I am using unemployment is twice as high initially, which means the consequences of a reform for the unemployed have a much higher weight in welfare evaluations.

following we will argue, however, that these strong announcement effects do not stem from improved self-insurance but rather from the redistribution of gains from firms to workers and that they are peculiar to the specification of recruitment costs used.

Table 8 gives exemplary results for a 20 percentage point reduction in the replacement ratio with various announcement periods for the baseline calibration.²¹ In addition to the welfare effects for the unemployed and the employed it reports the effect of the reform on the value of a filled job. Clearly, this value decreases in the length of the announcement period while the welfare gains of both worker types increase. In fact, the two opposite effects are intimately linked. For the self-insurance effect of announcing the reform in advance is only part of the reason for the welfare gains of individuals. The other part is due to the Nash bargaining through which announcement allows workers to extract (parts of) the windfall gains firms make in case of an unannounced reform. To see this, recall the expressions for the value of a filled job in equations (3.3) and (3.5). When an unannounced reform is introduced, the gross wage drops and vacancy creation shoots up (cf. figure 9). This implies that the value of a filled job jumps up as well. The intuition is that, given the new, high value of market tightness, recruiting a worker is very costly and hence the value of having one is higher than before.

With announcement, we still have that after the reduction in benefits the outside option of workers worsens and hence the value of a filled job increases. But now backward recursion implies that this effect feeds through into the value of a filled job, J , in all periods back to the point of announcement. Through Nash bargaining the workers appropriate part of this gain in J in the form of higher wages. The zero-profit condition at the same time implies that vacancies increase as we get closer to the enactment of the reform. These effects can be seen in figure 10. The longer the reform is announced, the less market tightness jumps and the lower the firms' windfall. Hence, the effect of announcement is not only to give the workers time to self-insure but moreover to transfer resources to them from the firms, which in turn

²¹For the alternative calibration mentioned above welfare effects are positive only with 24 months announcement.

improves their capacity to build up assets. Nevertheless, with view to the high future value of jobs vacancy creation sets in immediately, which further improves the lot of the workers.

The above dynamics logically arise from wage determination through Nash bargaining in combination with rents depending on the ratio of vacancies to unemployment. If the value of the firm were independent of market tightness we would not observe the arguably unrealistic spike in the wage. It may therefore be worth considering the opposite polar case of a fixed cost per job. This case obtains in a world in which, instead of periodic vacancy costs, an initial training is required for each worker. Denoting the fixed cost per job by K , equation (3.4) then turns into

$$O_t = \frac{1}{1+r} [q(\theta_t)(J_{t+1} - K) + (1 - q(\theta_t))O_{t+1}], \quad (3.11)$$

which by free entry clearly implies

$$J = K. \quad (3.12)$$

The gross wage is obviously constant at $w = p - (r + \lambda)K$. Vacancies and hence unemployment respond much more strongly to unemployment benefit changes because the dampening effect of vacancy costs is missing. In fact, for the same calibration used all along the elasticity of unemployment with respect to the replacement rate is about 1.5 (compared to 0.75 in the benchmark case). The welfare effects of lowering benefits are therefore much more positive than under the specification used above. For an unannounced one-off reduction of the replacement rate to 40% welfare gains are 4.3% for the unemployed and 4.9% for the employed.²² But in this setting not much is to be gained from announcing the reform in advance. Relative to a reform with no announcement, announcement two years ahead makes everybody worse off,

²²The reason they are so similar for unemployed and employed workers is that with the value of a filled job fixed Nash-bargaining holds the difference in values of the unemployed and the employed fixed for the median asset level, such that the divergence only stems from the different asset distributions for the two types.

and announcement one year ahead only benefits the unemployed. The big welfare effect from extracting the firms' windfalls being absent, it seems that the cost of delaying the gains from the reform dominates the gain in self-insurance.

3.5.3 The Optimal Path of Reform

The purpose of this section is to investigate whether more sophisticated, gradual reforms can increase the welfare gains by mitigating the trade-off between, on the one hand, delaying the efficiency gains and, on the other hand, improving self-insurance (and extracting firms' rents in the vacancy cost specification). Clearly, it is numerically infeasible to optimize over the entire path of the transition. Instead, we constrain the problem to searching over several classes of continuous functions of time with two degrees of freedom (after fixing beginning and endpoint). Further we impose that the new final replacement rate be reached after 240 periods (20 years) at the latest. The functional classes are monotonously decreasing polynomials of degree three, hyperbolae and linear transformations of decreasing segments of the density function of the normal distribution.²³ These functions allow to check globally concave and convex paths as well as paths with inflection points and announcement.

We search for the optimal path of reform for both specifications of turnover costs, the standard one with a periodic vacancy posting cost and the alternative of a fixed recruitment cost introduced in the last section. The results reflect the dichotomy that was already found for the effects of announcing one-off benefit reductions in advance: When hiring a worker involves a fixed recruitment cost and the goal is to maximize average worker welfare, the reform should be implemented immediately and at once. For all three functional classes tried the path converged (as far as possible)

²³Precisely, for $0 \leq t \leq 240$ we use

$$\rho_t = \begin{cases} \rho_{old} & t \leq x \\ \alpha + \beta f(t - x | \max(x, 0), \sigma) & t > x \end{cases}$$

where f is the Gaussian density function. $x > 0$ corresponds to a reform announced x periods in advance. $x < 0$ means the time path of the replacement rate is described by a part of the right half of the Gaussian density function.

towards the announced one-off reduction and never yielded a higher welfare. If the goal is to find the path that is optimal for the unemployed, small improvements can be achieved by announcing the policy four months ahead and then reducing benefits gradually over almost two years.

For the standard specification with market-dependent vacancy costs, by contrast, some announcement and/or gradualism in the implementation is desirable. Both forms of delaying the full reform allow workers to appropriate the gains that in the case of the unannounced one-off change in benefits accrued as windfalls to firms. For a reduction of the replacement rate from 60% to 50% a gradual but faster than linear phasing in of the reform seems almost equivalent in terms of welfare to a path that involves an announcement period of 16 months and then a very quick drop in the replacement rate.

In any case, it is noteworthy that neither 'optimal' path yields significant improvements over a one-off reform announced 24 months ahead. For the fixed-cost variant we already saw that allowing for sophisticated reform paths is not helpful. This suggests that fine-tuning the transition to reduce the welfare losses due to lack of appropriate self-insurance is not very important.²⁴ It seems that the effects working through (not) delaying efficiency gains and rent extraction dominate. Also, that part of the cost of transition that is due to foregone consumption in the process of asset accumulation cannot be avoided anyway. Thus, the welfare effects of reducing unemployment insurance that we found in sections 3.5.1 and 3.5.2 on the basis of simple reforms are actually very good approximations of what can be achieved. What also emerges is that ultimately the optimal reform size and path depend paramountly on what is the correct assumption about the nature of hiring costs. We have shown results for the two extreme cases. The truth is likely to be somewhere in between. Solutions of the model for intermediate specifications involving a fixed and a variable hiring cost turn out to be convex combinations of those two polar cases in all respects.

²⁴Our simulations suggest that this is true irrespective of the size of the reform.

3.6 Conclusion

In this paper the Mortensen-Pissarides matching model has been introduced in an incomplete markets setting in order to study the dynamics and welfare effects of reductions in unemployment benefits. The various numerical experiments performed for our calibration to Germany in the mid-1990s suggest that reductions of the replacement rate in the order of 10 or 20 percentage points would be welfare improving for workers. However, the welfare gains would be much smaller than simple steady-state comparisons suggest because there are significant transition costs associated with the need for individuals to increase their self-insurance capacity by accumulating higher savings. It was further shown that both the optimal size of reform and its optimal timing depend strongly on the assumptions one makes about the nature of turnover costs faced by firms. Under the standard Mortensen-Pissarides assumption that the costs of hiring a worker increase in the equilibrium ratio of vacancies to unemployed parts of the gains from reducing unemployment benefits are crowded out by the increase in (wasteful) vacancy costs. Moreover, since sudden increases in vacancies confer windfall gains upon firms with a filled job, a gradual or announced reform is preferable. It allows workers to appropriate most of those windfalls. Under the alternative assumption of a fixed, market independent cost of hiring a worker, on the other hand, a reduction in benefits should come as fast and fully as possible because the cost of delaying the efficiency gains dominates the gains from giving people time to self-insure before benefits are actually lowered.

Let us conclude with a final remark regarding limitations of the model used in this paper. It stems from the fact that we use Nash-bargaining to determine wages. This is widely done in the literature and there is as yet no well established alternative, but recently a quest for alternative mechanisms has set off because, in particular when used in business-cycle applications, Nash-bargaining causes difficulties in matching the data.²⁵ It seems that a method of wage determination that gives rise to more wage stickiness would be more appropriate. The consequence for our model would likely

²⁵Shimer (2002) carefully demonstrates the insufficiencies of Nash-bargaining.

be that employment dynamics would be slower, which would tend to reduce welfare gains, even though not by much. It could however worsen the lot of the unemployed and thereby accentuate equity issues involved in reducing unemployment insurance.

3.7 Appendix

3.7.1 Job Acceptance

Lemma 7 *With $\rho_t \in [0, 1]$ and $1 - \lambda \geq \theta_t q(\theta_t)$ an individual never rejects a job regardless of his asset level.*

Proof. The proof will make use of a revealed preference argument. Time indices are dropped for ease of notation.

Denote by $V(a, e)$ and $V(a, u)$ the maximized values of an employed and an unemployed worker respectively. Let $(c_u^*, a_u'^*)$ describe the unemployed individual's optimal allocation of his current income $b + (1 + r)a$ to consumption and savings. Given our assumption that b is no greater than the net wage, this choice is in the feasible set of a worker with assets a who holds a job as well. Suppose the employed worker chose $(c_u^*, a_u'^*)$. Then his utility from consumption today would be equal to that of the unemployed and next period he would hold the same assets as the unemployed. With the same assets, starting from next period he cannot be worse off than the one who was unemployed today because i) his probability of being matched no less than for the unemployed and ii) the individual could turn down a job offer if the value of unemployment were higher. Hence, for the same consumption and asset choice the employed worker is at least as well off as the unemployed worker. Since $(c_u^*, a_u'^*)$ is feasible for him we can conclude that for his optimal choice he is also at least as well off. Thus $V(a, e) \geq V(a, u)$ and a job offer is never rejected. ■

3.7.2 The Algorithms

Computation of the steady state

To find the steady state the following computational strategy is employed:

1. The value functions of the worker and the unemployed are approximated by Schumaker splines in a on a log-linear grid.

2. For a given θ , i.e. for given transition probabilities and wages, the stationary asset distributions of the unemployed and the employed are calculated by forward iterating the asset distributions. That is to say, in a first step next period's asset holdings are determined for each individual state vector (a, s) on a grid. Then these new asset levels are attributed to the support points of the grid adjusting the densities such as to preserve the means. Further the distributions are adjusted to reflect the transition probabilities between employment and unemployment (compare equation (3.10)). The median assets of employed workers, a^m , are derived from the asset distribution for employed workers.
3. The steady state version of equation (3.8) is used to update θ .

Steps 1-3 are repeated until convergence.

Computation of the transition dynamics

To characterize the transition towards steady-state following a policy shock I solve for the time paths of the variables as follows:

1. Solve for the steady-states under the old and the new policy regime.
2. Choose T as the number of out-of-steady-state periods considered. I.e. in the T th period before the new steady-state is reached the policy shock occurs.
3. Postulate a path for a^m and u in the $T - 1$ periods following the shock.
4. Calculate θ_1 from $J^{SS,new}$ (cf. equation 3.5).
5. Calculate the wage for period 1 from the bargaining problem.
6. Approximate the value function in period 1 given w_1 and θ_1 and the fact that from the next period on we are in steady state.
7. Repeat steps 4-6 for all T periods always using the last period's value function to calculate the new one.

8. Using the wealth distribution in the old steady-state and the calculated paths for θ , wages, and value functions, calculate the evolution of the wealth distribution and hence of a^m . Using the unemployment rate in the old steady-state and the calculated path for θ find a new path for unemployment.
9. Update the guess for the paths of median assets and unemployment and go back to step 4 until convergence.

Chapter 4

Optimal Capital and Labor Taxes with Heterogeneous Agents: Making Everybody Happy (with Albert Marcet)

4.1 Introduction

It has been known for a while that under a variety of circumstances optimal capital taxes are zero in the long run. This result, which originally goes back to Chamley (1986) and Judd (1985), has proven very resilient to many challenges. A quick summary of the literature is that there are ways to go away from the zero long run capital taxes, but often the deviation from zero is small and it can be positive or negative depending on minor changes to the model.

Surprisingly, even with heterogeneous agents capital should not be taxed in steady state no matter how a social planner weights the utilities of the different agents. Atkeson, Chari and Kehoe (1999), Chamley (1986), and Judd (1985) provide results of this kind.¹ This seems to suggest that there is no tradeoff between efficiency and equity: as long as the time path of capital taxes is chosen appropriately all agents can enjoy the efficiency gains that occur with capital taxes going to zero.

It is also well known that optimal capital taxes in the first few periods are not zero but in fact quite high.² Nevertheless, Lucas (1990) showed that even if the optimal transition were ignored, if capital taxes were abolished immediately and all tax revenue was collected from labor taxes, the welfare of the representative agent would increase. This result has been interpreted as suggesting that designing the transition optimally is not very important.

However, more recently several contributions have shown that this conclusion is not warranted if we acknowledge the heterogeneity of agents: Garcia-Milà, Marcet and Ventura (1995) show that for a reasonable calibration of inequality, if capital taxes were abolished immediately, output would increase but large parts of the population would lose a lot of utility relative to the status quo. Similar results have been obtained in different setups by Correia (1995), Domeij and Heathcote (2004), and Conesa and

¹Chamley and Judd allow for lump-sum transfers, which is a severe restriction. Moreover, it is not clear whether they correctly take into account the equilibrium conditions. The proof in Atkeson et al. (1999) does not suffer from these limitations. Note, though, that Atkeson, Chari, and Kehoe do not impose an upper bound on the admissible capital tax rates.

²For certain common utility functions it can be shown that capital taxes are high only one period and then drop to zero.

Krueger (2006).

Thus, once agents are heterogeneous and equity considerations play a role, the time path of capital taxes until their abolition does seem to matter. Abolishing them too fast is impossible in a Pareto improving way, but - given the steady state results - there should be a way of reaching the steady state with no capital taxes without harming any agent.

We study this issue and analyze the whole path for capital and labor taxes when fiscal policy is restricted to delivering allocations that are Pareto efficient AND Pareto superior to the status quo.³ As an additional restriction we impose an upper bound on capital taxes for all periods. We think of this bound as a requirement that a policy has to satisfy if it is to be credible. We know that optimal capital taxes under full commitment have a tendency to be initially very high, and we think it would be difficult for any government to convince investors that it is going to suppress capital taxes in the future if the first thing the government does is to hike capital taxes.⁴ Finally, to make the redistribution problem meaningful, we require tax rates to be common to all agents and exclude redistributive lump sum transfers.

We find that the optimal reform under all these constraints is quite different from the steady state analysis that much of the literature has concentrated on. In our baseline model the transition is much longer than absent distributive issues. Capital taxes take a very long time to be zero in order for *all* the agents to be better off than under the status quo - at least about ten years, if the degree of inequality is reasonably calibrated. The reason is that if taxes are abolished too quickly, as in Garcia-Milà et al. (1995), the redistributive effect is too strong and it makes the workers worse off despite the gain in aggregate efficiency. Also, we find that optimal labor taxes in the

³The recent work of Flodén (2006) studies optimal policy when the Ramsey planner maximizes the utility a certain agent. Therefore, the policies he studies do have a transition and capital taxes take a while before they hit zero. The results reinforce the view of the papers described in the previous paragraph: there are often large losses for a large part of the population, so it would seem that the efficiency/equity tradeoff is a big issue in the potential disappearance of capital taxes. But still, since Flodén does not look for a Pareto efficient/Pareto improving policy, it would seem he only studies part of the story.

⁴Lucas (1990) offered a similar motivation to study taxes that are constant over time.

initial periods are *lower* than in the status quo: in this way the planner can engineer growth while capital taxes are at the limit. Therefore, the policy recommendation for the short run is exactly the opposite of the steady state: lower labor taxes and high capital taxes are needed for the short-medium run (about 10 years).⁵ Thus, this model is a case where the so called 'timeless perspective', i.e. policy analysis neglecting the transition, would not only give the reverse recommendation to the actually optimal one, but it would yield very low welfare for some agents, as Garcia-Milà et al. (1995) and others found.⁶

We also find that the initially low labor taxes imply deficits that are typically not fully repaid, and that in the long run the government is in debt. There is recently a renewed interest in studying the determinants of the optimal level of debt in various setups.⁷ Our paper shows that implementing an optimal reform could be one such factor.

In our model two sources of distortions constrain the optimal policy, the absence of lump-sum taxes to raise revenues, and the absence of a lump-sum instrument for redistribution. In order to isolate the effects of each distortion we consider modifications of our baseline model that either allow for lump-sum transfers or fix labor supply. We thus show that aggregate welfare gains are much smaller in our baseline model than if a redistributive lump sum were available or if we were only interested in aggregate efficiency. The model with fixed labor supply also underlines how importantly equity concerns constrain the solution because the distortions from labor taxation are absent. Capital taxes can be abolished after 10 years at the earliest, while they would be suppressed immediately if there were transfers because labor taxes are non-distortive. Moreover, in this model the planner cannot engineer early growth by lowering early labor taxes. This is reflected in much slower accumulation

⁵The fact that the transition takes so long questions the validity in practice of such a policy, since the government has to make credible announcements that taxes are abolished while the economy is not growing.

⁶Also, this explains why they were finding such bad results when the long run optimum was implemented from period zero: it is a feature of optimal (Pareto improving) policy to not only suppress capital taxes in the long run, but also to lower labor taxes in the first periods.

⁷See Faraglia, Marcet and Scott (2006) and the references therein.

of capital during the periods with high capital taxes.

We find that in this setup the frontier of feasible equilibria sometimes has an increasing part, even in the range of equilibria that are Pareto superior to the status quo. This implies that the Pareto optimal frontier may not cover the range of all possible utilities and that the degree of redistribution is limited by the presence of distortionary taxes: if taxes are chosen optimally it may not be possible to find Pareto superior allocations that leave one agent indifferent and redistribute all the gains from optimality to the other agent. In our calibrated examples it was always the case that the 'capitalist' (i.e. the agent with a lot of physical wealth relative to human wealth) could be made to enjoy all the gains in increased efficiency while the worker could sometimes enjoy at most part of the benefit. In these cases, if the government insists on leaving the capitalist in the status quo it can only do so by lowering the utility gain of the worker also and, therefore, by pursuing a Pareto inefficient policy, even if it places itself at the boundary of the feasible equilibrium set.

The focus on Pareto optimal allocations implies that, for the right weight, the planner behaves as if he had a welfare function. But trying to interpret the weights if welfare functions in the objective of the planner per se could be quite misleading. We will show cases where the implied weight "seems" very large even though it does not achieve a large redistribution.

In solving the model we have to take care of a few technical issues. The upper bound in taxes introduces a forward looking constraint that requires the introduction of recursive contracts to be solved. The solution then depends on a second state variable in addition to capital.

The paper is organized as follows: In section 4.2 we lay out our baseline model as well as the modifications we consider. Section 4.3 discusses some properties of the models. Among others we provide a proof that capital taxes are zero in steady state in our setup, to which existing proofs do not fully apply. Our numerical results are discussed in section 4.4.

4.2 Models

4.2.1 The Baseline Model

We lay out a standard dynamic competitive equilibrium model with two agents that differ in their sources of income, a government that taxes capital and labor, no uncertainty. The model is as in GMV with only two agents. Tax rates τ_t^l and τ_t^k are allowed to be time dependent. Also assume no growth, so $\mu = 1$ and g constant.

The environment

More precisely, there are two consumers $j = 1, 2$ with utility $\sum_{t=0}^{\infty} \delta^t [u(c_{j,t}) + v(l_{j,t})]$ where c is consumption and l is labor of each agent each period. Agents differ in their initial wealth $k_{j,-1}$ and their labor productivity ϕ_j . Agent j obtains income from renting his/her capital at the rental price r_t and from selling his/her labor for a wage $w_t \phi_j$, pays labor taxes τ_t^l on labor income and capital taxes τ_t^k on capital income net of depreciation allowances. The period- t budget constraint is given by

$$c_{j,t} + k_{j,t-1} = w_t \phi_j l_{j,t} (1 - \tau_t^l) + k_{j,t-1} (1 + (r_t - d)(1 - \tau_t^k)) \quad \text{for } j = 1, 2 \quad (4.1)$$

capturing the fact that consumers are responsible for investment.

Firms maximize profits, have a production function $F(k_{t-1}, e_t)$ where e is total efficiency units of labor, k is total capital.

Government chooses capital and labor taxes and consumes g every period, has the standard budget constraint, it saves in capital and has initial capital k_{-1}^g .

We normalize each agents' mass to be $1/2$. Market clearing conditions are

$$\frac{1}{2} \sum_{j=1}^2 \phi_j l_{j,t} = e_t \quad (4.2)$$

$$k_t = k_t^g + \frac{1}{2} \sum_{j=1}^2 k_{j,t} \quad (4.3)$$

$$\frac{1}{2} \sum_{j=1}^2 c_{j,t} + g + k_t - (1-d)k_{t-1} = F(k_{t-1}, e_t) \quad (4.4)$$

Competitive Equilibria

The equilibrium concept is standard, agents take prices and taxes as given, maximize their own utility, markets clear, the budget constraint of the government is satisfied. Agents' maximization implies

$$u'(c_{j,t}) = \delta u'(c_{j,t+1}) (1 + (r_{t+1} - d)(1 - \tau_{t+1}^k)) \quad (4.5)$$

$$u'(c_{j,t}) w_t (1 - \tau_t^l) \phi_j + v'(l_{j,t}) = 0 \quad (4.6)$$

for all t and j . Firms' maximization implies factor prices equal marginal product to set $r_t = F_1(k_{t-1}, e_t)$ and $w_t = F_2(k_{t-1}, e_t)$.

The budget constraints of the agents written in present value form are

$$\sum_{t=0}^{\infty} \delta^t \frac{u'(c_{1,t})}{u'(c_{1,0})} (c_{j,t} - w_t \phi_j l_{j,t} (1 - \tau_t^l)) = k_{j,-1} (1 + (r_0 - d)(1 - \tau_0^k)) \quad \text{for } j = 1, 2 \quad (4.7)$$

Due to Walras' law the budget constraint of the government is implied by these and market clearing, so we ignore it.

Assuming further CRRA u and v , each with risk aversion $\sigma_c, \sigma_l < 0$, FOC for capital and labor imply

$$\frac{c_{2,t}}{c_{1,t}} = \lambda \quad \text{and} \quad \frac{1 - l_{2,t}}{1 - l_{1,t}} = \lambda \frac{\sigma_c}{\sigma_l} \left(\frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}} \quad \text{for all } t$$

for some λ that is constant through time.

Using the above equation and the primal approach in the usual way it is easy to see that for a competitive equilibrium to hold it is necessary and sufficient to find a

constant λ and a sequence $\{c_t^1, k_t, l_t^1\}$ satisfying

$$\sum_{t=0}^{\infty} \delta^t (u'(c_{1,t})c_{1,t} + v'(l_{1,t}) l_{1,t}) = u'(c_{1,0}) k_{1,-1}(1 + (r_0 - d)(1 - \tau_0^k)) \quad (4.8)$$

$$\sum_{t=0}^{\infty} \delta^t \left(u'(c_{1,t})\lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f(\lambda, l_{1,t}) \right) = u'(c_{1,0}) k_{2,-1}(1 + (r_0 - d)(1 - \tau_0^k)) \quad (4.9)$$

and feasibility. Here $f(\lambda, l_{1,t})$ is defined as

$$f(\lambda, l_{1,t}) \equiv 1 - (1 - l_{1,t}) \lambda^{\frac{\sigma_c}{\sigma_l}} \left(\frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}}$$

and it is the value of $l_{2,t}$ that solves (4.6) for each possible value of the endogenous variables λ, l_t^1 .

Taxes are then found as a residual from (4.5) and (4.6). Consumption and labor of agent 2 is found from λ and f and individual capital is backed out from the budget constraint period by period.⁸

Constraints on Policy

As usual, the Ramsey optimizer is restricted to choosing allocations, taxes and prices that are compatible with the above equilibrium conditions.

We now introduce some additional constraints to the choice of policy. First of all, we assume that the planner chooses Pareto efficient allocations. With the usual argument, this is achieved by assuming that the planner maximizes the utility of agent 1 subject to the constraint that agent 2 has a minimum value of utility:

$$\sum_{t=0}^{\infty} \delta^t [u(c_{2,t}) + v(l_{2,t})] \geq \underline{U}_2 \quad (4.10)$$

Varying the value of the minimum utility \underline{U}_2 we can trace the frontier of Pareto

⁸For the details on this model see Garcia-Milà et al. (1995). For details on the general primal approach, see Chari and Kehoe (n.d.).

efficient allocations. We will concentrate our attention on values of \underline{U}_2 that guarantee a Pareto improvement over some status quo utility U_{SQ}^j that would be achieved with some taxation scheme that is already in place.⁹ We call these POPI (=Pareto optimal Pareto improving) allocations, and they can be found by considering minimum utility values such that $\underline{U}_2 \geq U_{SQ}^2$ and by checking, after the problem has been solved, that

$$\sum_{t=0}^{\infty} \delta^t [u(c_{1,t}^*) + v(l_{1,t}^*)] \geq U_{SQ}^1$$

where * denotes the optimized value.

Also, we introduce tax limits ensuring that capital taxes never go beyond a certain level, so we introduce the constraint $\tau_t^k \leq \tilde{\tau}$ for all t and some given constant $\tilde{\tau}$. This tax limit is introduced to avoid the usual pattern that optimal capital taxes usually shift all the tax burden to early capital income, and to avoid the criticism that the Ramsey optimal tax is not a credible tax to announce (see Lucas (1990)). If the upper bound $\tilde{\tau}$ is equal to the status quo the government will only choose decreasing paths for capital taxes.

To enforce this tax limit it is necessary and sufficient to introduce the following constraint

$$u'(c_{1,t}) \geq \delta u'(c_{1,t+1}) (1 + (r_{t+1} - d)(1 - \tilde{\tau})) \quad \text{for all } t > 0 \quad \text{and} \quad (4.11)$$

$$\tau_0^k \leq \tilde{\tau} \quad (4.12)$$

The first equation insures that the actual capital tax that is backed out from (4.5) satisfies the limit and it allows us to use the primal approach where taxes do not appear explicitly.¹⁰ The limit for time 0 (4.12) is standard in models of optimal capital taxes (see Chari and Kehoe (n.d.)) and has to be specified separately.

⁹The status quo utility depends on the distribution of initial capital but we leave this dependence implicit.

¹⁰Atkeson et al. (1999) considered a similar bound explicitly.

To summarize, the planner solves

$$\begin{aligned}
& \max_{\lambda, \{c_t^1, k_t, l_t^1\}_{t=0}^\infty} && \sum_{t=0}^{\infty} \delta^t [u(c_{1,t}) + v(l_{1,t})] \\
& \text{s.t.} && \sum_{t=0}^{\infty} \delta^t [u(c_{2,t}) + v(l_{2,t})] \geq \underline{U}^2
\end{aligned} \tag{4.13}$$

subject to feasibility (4.4) for all t , the implementability constraints (4.8) and (4.9) (for period 0 only) and the tax limit constraint (4.11) for all periods $t > 0$. Notice that λ is a choice variable that has to be maximized over. Initial wealth, the tax bound $\tilde{\tau}$ and the utility bound \underline{U}^2 are given constants in this problem.

Letting α be the lagrange multiplier of the minimum utility constraint (4.13), letting Δ_1, Δ_2 be the multipliers of the (4.8) and (4.9) normalized by $u'(c_{1,0})$, and letting γ_t be the multiplier of (4.11), the Lagrangian is

$$\begin{aligned}
L = & \sum_{t=0}^{\infty} \delta^t \left[(u(c_{1,t}) + v(l_{1,t})) + \alpha(u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))) + \right. \\
& \Delta_1 (u'(c_{1,t})c_{1,t} + v'(l_{1,t}) l_{1,t}) + \\
& \Delta_2 (u'(c_{1,t})\lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t})f(\lambda, l_{1,t})) + \\
& \gamma_t (u'(c_{1,t}) - \delta u'(c_{1,t+1})(1 + (r_{t+1} - d)(1 - \tilde{\tau}))) - \\
& \left. \mu_t \left(\frac{1 + \lambda}{2} c_{1,t} + g + k_t - (1 - d)k_{t-1} - F(k_{t-1}, e_t) \right) \right] - \mathbf{A}
\end{aligned} \tag{4.14}$$

where $\mathbf{A} = u'(c_{1,0})[\Delta_1 k_{1,-1}(1 + (r_0 - d)(1 - \tau_0^k)) + \Delta_2 k_{2,-1}(1 + (r_0 - d)(1 - \tau_0^k))]$. Further, $\gamma_t, \alpha \geq 0$ and they satisfy the usual slackness conditions.

The first line of this Lagrangian has the usual interpretation: a Pareto efficient allocation amounts to solving a welfare function where the planner weighs linearly the utility of both agents, where the weight of agent 1 is normalized to one and the weight of agent two is the lagrange multiplier of the minimum utility constraint. The other lines represent all the constraints in the problem of the planner.

First order conditions are derived as usual with respect to capital, labor and consumption. They are shown in the appendix. Here we only comment on features of these first order conditions that differ from other papers on dynamic taxation.

Notice that λ is a constant to be found, just as the Δ, α 's. While the optimality of the Δ 's and α simply insures feasibility of some policy, many values of λ are compatible with feasibility, and the derivative with respect to λ has to be set to zero to insure an optimal choice. This derivative is

$$\begin{aligned} \sum_{t=0}^{\infty} \delta^t [\alpha (u'(\lambda c_{1,t})c_{1,t} + v'(f(\lambda, l_{1,t}))f_{\lambda}(\lambda, l_{1,t})) + & \quad (4.15) \\ \Delta_2(u'(c_{1,t})c_{1,t} + \frac{\phi_2}{\phi_1}v'(l_{1,t})f_{\lambda}(\lambda, l_{1,t})) - & \\ \mu_t \frac{1}{2}(c_{1,t} - F_e(k_{t-1}, e_t)\phi_2 f_{\lambda}(\lambda, l_{1,t}))] = 0 & \end{aligned}$$

This takes into account the fact that the planner can vary the ratio of consumptions of the agents, in effect, by varying the total tax burden of labor or capital. α is also a constant but it has to be set to a level insure the minimum utility constraints.

Solving the model involves iterating on the constants $\alpha, \Delta_1, \Delta_2, \lambda$ until a solution series is found that satisfies the period- t FOC and it insures that the all the conditions involving infinite sums computed from period zero (namely, the implementability constraints, the minimum utility constraint, and the FOC for λ) hold.

The multipliers have to satisfy the necessary slackness conditions. Although these are standard, since they are key for some important features of the solution, we now state in detail these conditions. In particular, the slackness condition for α (the multiplier of (4.13)) is

$$\begin{aligned} \text{either } \alpha > 0 \text{ and } \sum_{t=0}^{\infty} \delta^t [u(c_{2,t}) + v(l_{2,t})] &= \underline{U}^2 \\ \text{or } \alpha = 0 \text{ and } \sum_{t=0}^{\infty} \delta^t [u(c_{2,t}) + v(l_{2,t})] &\geq \underline{U}^2 \end{aligned}$$

In other words, either the minimum utility constraint is binding and in the Lagrangian the planner maximizes the weighted utility of both agents with weight 1 for agent 1 and weight α for agent 2, or the minimum utility constraint is NOT binding and the planner gives zero weight to agent 2. Even though the latter case is not usually a relevant case in studying PO allocations in models without frictions, we will see that it can arise in the type of model we are considering. The reason is that the frontier of the set of possible equilibria will have an increasing part.

Similarly, for the γ 's and for each t , we have

$$\begin{aligned} \text{either } \gamma_t &> 0 \text{ and } u'(c_{1,t}) = \delta u'(c_{1,t+1})(1 + (r_{t+1} - d)(1 - \tilde{\tau})) \\ \text{or } \gamma_t &= 0 \text{ and } u'(c_{1,t}) \geq \delta u'(c_{1,t+1})(1 + (r_{t+1} - d)(1 - \tilde{\tau})) \end{aligned}$$

It turns out that the Δ_i 's may be positive or negative, since the corresponding PVBCs have to be satisfied with equality. This becomes clear by looking at their economic interpretation. With two agents the marginal utility cost of distortive taxation is $\frac{\partial L}{\partial \tau_0^k} = u'(c_{1,0})[\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}](r_0 - d)$. Thus,

$$\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} \geq 0$$

with the inequality being strict as long as any taxes are raised after the initial period. This does not preclude one of the Δ_i being negative, which will in fact be the case whenever the constraints on redistribution that are imposed by the competitive equilibrium conditions are sufficiently severe. To see this consider a slightly modified model in which the social planner is allowed to redistribute initial wealth between agents by means of a transfer T .¹¹ All this modification does to the Lagrangian is to change the implementability constraints such that $\mathbf{A} = u'(c_{1,0})[\Delta_1(k_{1,-1}(1 + (r_0 - d)(1 - \tau_0^k)) - T) + \Delta_2(k_{2,-1}(1 + (r_0 - d)(1 - \tau_0^k)) + T)]$. Now the derivative of the Lagrangian with respect to the lump-sum transfer between

¹¹Below we will further discuss this modification as "modified model 1".

agents is $\frac{\partial L}{\partial T} = u'(c_{1,0})(\Delta_1 - \Delta_2)$. For any given T , and in particular for $T = 0$ as in our baseline model, this expression is a measure of the marginal utility cost of the transfer not being optimal. If the planner were free to choose T optimally, we would have $\Delta_1 = \Delta_2 > 0$. If the planner would like to redistribute more towards agent 2, $\Delta_1 - \Delta_2 > 0$ and vice versa. If the transfer is much too low (high) the derivative will be large in absolute value and Δ_2 (Δ_1) will be negative. In sum, while the weighted sum of the multipliers on the PVBCs is related to the cost of distortive taxation, their difference indicates the cost of not being able to redistribute lump sum. These multipliers thus capture in a simple way the two forces that drive the solution to our model away from the first best, the absence of lump-sum taxes and of agent-specific lump-sum transfers.

4.2.2 Modifications of the model

In order to learn about different aspects of the optimal policy it will be useful to consider three modifications of the model:

Modified Model 1: redistributive transfers (MM1)

We are mainly interested in the baseline model where a sequence of capital and labor taxes has an effect both in terms of redistribution of wealth and efficiency. Ideally, the tax authority would like to resolve these two issues separately and it would rather have another instrument to handle redistribution of wealth (just as in, say, intermediate micro, one shows that any Pareto efficient allocation can be supported by competitive equilibrium with redistribution of wealth). In order to study how the optimal program is affected by the presence or absence of this extra instrument, we modify the problem by assuming the government can resort to a lump sum redistributive transfer across agents. We denote by T the amount of goods that the planner transfers from agent 1 to agent 2. This changes the budget constraints by subtracting T from the right side of the budget constraint of agent 1, namely, equation (4.7) for $j = 1$ (and adding

T to the same constraint for 2). Correspondingly, we subtract T on the right side of constraint (4.8) and we add T in (4.9).

Also, we add some limits to these transfers:

$$\underline{T} \leq T \leq \bar{T}$$

Here, all of $\underline{T}, T, \bar{T}$ could be positive or negative. Subject to this, the planner chooses T optimally.

That is, the government can redistribute wealth but within some limits. It is clear that if $-\underline{T}, \bar{T}$ are sufficiently large the government can achieve a kind of first second best, where there are no redistributive issues, and the government faces only the usual tradeoff between taxes today or tomorrow and versus capital and labor. This is stated formally in the following

Result: If $-\underline{T}, \bar{T}$ are sufficiently large the solution to the modified problem is as in the baseline model ignoring the budget constraints (4.8) and (4.9) of the agents but considering the budget constraint of the government instead

The main model described in the previous section is, of course, a special case of this modified model with $\underline{T} = \bar{T} = 0$.

Modified Model 2: fixed labor input (MM2)

Modified model 2 is complementary to MM1. While modified model 1 serves to illustrate what the optimal program would look like absent distributional concerns that constrain the solution, modified model 2 isolates the effect of precisely these distributional concerns. By fixing labor input we eliminate the distortions from labor taxation that by themselves would lead the tax authority not to abolish capital taxes too fast. In this model, the transition to zero capital taxes would be immediate if it were not for the distributive effects of abolishing capital taxes.

Modified Model 3: time varying tax limits (MM3)

In the characterization of equilibrium it will be useful to study other types of tax limits. Instead of a uniform limit for all periods we assume capital taxes have to satisfy

$$\tau_t^k \leq \tilde{\tau}_t$$

where $\tilde{\tau}_t$ is a sequence of pre-specified time-varying tax limits, given to the planner. The case where $\tilde{\tau}_t$ is very large in the initial periods will be useful to study how much utility and redistributive power is lost by the fact that the capital taxes cannot be front-loaded in the usual way in our main model. The case where $\tilde{\tau}_t$ is very large in the last periods will be useful in the proof of asymptotic behavior of the model.

4.3 Characterization of equilibria

4.3.1 Steady state and zero capital taxes

First of all we derive the behavior in steady state. To the best of our knowledge there is no previous proof of zero long-run capital taxes that fully applies to our model, which features both a tax limit and heterogeneous agents but no lump-sum transfers or agent-specific taxes.¹²

Result: Assume log utility of consumption and $\tilde{\tau} > 0$. Further assume that there is free disposal on the part of the government. I.e. the government can collect more taxes than necessary to finance g and dump the rest. Then capital taxes are zero in

¹²When they consider heterogeneous agents, both Chamley (1986) and Judd (1985) allow for redistributive lump-sum transfers. Moreover, only Chamley limits frontloading of capital taxes by imposing a limit (of 100%) on the tax rate. Both Chamley and Judd are very vague about how they extend their results for representative agents to the heterogeneous agent case. Only in Chari and Kehoe (n.d.) and Atkeson et al. (1999) it is clear that indeed all competitive equilibrium constraints are taken into account. But for the case with heterogeneous agents and no transfers or agent specific taxes they only prove the case without a tax limit.

the long run. Moreover, for some finite N

$$\begin{aligned}\tau_t^k &= \tilde{\tau} \quad \text{for all } t \leq N, \\ &= 0 \quad \text{for all } t \geq N + 2,\end{aligned}$$

In other words, at some point, the capital tax rate jumps from the tax limit to zero in two periods.

Proof of result:

We proceed in two steps. First we show that it is not possible for all the FOC to be satisfied if the tax limit is binding forever. Then we show that capital taxes go from the limit to zero in two periods.

First of all, notice that in the log case the first order condition with respect to consumption for $t > 0$ becomes

$$\begin{aligned}c_{1,t}^{-1}(1 + \alpha) & - & (4.16) \\ -c_{1,t}^{-2}(\gamma_t - \delta\gamma_{t-1}(1 + (F_k(k_{t-1}, e_t) - d)(1 - \tilde{\tau}))) & = \mu_t \frac{1 + \lambda}{2}\end{aligned}$$

Then the FOC for consumption and labor for $t > 0$ at steady state for the variables imply

$$\begin{aligned}\bar{c}_1^{-1}(1 + \alpha) & + & (4.17) \\ -\bar{c}_1^{-2}(\gamma_t - \delta\gamma_{t-1}(1 + (F_k(\bar{k}, \bar{e}) - d)(1 - \tilde{\tau}))) & = \mu_t \frac{1 + \lambda}{2}\end{aligned}$$

$$\begin{aligned}
& -B(1 - \bar{l}_1)^{\sigma_l} \left(1 + \alpha \lambda^{\sigma_c} \frac{\phi_2}{\phi_1} f'(\lambda, \bar{l}_1) + \Delta_1 + \Delta_2 \frac{\phi_2}{\phi_1} f'(\lambda, \bar{l}_1) \right) + & (4.18) \\
& \sigma_l B (1 - \bar{l}_1)^{\sigma_l - 1} \left(\Delta_1 \bar{l}_1 + \Delta_2 \frac{\phi_2}{\phi_1} f(\lambda, \bar{l}_1) \right) + \\
& \gamma_{t-1} \bar{c}_1^{\sigma_c} (1 - \tilde{\tau}) F_{k,e}(\bar{k}, \bar{e}) \frac{1}{2} (\phi_1 + \phi_2 f'(\lambda, \bar{l}_1)) = \\
& -F_e(\bar{k}, \bar{e}) \frac{1}{2} (\phi_1 + \phi_2 f'(\lambda, \bar{l}_1)) \mu_t
\end{aligned}$$

Notice that we are only imposing steady state on the variables, not on the multipliers. This is natural because the real variables have natural bounds but the multipliers should not have bounds, otherwise there is no sense in which the Lagrangian is guaranteed to give a maximum.

At steady state and if the tax limit is at the bounds constraint (4.11) is satisfied as equality so that

$$\delta [1 + (F_k(\bar{k}, \bar{e}) - d)(1 - \tilde{\tau})] = 1$$

Also, collecting as constants those terms in (4.17) and (4.18) that do not depend on the multipliers γ or μ we have

$$\mathbf{A} - \bar{c}_1^{-2} (\gamma_t - \gamma_{t-1}) = \mu_t \frac{1 + \lambda}{2} \quad (4.19)$$

$$\mathbf{B} + \mathbf{C} \gamma_{t-1} = -\mu_t \quad (4.20)$$

for

$$\begin{aligned}
\mathbf{A} &= \bar{c}_1^{-1}(1 + \alpha) \\
\mathbf{B} &= \left[-B(1 - \bar{l}_1)^{\sigma_l} \left(1 + \alpha \lambda^{\sigma_c} \frac{\phi_2}{\phi_1} f'(\lambda, \bar{l}_1) \right) + \Delta_1 + \Delta_2 \frac{\phi_2}{\phi_1} f'(\lambda, \bar{l}_1) \right] + \\
&\quad \left. \sigma_l B(1 - \bar{l}_1)^{\sigma_l - 1} \left(\Delta_1 \bar{l}_1 + \Delta_2 \frac{\phi_2}{\phi_1} f(\lambda, \bar{l}_1) \right) \right] \frac{1}{F_e(\bar{k}, \bar{e})^{\frac{1}{2}} (\phi_1 + \phi_2 f'(\lambda, \bar{l}_1))} \\
\mathbf{C} &= \frac{\bar{c}_1^{\sigma_c} (1 - \tilde{\tau}) F_{k,e}(\bar{k}, \bar{e})}{F_e(\bar{k}, \bar{e})}
\end{aligned}$$

So, we have

$$\gamma_t = \bar{c}_1^2 \left[\mathbf{A} + \mathbf{B} \frac{1 + \lambda}{2} \right] + \gamma_{t-1} \left[\mathbf{C} \bar{c}_1^2 \frac{1 + \lambda}{2} \right] \quad (4.21)$$

Since $\mathbf{C} > 0$, γ goes to plus or minus infinity depending on the sign of the first bracket in this equation, regardless of the initial values of γ . If this bracket is negative this implies negative γ 's eventually, which is incompatible with an optimum. If this bracket is positive the above equation implies positive γ 's (going to infinite). Notice that according to (4.19) this explosive γ implies that μ is negative. In a sense the planner would prefer less output. But μ can never be negative if the government can just throw away tax revenues in excess of g and thus effectively reduce output. Thus, the tax limit cannot be binding forever.

Now we show that capital taxes go from limit to zero within two periods in finite time.

Let $N + 1$ be the *first* period where the tax limit is not binding, so that $\tau_{N+1}^k < \tilde{\tau}$ and $\tau_t^k = \tilde{\tau}$ for all $t \leq N$. Clearly, N is finite and well defined.

Now consider the "modified model 3" at the end of the previous section. Given N , consider the time-varying tax limits that leave $\tilde{\tau}_t^k = \tilde{\tau}$ for all $t \neq N + 1$ but where $\tilde{\tau}_{N+1}^k$ is very large (say, it is infinite) so as to insure beforehand the tax limit at period $N + 1$ can not be binding. Let us call this the "modified model 3.1". It is clear that the solution to this problem is equal to the solution of the baseline model, because we have just relaxed a tax limit that was not binding in the optimum of the baseline

model. Let us keep this fact in store for a while.

Now consider another modified model 3, let us call it modified model 3.2. The tax limits are now $\tilde{\tau}_t^k = \tilde{\tau}$ for all $t \leq N$ and $\tilde{\tau}_t^k$ is very large (say, infinite) to insure it is not binding for all $t > N$. Let us denote with $\hat{\cdot}$ the solution to this modified model 3.2.

Clearly the first order conditions for this modified model are the same as for the basic problem except that

$$\hat{\gamma}_t = 0 \quad \text{for all } t \geq N \quad (4.22)$$

(notice that γ_t is the multiplier associated with the constraint on τ_{t+1}^k , so that τ_{N+1}^k being unconstrained means $\gamma_N = 0$)

Combining (4.22) with (4.16), implies

$$\hat{c}_{1,t}^{-1}(1 + \hat{\alpha}) = \hat{\mu}_t \frac{1 + \hat{\lambda}}{2} \quad \text{for all } t \geq N + 1 \quad (4.23)$$

(notice, this last equation does not hold for $t = N$ because $\hat{\gamma}_{N-1} \neq 0$ appears in (4.16)). Plugging (4.22) in the FOC's with respect to capital we get

$$\hat{\mu}_t = \delta \hat{\mu}_{t+1} (1 + F_k(k_t, e_{t+1}) - d) \quad \text{for all } t \geq N$$

and using (4.23) we have

$$\hat{c}_{1,t}^{-1} = \delta \hat{c}_{1,t+1}^{-1} (1 + F_k(k_t, e_{t+1}) - d) \quad \text{for all } t \geq N + 1$$

Using the Euler equation of the consumer we conclude that $\hat{\tau}_t^k = 0$ for all $t \geq N + 2$.

Now, it is clear that the optimal solution for the modified model 3.2 is also feasible in the modified model 3.1, even though the latter is more restrictive, because the tax limit for $t > N$ in model 3.1 is positive so that $\hat{\tau}_t^k < \tilde{\tau}$. Therefore, the $\hat{\cdot}$ solution is also the solution to modified model 3.1.

Since we already argued that this had to be the solution to the baseline model as

well, this completes the proof.

4.3.2 Recursive Formulation

The model is recursive only after period 0. Formally, the structure of Marcet and Marimon (1998) does not apply starting at period $t = 0$ because the terms in the right side of the budget constraints have some endogenous variables (labor and consumption) that appear differently in period $T = 0$ than in all remaining periods.

To obtain a recursive formulation, we first observe that, by a similar argument as in Chari, Christiano and Kehoe (1994), the optimal solution is found by solving

$$\begin{aligned} \max_{\{c_t^1, k_t, l_t^1\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \delta^t [& u(c_{1,t}) + v(l_{1,t}) + \alpha(u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))) + \\ & \Delta_1(u'(c_{1,t})c_{1,t} + v'(l_{1,t}) l_{1,t}) + \\ & \Delta_2 \left(u'(c_{1,t})\lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) l_{2,t} \right) + \\ & \gamma_t u'(c_{1,t}) - \gamma_{t-1} u'(c_{1,t})(1 + (r_t - d)(1 - \tilde{\tau}))] \end{aligned}$$

subject to the tax limit and feasibility, and for fixed values of Δ 's, α , λ , γ_0 and k_0 . Notice that in this problem the series to be found starts at $t = 1$, and the choice in period zero is taken as given. Given the optimal choices in period zero for k, γ , given Δ 's, λ and α , we now consider the maximization for the periods $t > 0$. The results in Marcet and Marimon (1998) insure that an optimal policy function F

$$\begin{bmatrix} c_t^1 \\ k_t \\ l_t^1 \\ \gamma_t \end{bmatrix} = F(k_{t-1}, \gamma_{t-1})$$

could be found with the usual techniques to deliver the optimal policy.

Notice the dependence of F on the Δ 's, α and λ , this is left implicit.

4.3.3 Frontier of Equilibrium Set

Our approach allows us to trace out the whole utility possibilities frontier of the set of competitive equilibria. The Pareto frontier is only a subset of this frontier, and the set of utilities belonging to POPI plans is contained within the Pareto frontier. It is instructive to relate these sets in terms of the constraints of the optimal policy problem.

The frontier of the equilibrium set is found by, first, varying α in the Lagrangian from plus to minus infinity. We omit here the requirement to improve agents upon the status quo, this is trivial to find ex-post. For $\alpha \geq 0$ the solution is equivalent to solving for the Pareto optimal allocations imposing a minimum utility constraint for agent two, where \underline{U}^2 is the utility value computed from the solution of the Lagrangian. The points on the frontier that feature positive α constitute the Pareto frontier, since a positive α indicates that the minimum utility constraint is binding. For $\alpha < 0$ the solution of the Lagrangian corresponds to maximizing the first agent's utility and imposing \underline{U}^2 with *equality*. In this range the solution is not Pareto-optimal because, as is indicated by the negative Lagrange multiplier α , the first agent's utility could be increased by also increasing \underline{U}^2 . These points of the frontier would correspond to a welfare function where the planner would be willing to hurt agent 1 as long as agent 2 gets a sufficiently low utility. But in order to trace out the entire frontier we also have to switch agents one and two in the first line of the Lagrangian so that α multiplies the utility of agent one and then we have to vary α from zero to negative infinity again: this would correspond to points in the equilibrium frontier that are again *not* Pareto optimal and that are obtained by forcing the planner to give a certain utility to agent one.

Being a subset of the Pareto frontier, set of POPI plans clearly features positive α . However, note that non-optimal points on the feasible frontier, i.e. points where $\alpha < 0$ may also be Pareto-improving. If this is the case, it will be true for at least one agent that any POPI plan will strictly improve his utility. I.e. it is not possible to shift all the gains to the other agent in a Pareto optimal way. We will see an example

of this in section 4.4.4.

4.4 Results

In the following we present and discuss our numerical results, which are based on the parameter choices explained in the next subsection. We first describe the results for the baseline model of subsection 2.1, and then contrast them with those for the modified models as a way of gaining intuition for the forces at work.

4.4.1 Calibration

We calibrate the model to match several aspects of the status quo before the reform. All parameters except for the tax rates remain the same during the policy experiments. An overview of our parameter choices is provided in table 4.6.2. We evaluate the model at a yearly frequency.

Preferences: Agents have a CRRA-utility function over each consumption and leisure, ie. $u(c_{i,t}, l_{i,t}) = \frac{c_{i,t}^{1-\sigma_c}}{1-\sigma_c} + B \frac{(1-l_{i,t})^{1-\sigma_l}}{1-\sigma_l}$. Our choices for the risk aversion parameters σ_c and σ_l are standard. The same is true for the discount factor δ . The parameter B is chosen such that in a corresponding representative agent economy agents would work one third of their time in the steady state before the reform.

Heterogeneity: Our two types of agents are heterogeneous with respect to both their labor efficiency ϕ^j and their initial wealth $k_{j,-1}$. For simplicity we will from now on speak of "workers", indexed w , and "capitalists", indexed c . Capitalist are the group whose ratio of wealth to labor efficiency is higher. I.e. they are rich relative to their earnings potential. Note, however, that in absolute terms the capitalists are both richer AND more productive. In the status quo before the reform the heterogeneity parameters of table 4.6.2 translate into a relative consumption of the workers of $\lambda = .4$.

We base our choice of relative labor efficiency and wealth on the analysis of the

Panel Study of Income Dynamics performed in Garcia-Milà et al. (1995). They split their sample in five groups, while we have only two in order to facilitate computations. The degree of heterogeneity in our calibration is somewhat less than the difference between their two most extreme groups. Thus, we clearly understate heterogeneity. This is even more so as we are interested in the scope for Pareto improving tax reforms, which means we should not only take the poor and the rich as some group averages, but the poorest and the richest individual. This is important to keep in mind because it implies that our results must be regarded as lower bounds in the sense that in reality the constraint to improve everybody (or at least almost everybody) will be much tighter.

Production: We use a standard yearly calibration for technology. The production function is Cobb-Douglas with a capital income share of $\alpha^k = .36$. There is no productivity growth. The depreciation rate is $d = .08$. Initial capital is such that the corresponding representative agent economy would be in steady state before the reform.

Government: Before the reform the capital and labor income tax rates are 57% and 23% respectively. These are the average marginal tax rates calculated by McGrattan, Rogerson and Wright (1997) for the period 1947-87. Government spending per period g , is chosen to balance the budget intertemporally with these tax rates. It amounts to about 25% of output in the status quo. Note that the choice of tax rates in the status quo matters for two reasons. First of all, the capital tax rate influences the steady state (and hence initial) capital stock. Secondly, status quo utilities depend on the tax rates, and thus the scope for Pareto improvements.

We assume that during the reform the capital tax rate can never increase above its initial level. A justification for this assumption is political credibility: It should be hard for the government to convince agents that in the future capital taxes will be abolished if at the same time they are raised.

4.4.2 The set of Pareto Optimal-Pareto Improving Plans

Our goal is to explore the properties of tax policies that are not only Pareto optimal but, in addition, also improve all agents *vis a vis* the status quo described in section 4.4.1. Figure 11 traces out the welfare gains that each agent stands to reap from such POPI programs for the baseline model of section 4.2.1.¹³ Clearly, the gains are quite large, it is feasible to improve both agents significantly. Notice how the frontier of utilities is decreasing, meaning that we can find a Pareto optimal tax reform where the worker reaps the entire surplus of the reform while leaving the capitalist at his status quo utility (at the point where the solid line crosses the x-axis), or we can find a tax reform where only the capitalist enjoys all the gains (the solid line crossing the y-axis), or many tax reforms in between where both agents gain.¹⁴

As we vary the distribution of gains between the two agent types, many properties of the policies and allocation change. Figure 12 illustrates how several of them change with the welfare gain of the worker. First of all, the duration of the transition to zero capital taxation increases from about nine to nineteen years as we increase the welfare gain of the worker from zero to the maximum compatible with not hurting the capitalist (which is approximately 9%). Duration here is defined as the number of years with positive capital tax rates. As we know from the proof in section 4.3.1, capital taxes stay at their upper bound for all but the last period of the transition and then transit to zero with (at most) one intermittent period. A typical time path for capital taxes is drawn in figure 13. The fact that the duration of the transition is so sensitive to the distribution of the welfare gains is a reflection of the lack of redistributive instruments of the social planner: The worker contributes to the public coffers primarily through labor taxes, which means his burden stands to increase through the reform while the capitalist's burden decreases. This effect tends to distribute the efficiency gains that the reform permits asymmetrically. The earlier capital taxes

¹³In all the figures in this paper reporting results on welfare, the welfare gains for each agent are measured as the percentage, permanent increase in status quo consumption that would give the agent the same utility as in the optimal tax reform. Therefore, the origin of the graph represents the status quo utility, and the positive orthant contains Pareto improving allocations.

¹⁴We will see in section 4.4.4 that this is not generally true.

are suppressed, the more revenue has to be raised from labor taxes and the bigger the relative tax burden of the worker. This is why, by delaying the suppression of capital taxes, the welfare gains of the worker can be increased at the expense of the capitalist. The second panel in figure 12 further illustrates this mechanism: The more the worker gains the higher is the share of capital taxes in revenues in present value terms because they are suppressed only later.¹⁵

The final graph in figure 12 depicts α , the multiplier on the minimum utility constraint for the worker, and λ , the ratio of the worker's consumption to the capitalist's. We put these two graphs in the same picture because it can be shown that in a first best situation, i.e. if there is no distortionary taxation and no distributive conflict ($\Delta_1 = \Delta_2 = 0$) and hence the upper bound on capital taxes never binds ($\gamma_t = 0 \forall t$), and with logarithmic utility of consumption, we would have $\alpha = \lambda$.¹⁶ In the second best world of our model, by contrast, as we increase the welfare of the worker the marginal cost of doing so explodes, while his consumption share increases only mildly. In fact, it always remains very close to its value in the status quo, which is 0.4.

Another way of looking at α comes to mind by noting that it is the relative weight that agent 2 receives in the Lagrangean of the optimal program. This suggests that α is a measure of the bias of the social planner in favor of the workers. We do not favor this interpretation, however, because it invites to take the welfare function of the planner literally as a measure of what the planner should do. Keeping in mind Arrow's impossibility theorem, we think it is better to stay away from that interpretation, and this is why we confine our attention to studying POPI allocations, where the only unit of interest is the utility that each agent achieves through various tax reforms. Under this view, the weight α is not a measure of what a government should or should not do, it is just a Lagrange multiplier determined in equilibrium, it measures the cost of enforcing the minimum utility constraint. The fact that α has to increase so much to achieve a small redistribution is just a reflection of the difficulties that the planner finds in redistributing wealth from one agent to the other when only capital

¹⁵For comparison, the share of capital taxes in revenues is about .43 in the status quo.

¹⁶Generally, in the first best $\alpha = \lambda^{-\sigma_c}$.

or labor taxes are available.

4.4.3 The Time Path of the Economy under POPI Plans

To further describe the comparative dynamics of POPI tax reforms it helps to consider the time paths of capital, labor supply, the labor tax rate, and the government deficit that are pictured in figure 14. First note that qualitatively the paths are very similar across the set of POPI plans. The horizontal shifts in the graphs occur because under plans that shift the benefits to the worker capital taxes remain at their initial level for longer. The kinks in the paths of labor taxes and government deficit occur precisely in the intermediate period when capital taxes transit from their maximum to zero.

The paramount message from the set of graphs in figure 14 is that the optimal policy in our model goes against several traditional policy prescriptions that have been derived from studies of optimal dynamic taxation. We find the following:

- *Long transition.* It is known that with log utility of consumption the transition takes one period (capital taxes go to zero abruptly in period $t = 1$) in the standard model without tax limits. In our model the transition is very long. Part of this length is due to the tax limit (see section on MM1 model below) but as we explained in the description of Figure 12, the planner must lengthen the transition in order to insure that both agents gain, even more so in tax reforms more favorable to the worker.
- *Non-smooth taxes.* Both in capital and labor taxes there is an abrupt change at some point in the future. Capital taxes stay at the upper bound for a while and then go in one period to zero.¹⁷ Furthermore, labor taxes also change considerably over time.
- *It is optimal to reduce labor taxes initially.* Until now, the literature has concentrated on the long run results, which often imply very low capital taxes.

¹⁷This feature, for capital taxes, already was described in Atkeson et al. (1999) for a tax limit such that the after tax gross return is bounded below by $1 - d$.

But we find that the optimal policy is to reduce labor taxes initially, for some tax reforms even to a negative level, and then increase labor taxes above their status quo level. This, combined with the long transition, means that labor taxes often have to be low for a very long time.

What happens is the following: the planner wants to frontload capital taxes for the usual reasons that have been described at length in the literature. Therefore, it is optimal to have capital taxes at the upper limit in the first few periods and then let them go to zero. But in order to boost output and capital accumulation in the early periods, when capital taxes are still high, it is optimal to lower labor taxes to induce an increase in the return of capital. Figure 3 shows that labor supply of all agents is very high in the early periods.¹⁸ This engineers capital growth in the early periods when capital taxes are still at their high old level. Eventually the zero capital tax is the one promoting growth and helping the economy converge to the steady state where the golden rule holds. Absent this backloading of labor taxes early capital accumulation would take place only to the extent that the expectation of low future capital taxes raises incentives to save. This is much less, as will be confirmed in our analysis of model MM2 with fixed labor supply in section 4.4.4.

The fact that the short run policy is very different from the one applicable in the long run questions the validity of steady state analysis as a tool to understand what governments should do. In particular, the nowadays fashionable 'timeless perspective', which focuses on the analysis of optimal policies in the steady state, would give a recommendation that is the exact opposite from the optimal in the short run. This is the reason why Garcia-Milà et al. (1995) and related work found that implementing the long run policy with zero capital taxes immediately hurts poor agents (workers): It looks like it is not only the immediate suppression of capital taxes that is non-optimal, but also the high initial labor tax rates.

Furthermore, it would seem that the fact that the transition is so long implies that the government would need a very high degree of commitment in order to actually

¹⁸Capitalists, who also have higher labor productivity, always work less than workers.

achieve the long run zero capital tax.

A somewhat surprising pattern that emerges from the pictures is that the long run labor tax rate is higher the later capital taxes are abolished, i.e. the more the policy favors the worker. This may seem paradoxical because the worker is interested in low labor taxes. Note though, that even though the long run labor tax *rate* is higher if the worker is favored, the *share* of labor taxes in the present value of revenues is lower in this case, as figure 12 showed. This suggests that the long run labor tax rate is high for two reasons. First, when capital taxation is abandoned late the initial boost to capital accumulation comes mainly from extremely low initial labor taxes. I.e. the backloading of labor taxes is strongest in these cases. Second, the long run labor supply is lower the later capital taxes are suppressed, while the gross wage is always the same.¹⁹ Therefore, the revenue raised for a given labor tax rate is lower.

Since government expenditures are constant, the low initial labor taxes translate into government deficits. Only as labor taxes rise the government budget turns into surplus. Once capital taxes are suppressed and revenues fall again, the government deficit quickly reaches its long run value which can be positive or negative depending on whether during the transition the government accumulated wealth or not. We can see from figure 3 that most POPI policies imply that the government runs a primary surplus in the long run. This implies that the government is in debt in the long run, because the primary surplus is needed to pay the interest on debt. Therefore, for most tax reforms the low taxes in the initial periods generate a debt that is, in part, never repaid.

4.4.4 Isolating the Effects: Modified Models 1 and 2

As mentioned previously, in our baseline model there are two reasons not to abandon capital taxes too fast. One is the conventional one that the shortfall in government

¹⁹Since the long run real return on capital is determined by the rates of time preference and depreciation and the production function is Cobb-Douglas, the long run capital-labor ratio and wage are independent of the policy - as long as capital taxes are zero eventually.

revenues would have to be compensated through higher labor taxes, which are also distorting. So the tax authority has to balance off the two types of distortions. The second reason present in our model is the requirement to ensure a minimum utility level for both agents. In particular, abolishing capital taxes too fast would hurt the worker. In this section we want to look at each of these two factors separately.

Optimal Policies with Lump-Sum Transfers

By allowing for lump-sum transfers between the agents, Modified Model 1 gives the social planner an extra instrument to resolve the distributional conflicts, while retaining the problem of distortive labor taxes. Figure 15 traces out the set of Pareto optimal plans in terms of the welfare gains of each agent for MM1 (solid line) and, for comparison, for the baseline model (dashed line). Note that the POPI tax reforms are those that occupy the positive orthant, as they are the ones that imply Pareto improvement.

First of all note that the frontier of equilibria is decreasing, so the frontier of equilibria is also that of Pareto optimal allocations. Obviously the utility possibilities frontier in the baseline model is always dominated by that for the model with transfers. This indicates how the absence of transfers constrains the solution. It turns out that all POPI policies for the model MM1 imply a positive lump sum transfer from capitalists to workers, therefore, all POPI policies of the baseline model could be strictly improved upon if such a transfer took place. This transfer would be bigger the higher the required welfare gain of the worker. It is particularly interesting that the possibility of a transfer would allow to enhance the welfare gains even if the worker's welfare is to be kept at its status quo level (i.e. if the planner considered only tax reforms along the y-axis). In other words, the capitalist would be better off if he could make a payment to the worker and then choose a policy that leaves the worker indifferent to the situation before the reform. A zero transfer would be chosen by the policy maker only for a policy that harms the worker. This policy corresponds to the point in the north-western quadrant of figure 15 where the two lines just touch.

(If we move further up the line, i.e. benefit the capitalist even more at the expense of the worker, a transfer from worker to capitalist becomes optimal.)

How the possibility of a transfer takes care of the distributional component of our problem can also be seen from other properties of POPI plans with optimal transfers. As a benchmark, recall that in figure 12 all variables with the exception of λ change strongly as the distribution of the gains from reform changes. This is very different when optimal transfers are available. The graphs of duration and share of capital taxes in revenues analogous to the ones in Figure 2 would be almost flat: capital taxes are always suppressed after 5-6 years, the share of capital taxes is always 0.11. α , the multiplier on the worker's utility constraint, would increase only slowly with $\Pi(\textit{worker})$, while λ would rise much more than without the transfer. This pattern illustrates that the policies and the path of the economy would hardly depend on the distribution of the gains from reform. On the other hand, shifting welfare gains and consumption between agents would be much easier, as indicated by the behavior of α and λ .

Optimal Policies with a Fixed Labor Supply

Modified Model 2 features a fixed labor input, such that labor taxes are no longer distortive. What remains is the need to Pareto improve upon the status quo. Figure 16 traces out a part of the frontier of feasible equilibria for MM2 (solid line) as well as the corresponding frontier for a chimera of MM1 and MM2 (dashed line), ie. the model without labor AND with an optimal transfer, for comparison. This latter graph actually corresponds to the first best because neither distributional concerns nor inefficient taxes constrain the social planner's problem and hence capital taxes can be abolished immediately without any drawbacks. Clearly, MM2 gives rise to solutions that are far away from the first best, no matter how the gains are distributed between the agents. Correspondingly, it takes 10 to 20 years until the suppression of capital taxes. This shows that the need to improve all agents by itself imposes a severe constraint on the planner's problem, even when distortions of the labor supply

are not an issue.

Another interesting aspect of figure 16 is the shape of the frontier of feasible equilibria in MM2. It bends backward, which means not the entire frontier actually belongs to the set of Pareto optimal plans: The increasing part of the solid line contains equilibria that are at the frontier of the feasible equilibrium set but that are not Pareto optimal, since the point $\alpha_c = 0$ Pareto-dominates all these equilibria. The way we generate this frontier is by imposing a minimum utility constraint for the capitalist with equality.²⁰ Those plans that lie on the (non-optimal) backward bending part of the frontier go along with a negative multiplier on the utility constraint ($\alpha_c < 0$). I.e. in these cases the planner can only force the capitalist onto a certain (low) utility level by also harming the worker. Note, however, that this non-optimal part of the frontier still satisfies the requirement of Pareto superiority *vis a vis* the status quo. Notice also that this implies that in this model it is not possible to distribute all the gains from a Pareto optimal reform to the worker. It is perfectly feasible, on the other hand, to keep the worker at status quo and make only the capitalist benefit.²¹

MM2 also helps to illustrate the role of the path of labor taxes in the baseline economy. Recall that in section 4.4.3 we argued that it is important that the planner can initially boost labor supply and thus capital accumulation by lowering labor taxes while capital taxes are still high. Here in MM2 this option does not exist since labor supply is fixed and in fact the time path of labor taxes is indeterminate.²² Correspondingly, capital increases in the initial periods only to the extent that the expectation of low future capital taxes induces saving. Figure 17 illustrates that the resulting capital accumulation is much slower initially. We have plotted two illustrative time paths for capital. The solid line depicts the evolution of the capital stock for the fastest possible transition to zero capital taxes, namely for the plan

²⁰Details are given in section 4.3.3. Note that while before we were maximizing the utility of the capitalist subject to a min. utility constraint for the worker, now we are maximizing for the worker subject to a constraint on the capitalist's utility. To make this plain we now index the multiplier with "c".

²¹It must be expected that the same phenomenon occurs in the model with a labor choice if σ_l is sufficiently high. After all, in the limit as $\sigma_l \rightarrow \infty$ the labor supply becomes fixed.

²²All that matters is the total discounted sum of labor taxes paid.

that just leaves the worker at his status quo utility. The dashed line corresponds to the slowest possible POPI transition, i.e. the one that maximizes the utility gain of the workers and has $\alpha_c = 0$. All other paths belonging to POPI plans would be in between. Comparison with the first panel in figure 14 shows clearly that in the model without labor capital rises much less in the first periods while capital taxes are still in place. Thus the backloading of labor taxes in the baseline model seems to be an important part of the optimal policy.

4.5 Conclusion

Several recent papers suggest that the abolition of capital taxes - desirable as it may be for efficiency reasons - will harm large groups of agents, unless it is done very carefully in terms of timing. In this paper we respond to this observation by studying the optimal time path of capital and labor taxes in a model of heterogeneous agents, subject to the constraint that nobody lose from the reform. For reasons of political credibility we moreover impose an upper bound on capital tax rates corresponding to their level before the reform. The Pareto optimal and Pareto improving plans that solve our policy problem have very interesting properties. The time path of tax rates is highly non-smooth. Capital tax rates remain at their upper bound for at least 10 years and then drop to zero within two periods. Labor tax rates, by contrast, are initially very low, often even negative, and rise to their new long run level around the time when capital taxes are suppressed. As a consequence, the government typically accumulates debt that is never repaid. These time paths suggest that 1) a Pareto improving abolition of capital taxes requires a lot of credibility on part of the government because the transition is so long, and 2) the nowadays popular "timeless perspective", i.e. the exclusive focus on the steady state, is very misleading in our case. It would mean to miss all the dynamics that are necessary in order to distribute the gains from the reform broadly.

4.6 Appendix

4.6.1 The maximization problem and first order conditions

Using the derivations in section 4.2.1, the maximization problem to be solved becomes

$$\max_{\lambda, \{c_t^1, k_t, l_t^1\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \delta^t [u(c_{1,t}) + v(l_{1,t})] \quad (4.24)$$

$$\text{s.t.} \quad u'(c_{1,t}) \geq \delta u'(c_{1,t+1})(1 + (r_{t+1} - d)(1 - \tilde{\tau})) \quad \text{for all } t \quad (4.25)$$

$$\frac{1 + \lambda}{2} c_{1,t} + g + k_t - (1 - d)k_{t-1} = F\left(k_{t-1}, \frac{l_{1,t} + f(\lambda, l_{1,t})}{2}\right) \quad \text{for all } t \quad (4.26)$$

$$\sum_{t=0}^{\infty} \delta^t [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))] \geq \underline{U}^2 \quad (4.27)$$

$$\sum_{t=0}^{\infty} \delta^t (u'(c_{1,t})c_{1,t} + v'(l_{1,t})l_{1,t}) = u'(c_{1,0})k_{1,-1}(1 + (r_0 - d)(1 - \tau_0^k)) \quad (4.28)$$

$$\sum_{t=0}^{\infty} \delta^t \left(u'(c_{1,t})\lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t})f(\lambda, l_{1,t}) \right) = u'(c_{1,0})k_{2,-1}(1 + (r_0 - d)(1 - \tau_0^k)) \quad (4.29)$$

letting $\alpha, \Delta_1, \Delta_2$ be the lagrange multipliers for the constraints involving discounted sums (4.27), (4.28) and (4.29), the Lagrangian is given by (4.14).

The first order conditions for the Lagrangian are:

- for consumption, $t > 0$:

$$\begin{aligned} u'(c_{1,t}) + \alpha \lambda u'(\lambda c_{1,t}) + (\Delta_1 + \lambda \Delta_2)[u'(c_{1,t}) + u''(c_{1,t})c_{1,t}] + \\ \gamma_t u''(c_{1,t}) - \delta \gamma_{t-1} u''(c_{1,t})(1 + (r_t - d)(1 - \tilde{\tau})) = \mu_t \frac{1}{2}(1 + \lambda) \end{aligned}$$

- for consumption, $t = 0$:

$$\begin{aligned}
& u'(c_{1,0}) + \alpha \lambda u'(\lambda c_{1,0}) + (\Delta_1 + \lambda \Delta_2)[u'(c_{1,0}) + u''(c_{1,0})c_{1,0}] - \\
& u''(c_{1,0})((\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1})(1 + (r_0 - d)(1 - \tau_0^k)) + \\
& \Delta_1 m_{1,1} + \Delta_2 m_{2,-1}) + \gamma_0 u''(c_{1,0}) = \mu_0 \frac{1 + \lambda}{2}
\end{aligned}$$

- for labor, $t > 0$:

$$\begin{aligned}
& v'(l_{1,t}) + \alpha v'(f(\lambda, l_{1,t}))f'(\lambda, l_{1,t}) + \\
& \Delta_1[v'(l_{1,t}) + v''(l_{1,t})l_{1,t}] + \Delta_2 \frac{\phi_2}{\phi_1}[v'(l_{1,t})f'(\lambda, l_{1,t}) + v''(l_{1,t})f(\lambda, l_{1,t})] - \\
& \gamma_{t-1} u'(c_{1,t})(1 - \tilde{\tau})F_{k,e}(k_{t-1}, e_t) \frac{1}{2}(\phi_1 + \phi_2 f'(\lambda, l_{1,t})) = \\
& -F_e(k_{t-1}, e_t) \frac{1}{2}(\phi_1 + \phi_2 f'(\lambda, l_{1,t}))\mu_t
\end{aligned}$$

- for labor, $t = 0$:

$$\begin{aligned}
& v'(l_{1,0}) + \alpha v'(f(\lambda, l_{1,0}))f'(\lambda, l_{1,0}) + \\
& \Delta_1[v'(l_{1,0}) + v''(l_{1,0})l_{1,0}] + \Delta_2 \frac{\phi_2}{\phi_1}[v'(l_{1,0})f'(\lambda, l_{1,0}) + v''(l_{1,0})f(\lambda, l_{1,0})] - \\
& (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1})u'(c_{1,0})F_{k,e}(k_{t-1}, e_t) \frac{1}{2}(\phi_1 + \phi_2 f'(\lambda, l_{1,t}))(1 - \tau_0^k) = \\
& -F_e(k_{t-1}, e_t) \frac{1}{2}(\phi_1 + \phi_2 f'(\lambda, l_{1,t}))\mu_0
\end{aligned}$$

- for capital, $t \geq 0$:

$$\mu_t + \gamma_t \delta u'(c_{1,t+1})(1 - \tilde{\tau})F_{k,k}(k_t, e_{t+1}) = \delta \mu_{t+1}(1 + F_k(k_t, e_{t+1}) - d)$$

- for λ :

$$\begin{aligned} \sum_{t=0}^{\infty} \delta^t & \left[\alpha (u'(\lambda c_{1,t})c_{1,t} + v'(f(\lambda, l_{1,t}))f_{\lambda}(\lambda, l_{1,t})) + \right. \\ & \Delta_2(u'(c_{1,t})\lambda c_{1,t} + \frac{\phi_2}{\phi_1}v'(l_{1,t})f_{\lambda}(\lambda, l_{1,t})) - \\ & \left. \mu_t \frac{1}{2}(c_{1,t} - F_e(k_{t-1}, e_t)\phi_2 f_{\lambda}(\lambda, l_{1,t})) \right] = 0 \end{aligned}$$

4.6.2 Computational strategy: Approximation of the time path

1. Fix T as the number of periods after which the steady-state is assumed to have been reached. (We use $T = 150$.)
2. Propose a $3*T+3$ -dimensional vector $X = \{k_0, \dots, k_{T-1}, l_0, \dots, l_{T-1}, \gamma_0, \dots, \gamma_{T-1}, \Delta_1, \Delta_2, \lambda\}$. (This is not the minimal number of variables to be solved for as a fixed point problems. $2 * T + 3$ would be sufficient, however, convergence is better if the approximation errors are spread over a larger number of variables.)
3. With k_{-1} and g known, find $\{c_t, F_{kt}, F_{lt}, F_{klt}, F_{kkt}\}$ from the resource constraint and the production function.
4. Calculate $\{\mu_t\}$ from the FOC for labor.
5. Calculate $\{\gamma_t\}$ from the FOC for consumption, making use of $\{\mu_t\}$ and the guess for $\{\gamma_t\}$ from the X-vector. (The guess is plugged into γ_{t-1} , γ_t is backed out.)
6. Form the $3 * T + 3$ residual equations to be set to 0:
 - The FOC for capital (Euler equation) has to be satisfied. (T equations)
 - The vector $\{\gamma_t\}$ has to converge, ie old and new guess have to be equal. (T equations)
 - Check for each period whether the constraint on τ^k is satisfied. If yes, impose $\gamma_t = 0$. Otherwise, the constraint on capital taxes has to be satisfied with equality. (T equations)
 - The remaining 3 equations come from the present value budget constraints (PVBC) and the FOC for λ . The discounted sums in the PVBCs are calculated using the time path of the variables for the first T periods and adding the net present value of staying in steady-state forever thereafter.

7. Iterate on X to set the residuals to 0. (We use Broydn's algorithm to solve this $3 * T + 3$ -dimensional fixed point problem.)

Bibliography

- Abel, A. (1990), ‘Asset prices under habit formation and catching up with the joneses’, *American Economic Review* **80**(1), 38–42.
- Abel, A. (1999), ‘Risk premia and term premia in general equilibrium’, *Journal of Monetary Economics* **43**(1), 3–33.
- Acemoglu, D. and Shimer, R. (1999), ‘Efficient unemployment insurance’, *Journal of Political Economy* **107**(5), 893–928.
- Adam, K., Marcet, A. and Nicolini, J. (2006), Learning and stock market volatility, *Computing in Economics and Finance* 2006 15, Society for Computational Economics.
- Aiyagari, S. (1994), ‘Uninsured idiosyncratic risk and aggregate saving’, *Quarterly Journal of Economics* pp. 659–684.
- Atkeson, A., Chari, V. and Kehoe, P. (1999), ‘Taxing capital income: A bad idea’, *Federal Reserve Bank of Minneapolis Quarterly Review* **23**(3), 3–17.
- Bansal, R. and Yaron, A. (2004), ‘Risks for the long run: A potential resolution for asset pricing puzzles’, *Journal of Finance* **59**(4), 1481–1509.
- Bedau, K. (1999), ‘Entwicklung der einkommenspositionen von haushalten in deutschland’, *DIW-Wochenbericht* (3).

- Bedau, K. and Teichmann, D. (1995), ‘Die belastung der privaten haushalte in west- und ostdeutschland durch direkte und indirekte steuern’, *DIW-Wochenbericht* (46).
- Bundesregierung (2001), ‘Lebenslagen in deutschland: Der erste armuts- und reich- tumsbericht der bundesregierung’.
- Campbell, J. (2003), Consumption-based asset pricing, in G. Constantinides, M. Har- ris and R. M. Stulz, eds, ‘Handbook of the Economics of Finance’, Vol. 1, Else- vier, chapter 13, pp. 803–887.
- Campbell, J. and Cochrane, J. (1999), ‘By force of habit: A consumption-based explanations of aggregate stock market behavior’, *Journal of Political Economy* **107**(2), 205–251.
- Campbell, J. and Shiller, R. (1988a), ‘The dividend-price ratio and expectations of future dividends and discount factors’, *Review of Financial Studies* **1**, 195–228.
- Campbell, J. and Shiller, R. (1988b), ‘Stock prices, earnings, and expected dividends’, *The Journal of Finance* **43**(3), 661–676.
- Campbell, J. and Viceira, L. (2002), *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press.
- Carroll, C. and Samwick, A. (1997), ‘The nature of precautionary wealth’, *Journal of Monetary Economics* **40**, 41–71.
- Chamley, C. (1986), ‘Optimal taxation of capital income in general equilibrium with infinite lives’, *Econometrica* **54**, 607–622.
- Chan, Y. and Kogan, L. (2002), ‘Catching up with the joneses: Heterogeneous preferences and the dynamics of asset prices’, *Journal of Political Economy* **110**(6), 1255–1285.

- Chari, V., Christiano, L. and Kehoe, P. (1994), ‘Optimal fiscal policy in a business cycle model’, *Journal of Political Economy* **102**(4), 617–52.
- Chari, V. and Kehoe, P. (n.d.), Optimal fiscal and monetary policy, *in* ‘Handbook of Macroeconomics’.
- Cochrane, J. (2005), *Asset Pricing*, 2nd edn, Princeton University Press.
- Cochrane, J. (2006), Financial markets and the real economy. unpublished.
- Coen-Pirani, D. (2004), ‘Effects of differences in risk aversion on the distribution of wealth’, *Macroeconomic Dynamics* **8**, 617–632.
- Coen-Pirani, D. (2005), ‘Margin requirements and equilibrium asset prices’, *Journal of Monetary Economics* **52**, 449–475.
- Cogley, T. and Sargent, T. (2005), The market price of risk and the equity premium: A legacy of the great depression? unpublished.
- Conesa, J. and Krueger, D. (2006), ‘On the optimal progressivity of the income tax code’, *Journal of Monetary Economics* **53**(7), 1425–1450.
- Constantinides, G. (1990), ‘Habit formation: A resolution of the equity premium puzzle’, *Journal of Political Economy* **98**, 519–43.
- Constantinides, G. and Duffie, D. (1996), ‘Asset pricing with heterogeneous consumers’, *Journal of Political Economy* **104**(2), 219–240.
- Correia, M. I. H. (1995), Efficiency and equity: Is there a trade-off?, CEPR Discussion Papers 1218, C.E.P.R. Discussion Papers.
- Costain, J. (1997), ‘Unemployment insurance with endogenous search intensity and precautionary saving’, UPF Working Paper.
- Costain, J. and Reiter, M. (2003), Business cycles, unemployment insurance, and the calibration of matching models. unpublished.

- Domeij, D. and Heathcote, J. (2004), ‘On the distributional effects of reducing capital taxes’, *International Economic Review* **45**(2), 523–554.
- Epstein, L. and Zin, S. (1989), ‘Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework’, *Econometrica* **57**, 937–969.
- Fama, E. and French, K. (1988), ‘Dividend yields and expected stock returns’, *Journal of Financial Economics* **22**, 3–25.
- Fama, E. and French, K. (1989), ‘Business conditions and expected returns on stocks and bonds’, *Journal of Financial Economics* **25**, 23–49.
- Faraglia, E., Marcet, A. and Scott, A. (2006), In search of a theory of debt management. working paper.
- Flodén, M. (2006), Why are capital income taxes so high?, Working Paper Series in Economics and Finance 623, Stockholm School of Economics.
- Garcia-Milà, T., Marcet, A. and Ventura, E. (1995), Supply side interventions and redistribution, Economics Working Papers 115, Department of Economics and Business, Universitat Pompeu Fabra.
- Hall, R. (1988), ‘Intertemporal substitution in consumption’, *Journal of Political Economy* **96**, 339–357.
- Heaton, J. and Lucas, D. (1996), ‘Evaluating the effects of incomplete markets on risk sharing and asset pricing’, *Journal of Political Economy* **104**(3), 443–487.
- Hopenhayn, H. and Nicolini, J. (1997), ‘Optimal unemployment insurance’, *Journal of Political Economy* **105**(2), 412–438.
- Hosios, A. (1990), ‘On the efficiency of matching and related models of search and unemployment’, *Review of Economic Studies* **57**, 279–98.

- Joseph, G. and Weitzman, T. (2003), ‘Optimal unemployment insurance: transitional dynamics vs. steady state’, *Review of Economic Dynamics* **6**(4), 869–884.
- Judd, K. (1985), ‘Redistributive taxation in a simple perfect foresight model’, *Journal of Public Economics* **28**, 59–83.
- Krusell, P. and Smith, A. (1998), ‘Income and wealth heterogeneity in the macroeconomy’, *Journal of Political Economy* **106**(5), 867–896.
- Lentz, R. (2003), Optimal unemployment insurance in an estimated job search model with savings. unpublished.
- Lucas, R. (1990), ‘Supply side economics: An analytical review’, *Oxford Economic Papers* **42**, 293–316.
- Machin, S. and Manning, A. (1999), The causes and consequences of longterm unemployment in europe, in O. Ashenfelter and D. Card, eds, ‘Handbook of Labor Economics’, Vol. 3c, Elsevier, chapter 47, pp. 3085–3139.
- Marcet, A. and Marimon, R. (1998), Recursive contracts, Technical report.
- Marcet, A. and Singleton, K. (1999), ‘Equilibrium asset prices and savings of heterogeneous agents in the presence of incomplete markets and portfolio constraints’, *Macroeconomic Dynamics* **3**, 243–277.
- Marimon, R. and Zilibotti, F. (1999), ‘Unemployment vs. mismatch of talents: Reconsidering unemployment benefits’, *Economic Journal* **109**(455), 266–291.
- Martin, J. (1996), ‘Measures of replacement rates for the purpose of international comparisons: a note’, *OECD Economic Studies* **26**(1), 98–118.
- McGrattan, E., Rogerson, R. and Wright, R. (1997), ‘An equilibrium model of the business cycle with household production and fiscal policy’, *International Economic Review* **38**(2), 267–90.

- Mehra, R. and Prescott, E. (1985*a*), ‘The equity premium: A puzzle’, *Journal of Monetary Economics* **15**, 145–161.
- Mehra, R. and Prescott, E. (1985*b*), ‘The equity premium: A puzzle’, *Journal of Monetary Economics* **15**(2), 145–162.
- Mnnich, M. (2001), ‘Einkommens- und geldvermögensverteilung privater haushalte in deutschland - teil 2. ergebnis der einkommens- und verbrauchsstichprobe 1998’, *Statistisches Bundesamt, Wirtschaft und Statistik* (2), 121–137.
- Nickell, S. and Nunziata, L. (2001), ‘Labor market institutions database’.
- Petrongolo, B. and Pissarides, C. (2001), ‘Looking into the black box: A survey of the matching function’, *Journal of Economic Literature* **39**, 390–431.
- Pissarides, C. (1990), *Equilibrium Unemployment Theory*, MIT Press, Cambridge, MA.
- Rebelo, S., Gomes, J. and Greenwood, J. (2003), ‘Equilibrium unemployment’, *Journal of Monetary Economics* **48**(1), 109–152.
- Shavell, S. and Weiss, L. (1979), ‘The optimal payment of unemployment insurance benefits over time’, *Journal of Political Economy* **87**(6), 1347–62.
- Shimer, R. (2002), ‘The cyclical behavior of equilibrium unemployment and wages: Evidence and theory’, *NBER Working Papers* (9536).
- Telmer, C. (1993), ‘Asset-pricing puzzles and incomplete markets’, *The Journal of Finance* **48**(5), 1803–1832.
- Vissing-Jorgensen, A. (2002), ‘Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures’, NBER Working Paper 8884.

Wang, C. and Williamson, S. (1996), ‘Unemployment insurance with moral hazard in a dynamic economy’, *Carnegie-Rochester Conference Series on Public Policy* **44**, 1–41.

Weil, P. (1989), ‘The equity premium puzzle and the risk-free rate puzzle’, *Journal of Monetary Economics* **24**(3), 401–421.

Weitzman, M. (2007), ‘Prior-sensitive expectations and asset-return puzzles’, *American Economic Review* . Forthcoming.

Tables and Figures

Tables for Chapter 2

PARAMETER	VARIABLE	VALUE	
		Benchmark calibration	Dividend calibration
Mean consumption growth	μ	1.7%p.a.	1.25%p.a.
Std. dev. of consumption growth	σ	3.3 % p.a.	12.8 % p.a.
Discount factor	δ	.96	.96
Intertemporal elasticity of substitution	ψ	1.5	1.5
Average risk aversion	$\bar{\gamma}$	42.74	3.31
L-type's risk aversion*	γ_L	36	2.335
H-type's risk aversion*	γ_H	97.51	5.683

*These risk aversion levels translate into the above average risk aversion if on average the L-type owns 75% of the wealth.

Table 1: Benchmark Parameter Values

STATISTIC	DATA	MODEL		
		$w_{L,0} = 0.1$	$w_{L,0} = 0.2$	$w_{L,0} = 0.3$
$E(\Delta c)$	1.7	1.7	1.7	1.7
$\sigma(\Delta c)$	3.3	3.3	3.3	3.3
$E(r^f)$	1.7	1.3	1.4	1.5
$\sigma(r^f)$	5.1	0.9	0.7	0.5
$E(r - r^f)$	4.6	4.9	4.7	4.6
$\sigma(r - r^f)$	18.5	3.6	3.5	3.5
$E(PD)$	22.7	23.4	23.5	23.6
$\sigma(PD)$	6.2	1.0	0.8	0.7
$\rho(PD, PD_{-1})$	0.81	0.85	0.78	0.70

Note: The model is simulated at a quarterly frequency. Statistics are calculated from time-averaged data at an annual frequency. All returns are annual percentages. Small letter are logs.

Table 2: Unconditional statistics of simulated and historical data

HORIZON (Years)	DATA		MODEL					
			$w_{L,0} = 0.1$		$w_{L,0} = 0.2$		$w_{L,0} = 0.3$	
	10 x Coefficient	R^2	10 x Coefficient	R^2	10 x Coefficient	R^2	10 x Coefficient	R^2
1	-1.65	.06	-2.77	.11	-2.51	.06	-2.23	.04
2	-3.43	.11	-5.41	.20	-4.93	.12	-4.40	.07
3	-3.92	.13	-7.90	.27	-7.22	.17	-6.47	.11
5	-7.91	.35	-12.40	.38	-11.39	.26	-10.29	.17
7	-7.92	.20	-16.58	.46	-15.39	.33	-14.13	.22
10	-13.19	.59	-21.69	.55	-20.29	.41	-18.88	.29

Table 3: Long-horizon return regressions

VOLATILITIES			CORRELATIONS		
$\sigma(\Delta c_L)$	$\sigma(\Delta c_H)$	$\sigma(\Delta c)$	$\rho(\Delta c_L, \Delta c)$	$\rho(\Delta c_H, \Delta c)$	$\rho(\Delta c_L, \Delta c_H)$
6.7%	1.2%	3.3%	.57	.96	.59

Table 4: Volatilities and correlations of individual and aggregate consumption growth ($w_{L,0} = 0.1$)

MEANS AND STD. DEV.'S			LONG-HORIZON REGRESSIONS				
STATISTIC	DATA	MODEL	HORIZON (Years)	DATA		MODEL	
				10 x Coeff.	R^2	10 x Coeff.	R^2
$E(r^f)$	1.7	1.1	1	-1.65	.06	-1.32	.006
$\sigma(r^f)$	5.1	1.0	2	-3.43	.11	-2.50	.011
$E(r - r^f)$	4.6	4.8	3	-3.92	.13	-3.51	.015
$\sigma(r - r^f)$	18.5	13.3	5	-7.91	.35	-5.86	.024
$E(PD)$	22.7	23.6	7	-7.92	.20	-8.75	.039
$\sigma(PD)$	6.2	1.8	10	-13.19	.59	-12.04	.052

Note: The model is simulated starting at $w_{L,0} = 0.1$.

Table 5: Means, standard deviations, and long horizon return regressions for the calibration to dividend data

Tables for Chapter 3

γ	2
β	0.9955 ($\hat{=}$ 0.9479 p.a.)
r	0.0017 ($\hat{=}$ 0.02 p.a.)
y	1
κ	2.6534
λ	0.0072
χ	0.254
η	-0.5
σ	0.5

Table 6: Parameters

new rate	status	static	dynamic	transition cost
59%	unemployed	0.04%	-0.01%	0.05%
	employed	0.12%	0.06%	0.06%
50%	unemployed	0.16%	-0.48%	0.64%
	employed	0.98%	0.29%	0.69%
40%	unemployed	-0.01%	-1.73%	1.72%
	employed	1.63%	0.03%	1.60%
30%	unemployed	-0.34%	-3.88%	4.21%
	employed	2.10%	-0.59%	2.69%

Table 7: Welfare effects of lowering the replacement rate (old level: 60%) without announcement

new rate	agent type	static	dynamic, announcement		
			none	1 yr	2 yrs
40%	unemployed	-0.01%	-1.73%	0.35%	0.71%
	employed	1.63%	0.03%	0.63%	0.75%
	firms	14.32%	35.74%	8.54%	0.78%

Table 8: Welfare effects of reducing the replacement rate from 60% to 40% depending on the length of the announcement period

Tables for Chapter 4

<i>Preference parameters</i>	γ_c	-1
	γ_l	-3
	B	.76
	δ	.96
<i>Heterogeneity Parameters</i>	ϕ_c/ϕ_w	1.05
	$k_{c,-1}$	5.49
	$k_{w,-1}$	-3.47
<i>Production parameters</i>	α^k	.36
	d	.08
	k_{-1}	1.01
<i>Government spending</i>	g	.13
<i>Tax rates before reform</i>	τ^l	.23
	τ^k	.57
<i>Upper bound on cap. tax rate</i>	$\tilde{\tau}$.57

Table 9: Parameter Values of the Baseline Economy.

Figures for Chapter 2

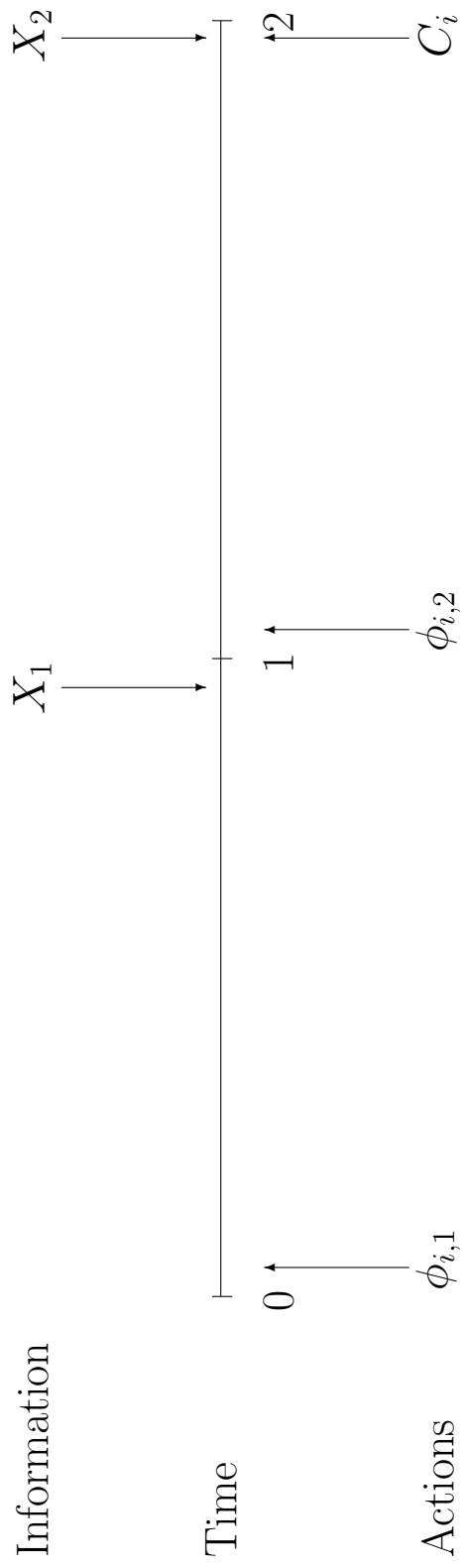


Figure 1: The timing of events in the two-period model

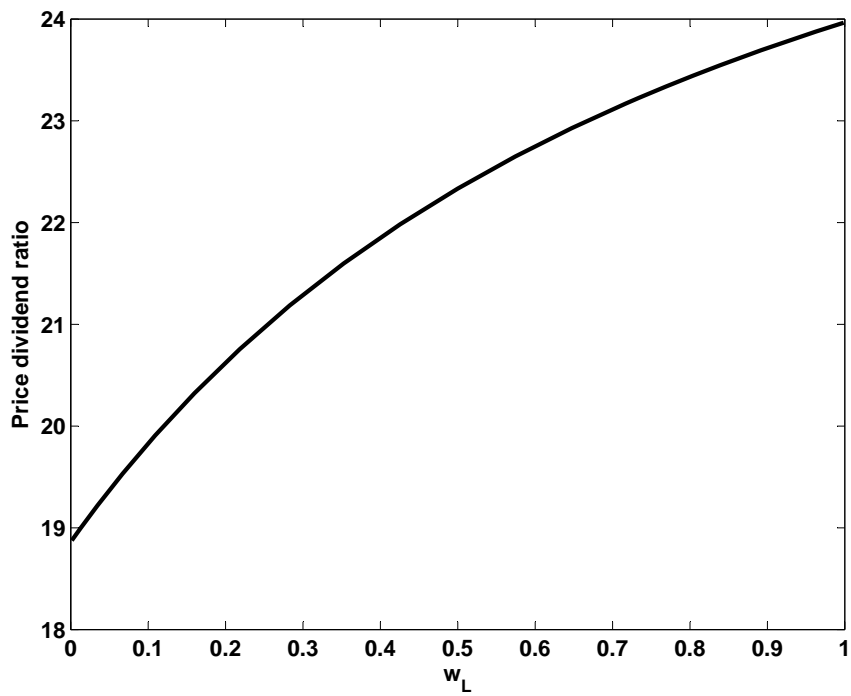
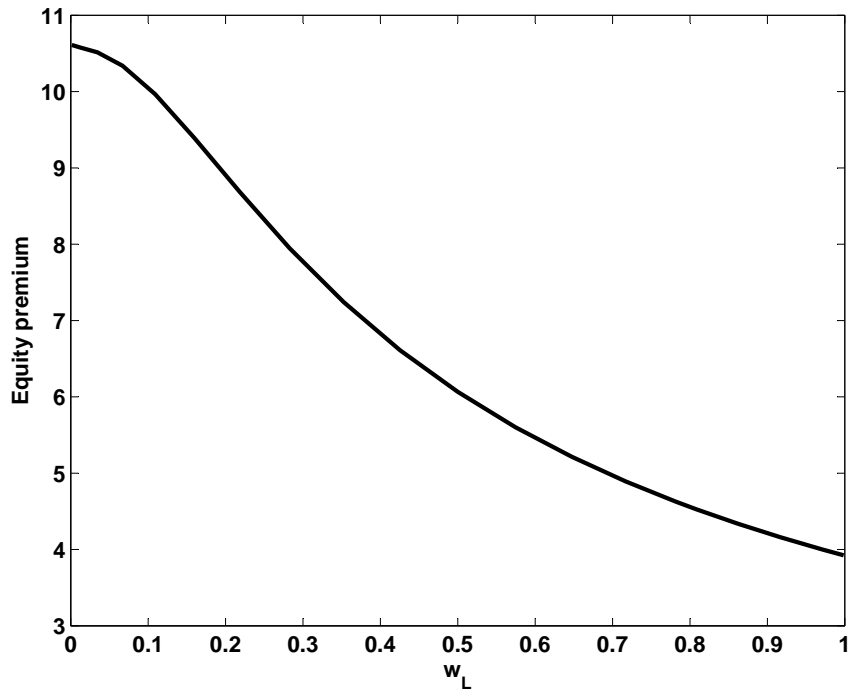


Figure 2: Equity premium (in % p.a.) and price-dividend ratio as functions of the wealth distribution

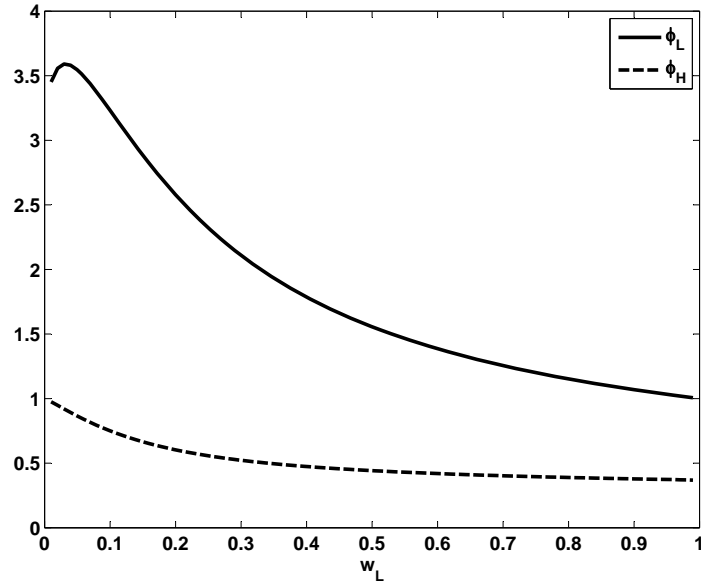


Figure 3: The share of the stock in each agent type's portfolio as a function of the wealth distribution

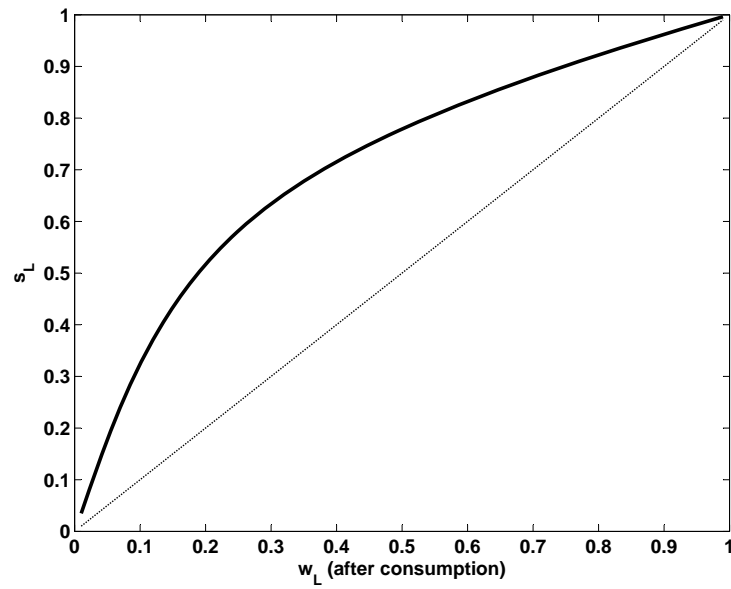


Figure 4: Stock holdings of the L-type as a function of his wealth share. (The dotted line represents $s_L = w_L$, i.e. the counterfactual of no asset trade.)

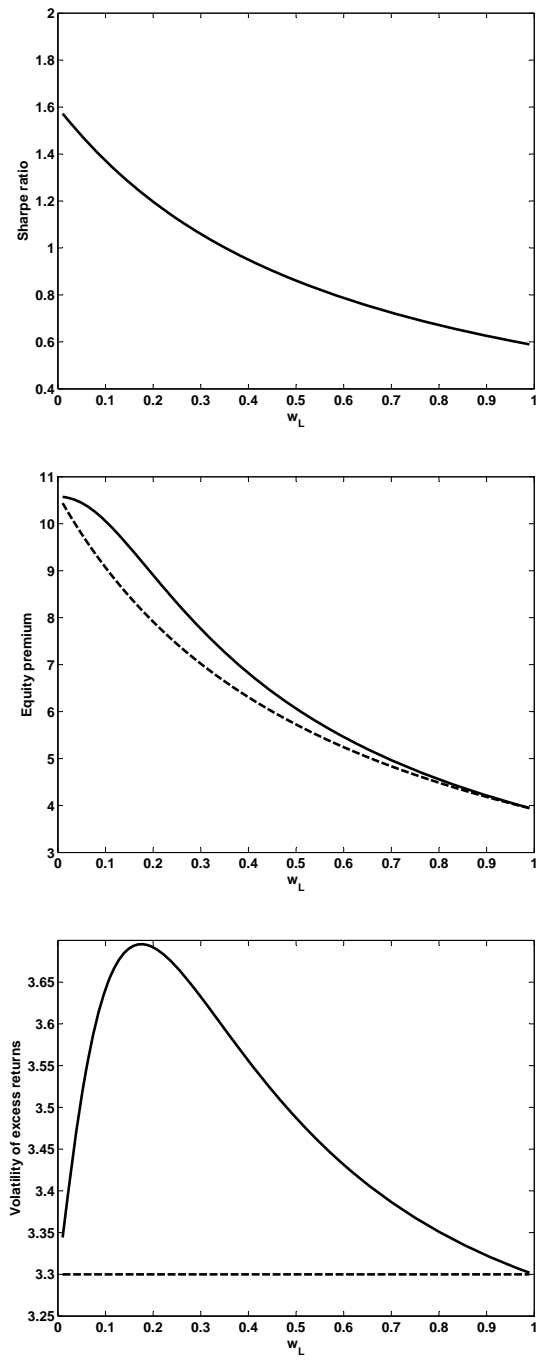


Figure 5: Sharpe ratio, equity premium (in % p.a.), and volatility of excess returns as functions of the wealth distribution
 Solid lines: heterogeneous agent model
 Dashed lines: emulation in the representative agent model

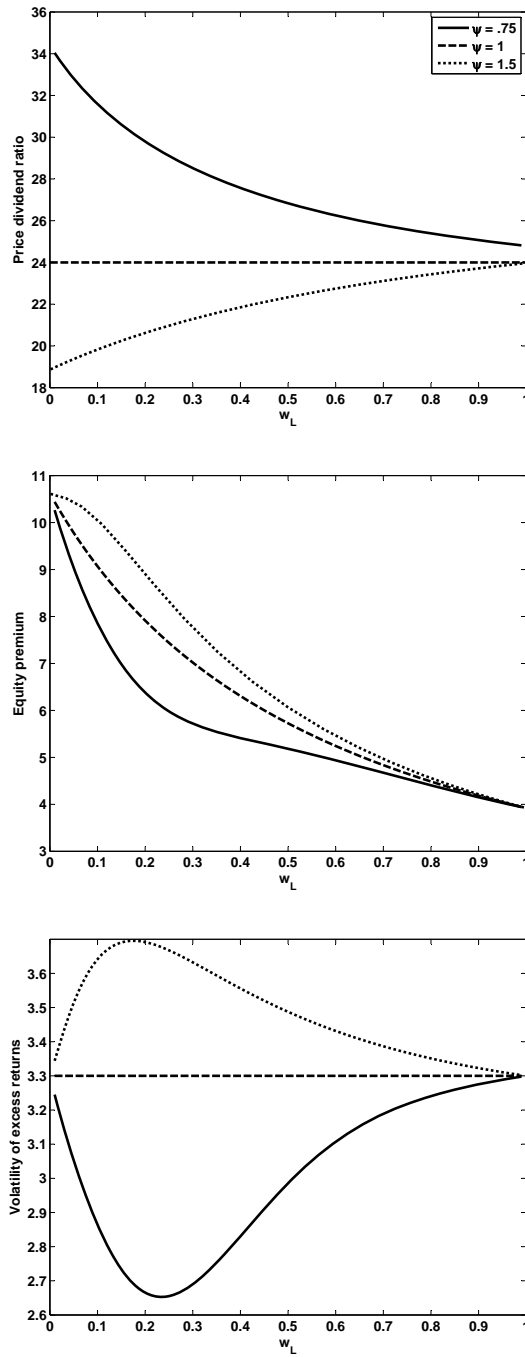


Figure 6: The Role of the intertemporal elasticity of substitution

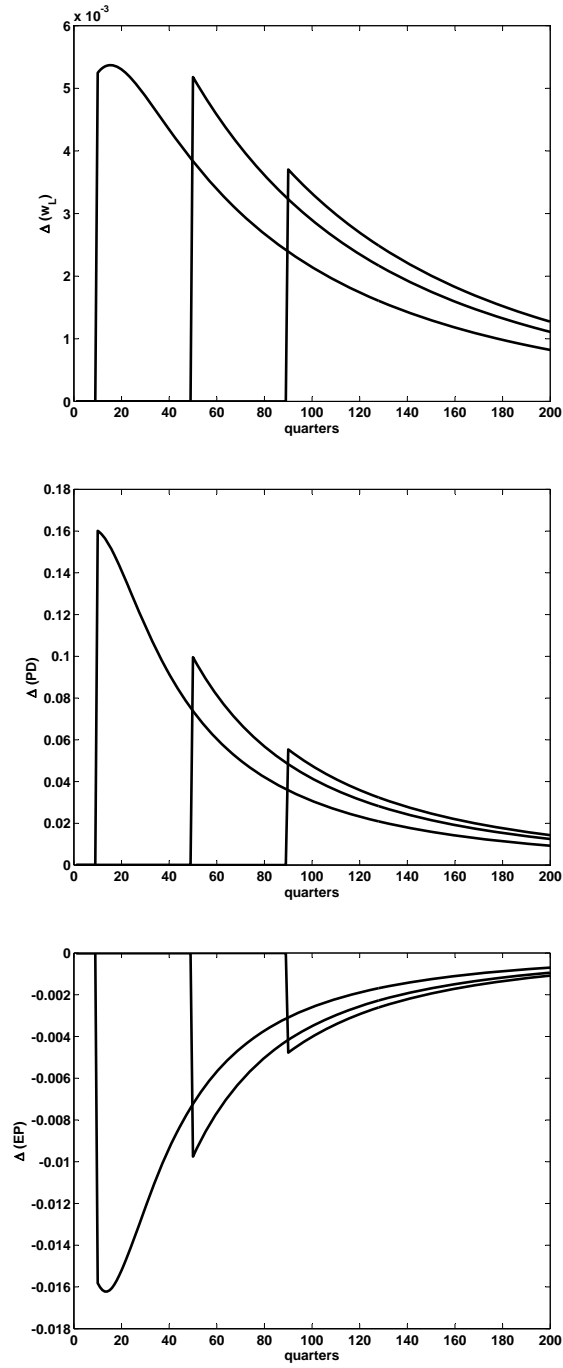


Figure 7: Impulse responses: Difference between simulation and without a dividend growth shock of one standard deviation.

Figures for Chapter 3

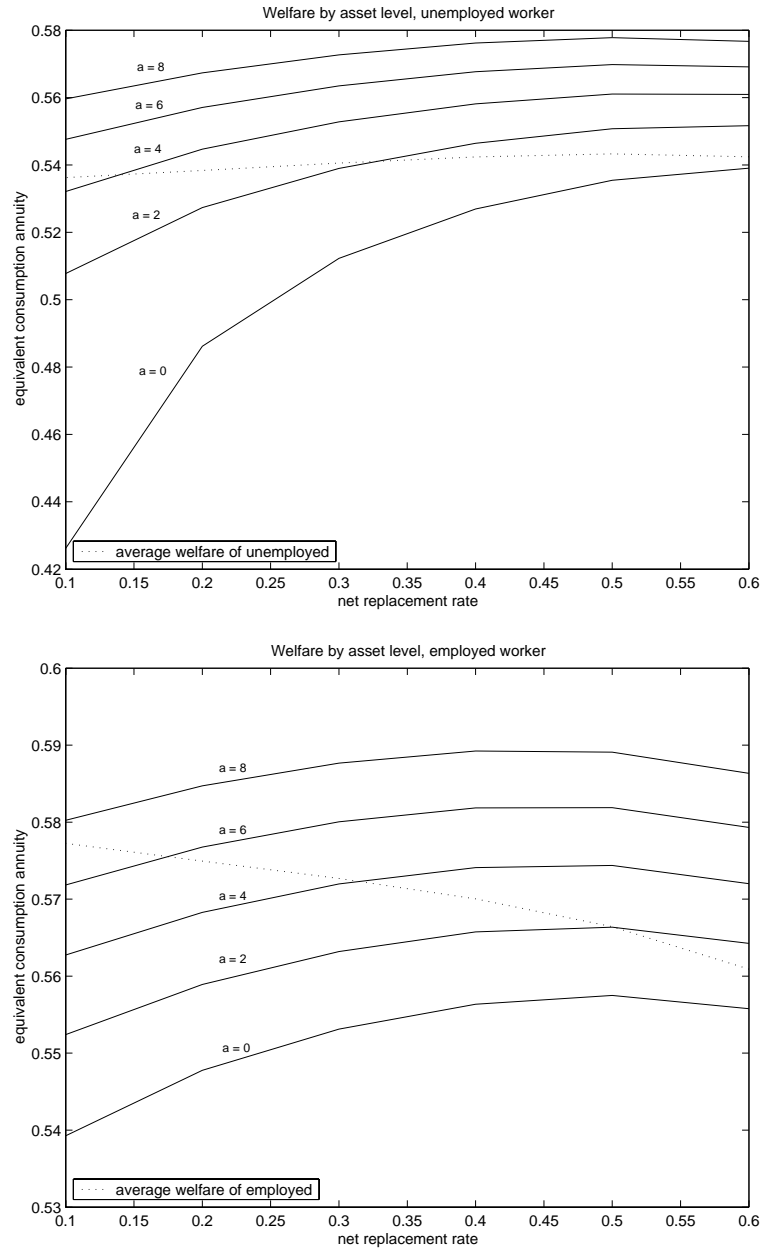


Figure 8: Steady-state welfare by asset level and employment status across replacement rates

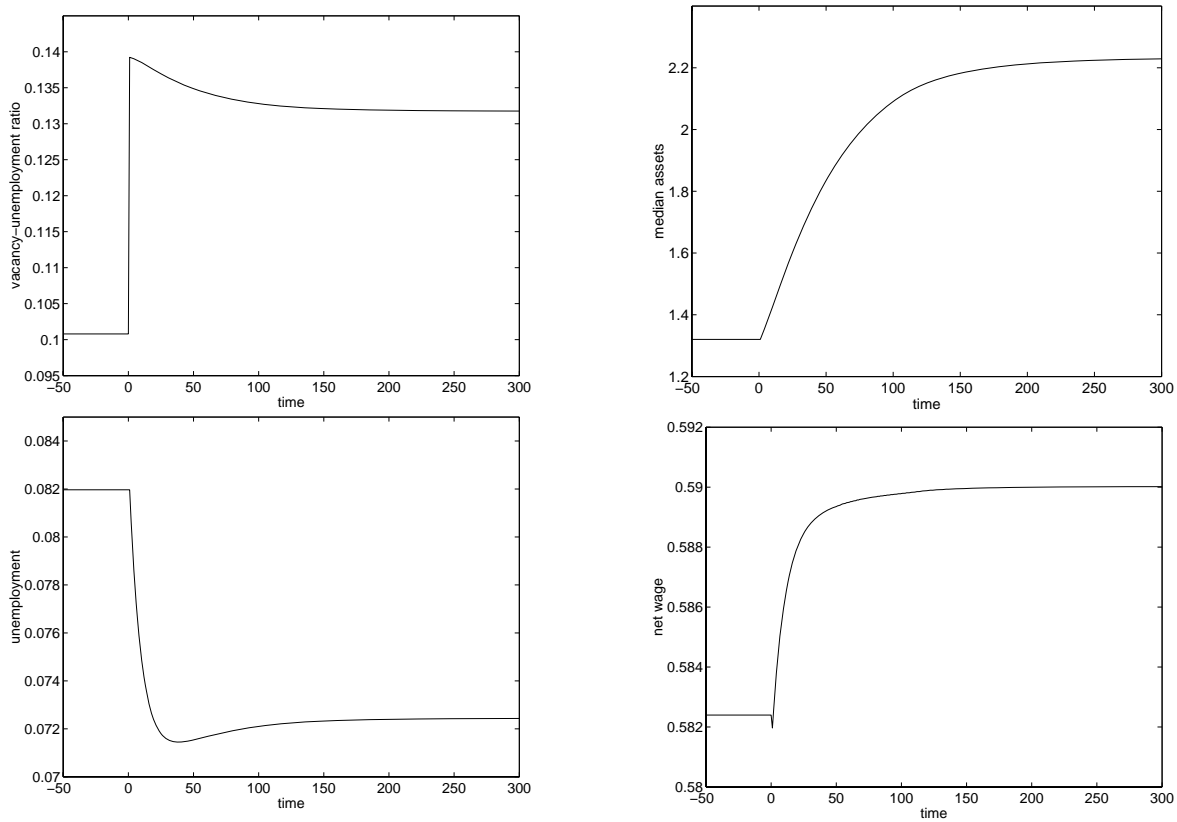


Figure 9: The transition paths of vacancy-unemployment ratio, median asset holdings, unemployment and net wages

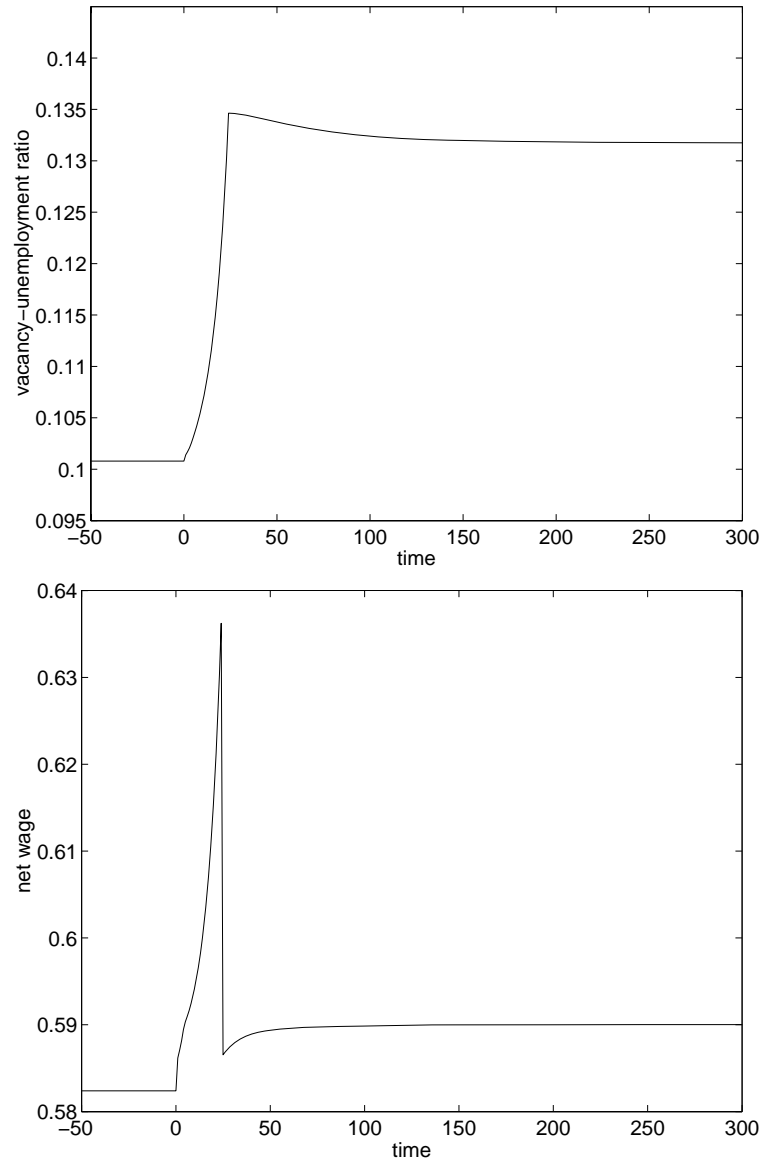


Figure 10: The transition paths of vacancy-unemployment ratio and net wages with announcement

Figures for Chapter 4

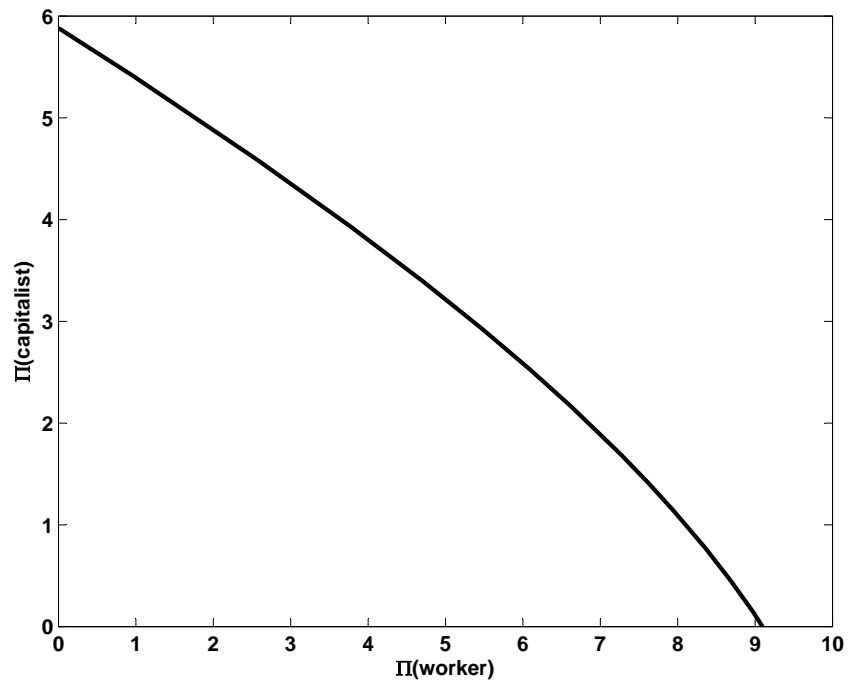


Figure 11: The set of Pareto improvements in terms of welfare gains (in %) in the baseline model

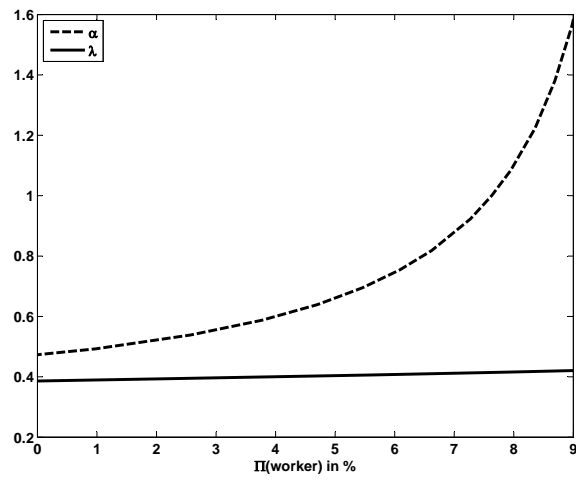
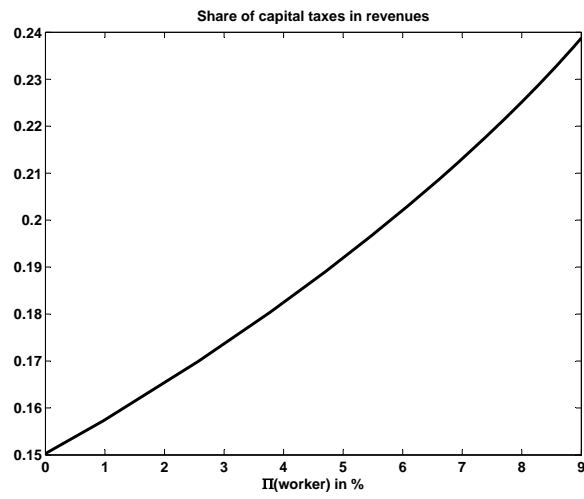
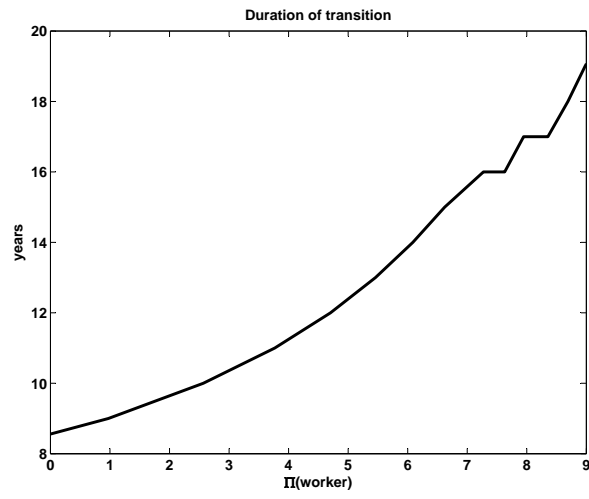


Figure 12: Properties of different POPI programs (baseline model)

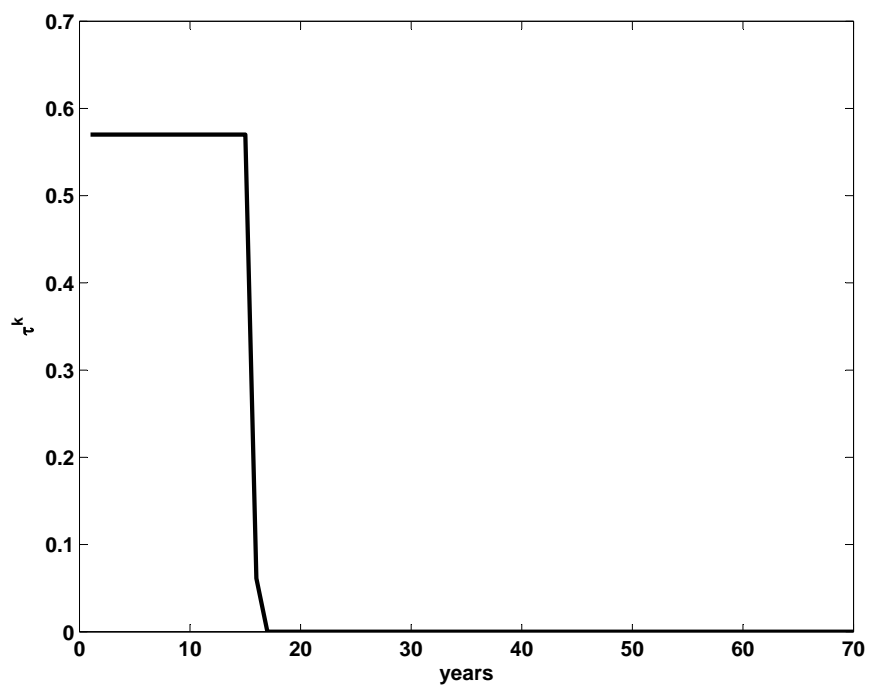


Figure 13: A typical time path for capital taxes

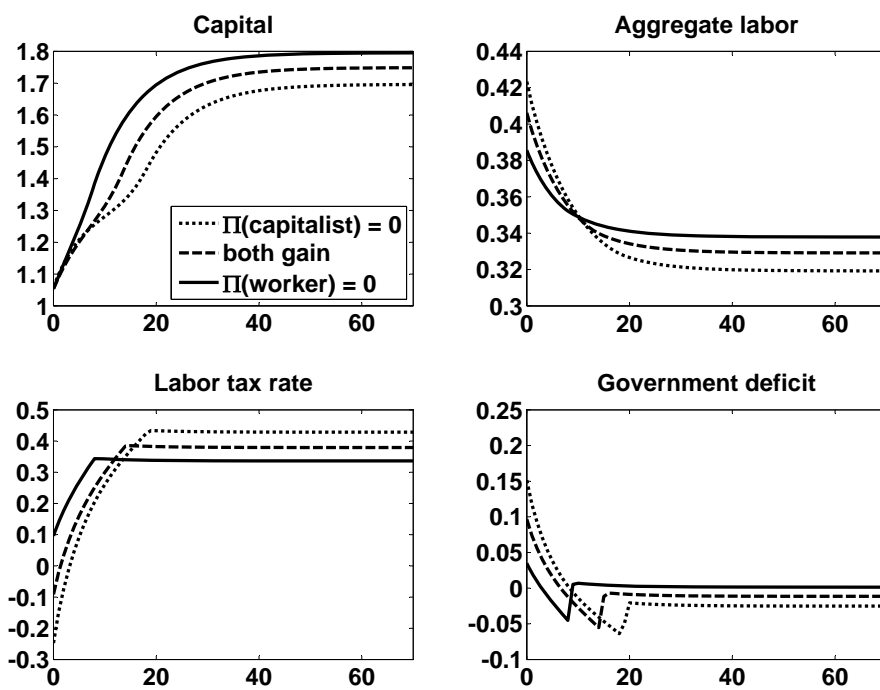


Figure 14: The time paths of selected variables for three POPI plans in the baseline model. (Time is in years.)

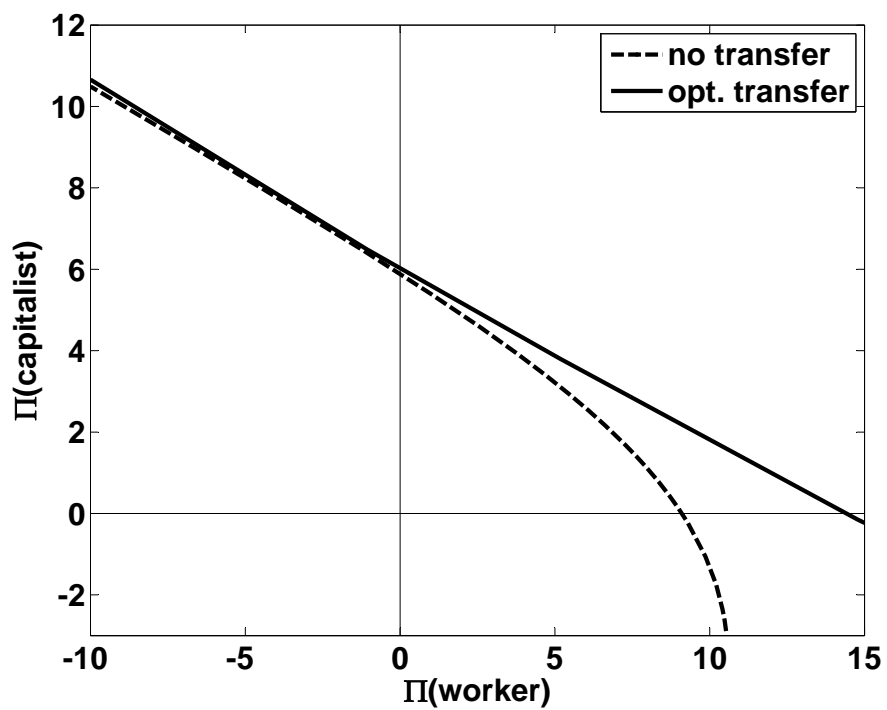


Figure 15: The frontier of feasible equilibria in terms of welfare gains (in %) in the baseline model and in MM1

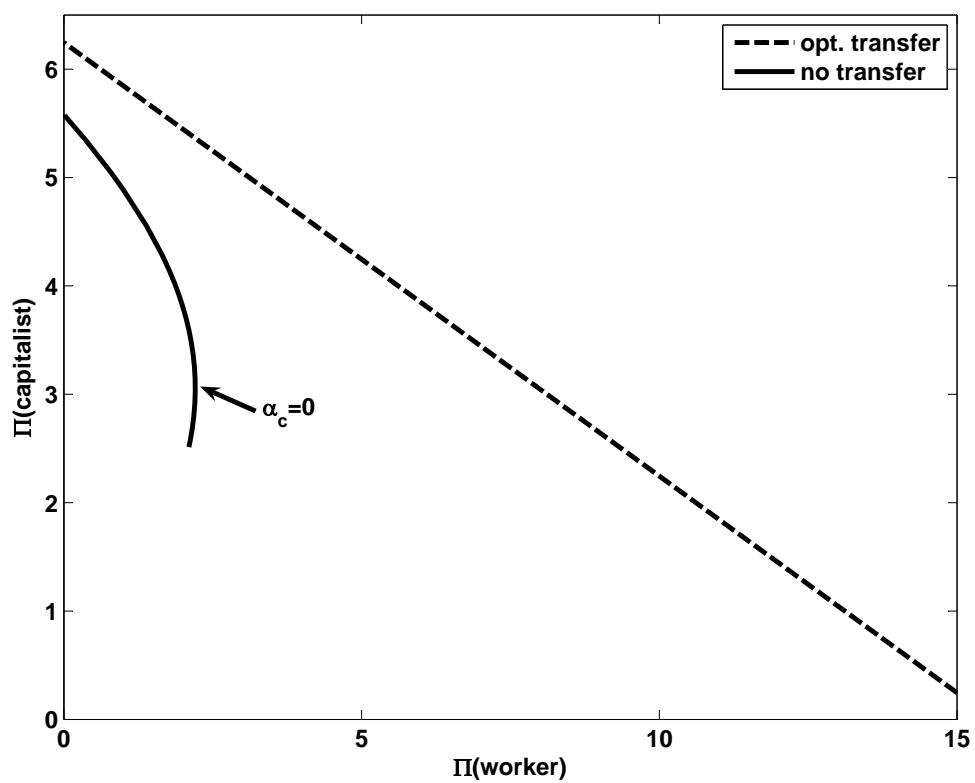


Figure 16: The frontier of Pareto improving feasible equilibria in terms of welfare gains (in %) in the model without labor choice (MM2)

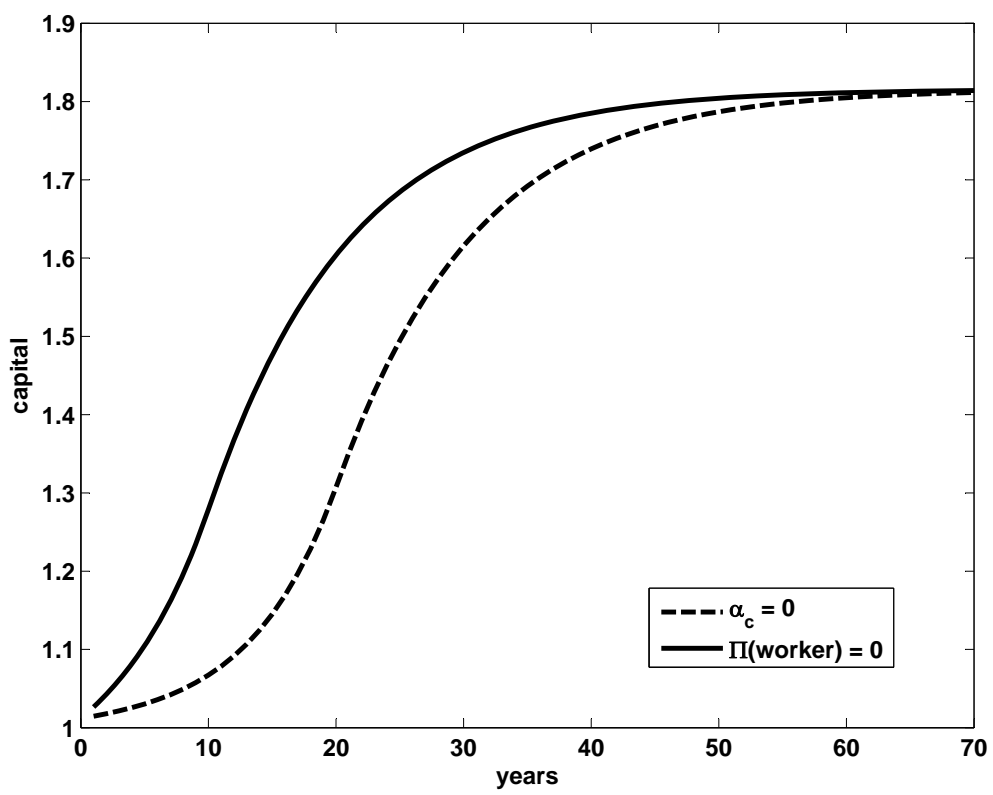


Figure 17: Illustrative paths of capital in the model without labor choice (MM2)