PhD

Management & Business Administration

Economia y Empresa

Three Essays On Liquidity and Contagion

By

Andrea TOTO

April 2016

Supervisors:

Mario Eboli
Simone Alfarano

to my father Elio,
to my mother Liliana,
to my sister Claudia,
to Massimiliano,
to my nephew Davide,
and to my little niece Sofia

Contents

RESUMEN EXTENSO

La presente tesis doctoral, "Tres ensayos sobre la liquidez y el contagio", se compone de cuatro capítulos. El primer capítulo introductorio contribuye a la finalidad de introducir los conceptos básicos de fondo de los tres capítulos siguientes. Más precisamente, se explica cómo la literatura fundamental define los conceptos de contagio financiero, riesgo de contraparte y riesgo sistémico. El contagio financiero se define en términos generales como la transmisión de las dificultades financieras a través de los agentes, sectores o regiones de la economía. La literatura distigue diferentes formas de contagio financiero, que corresponden a diferentes posibles canales de propagación. Dos tipos principales de dificultades financieras pueden propagarse por contagio. Un tipo conduce a la insolvencia, es decir, el tipo de defecto que se produce cuando el valor de los activos cae por debajo del valor de los pasivos o, equivalentemente, cuando el capital se vuelve negativo. El otro tipo de problema financiero implica la falta de liquidez. A continuación, se introduce cómo la literatura aplica teorías de redes y técnicas desarrolladas en las matemáticas y la física teórica para estudiar el contagio y el riesgo sistémico en los sistemas financieros. Los sistemas financieros modernos exhiben un alto grado de interdependencia. Existen diferentes fuentes de conexiones entre las instituciones financieras, derivadas tanto de la parte del activo como del pasivo de su balance. Los bancos están conectados directamente a través de las exposiciones recíprocas adquiridas en el mercado interbancario. Del mismo modo, tener carteras similares o compartir la misma masa de depositantes crea vínculos indirectos entre instituciones financieras. Pensado como un conjunto de nodos y enlaces entre los nodos, las redes pueden ser una representación útil de los sistemas financieros. Al proporcionar medios para modelar las interacciones económicas, el análisis de redes puede explicar ciertos fenómenos económicos. El uso de las teorías de red puede enriquecer la comprensión de los sistemas financieros. Introducimos algunas direcciones recientes en las redes financieras relativas a las sinergias creadas a partir de la aplicación de la teoría de redes para responder a cuestiones financieras sobre el tema del riesgo sistémico. En este contexto, se plantean dos preguntas: cómo de resistentes son las redes financieras al contagio, y cómo las instituciones financieras forman conexiones cuando se exponen al riesgo de contagio.

En el segundo capítulo se revisan los modelos de riesgo de crédito y los modelos de riesgo de contraparte y contagio y su aplicación en la gestión del riesgo de crédito, y se comparan los dos tipos principales de modelos en la literatura que tratan de describir los procesos predeterminados para las obligaciones de deuda y otros instrumentos financieros que son *"defaultable"* (que son susceptibles de incumplimiento); estos modelos normalmente se conocen como modelos estructurales y de forma (o intensidad) reducida. Al hacer esto, ponemos énfasis en sus principales diferencias y posibles conjunciones y aspectos comunes. Además, se discuten los desafíos y posibles progresos a ser alcanzados al reducir la distancia entre los modelos estructurales y de forma reducida en modelar el riesgo de contraparte y riesgo de crédito, sobre todo dentro de una perspectiva basada en la información, al estilo de Jarrow y Protter (2004). Esta perspectiva permite hacer más fácil la comparación entre los dos enfoques y también es una buena fuente de creación de una nueva corriente de literatura que es capaz de incorporar y reconciliar ambos enfoques dentro de los modelos híbridos, con el fin de conseguir mejoras en la evaluación de la contraparte y el riesgo crediticio. En ambas corrientes separadas de la literatura, el objetivo principal es la predicción por defecto; por tanto el debate entre los modelos debería naturalmente concentrarse en él. Una visión unitaria de los dos enfoques debería considerar las dos corrientes de modelos como la misma, pero utilizando diferentes supuestos acerca de la información disponible para el modelador. Desde el punto de vista informativo, los modelos estructurales suponen un conocimiento completo (muy detallado y de información continua), cerrados a la de la visibilidad de los gerentes. La implicación principal es que este tipo de modelo es muy apropiado para los ámbitos internos y también para los reguladores (en el caso de los bancos comerciales). Para aclarar esto, se considera que es evidente que para las necesidades internas los modelos más apropiados deberían utilizar tanta información disponible para los administradores como sea posible; la pregunta sigue siendo "¿Por qué los reguladores podrían (y deberían) utilizar modelos estructurales?". La respuesta es "Porque tienen derecho, por ley, (y acceso) a casi la misma información que los "*insiders*" (gerentes), por lo que deberían beneficiarse de esta situación". Además, en el análisis de los modelos estructurales, también hay que tener en cuenta un punto débil: si no se permiten saltos, el tiempo predeterminado es predecible, o en otras palabras, en teoría sugieren una propagación de cero a corto plazo, lo cual contradice el comportamiento empíricamente observado. Por otra parte, los modelos de forma reducida presuponen un conocimiento de un conjunto de información menos detallada, como el observado (imitando) por el mercado (los inversores). La principal consecuencia es que el tiempo predeterminado es impredecible. La información de mercado debería tenerse en cuenta, y, en consecuencia, los modelos reducidos son importantes en vigor de razones realistas. Cuando los precios – y por tanto la evolución de las empresas - son determinados, en equilibrio, en el mercado, se sienten atraídos por los agentes que conocen (la mayoría de ellos) sólo información pública. Por ejemplo, el proceso de valor de activos no es observable a los "*outsiders*", más exactamente se observa de forma discontinua; cuando la empresa revela información contable y otros temas relevantes. Este punto de vista hace que los modelos de forma reducida sean muy apropiados para la fijación de precios y la cobertura de riesgos de crédito. Para un análisis correcto de estos modelos también hay que tener en cuenta un gran problema que inducen, comparando con la primera corriente: ahora el valor predeterminado se plantea de forma exógena, no endógena, al igual que en los modelos estructurales (llamados por esta característica modelos de causa-efecto). Esto también se contradice con la realidad, ya que los modelos de forma reducida asumían que el incumplimiento financiero no está relacionado en absoluto con las características de las empresas, lo que es claramente un defecto de estos modelos. La intuición detrás de la perspectiva de la reconciliación surge como una forma de mejorar las corrientes separadas anteriores, y también para tener en cuenta sus puntos débiles y para encontrar la manera de tratar con ellos. Básicamente, el objetivo es poner en práctica modelos que debería ser más realistas desde el punto de vista general del "*outsider*", que tratan con información incompleta o al menos con supuestos relajados sobre información completa, donde el incumplimiento financiero debería ser impredecible, pero también endógenamente influenciado. El desarrollo reciente del tema puede ser una explicación de la complejidad e importancia de estos modelos, pero también puede sugerir que muchas cosas acerca de estos modelos no fueron reveladas hasta ahora. Los modelos de conciliación o híbridos deberían estar dirigidos a relajar el supuesto de conocimiento de información completa, de tal manera que a partir de los modelos estructurales con valores predeterminados predecibles para obtener modelos de tasa de riesgo con un valor predeterminado inaccesible. Duffie y Lando (2001), Giesecke y Goldberg (2004) y Cetin et al. (2004) representan tres enfoques diferentes con el fin de conciliar los dos enfoques principales. El modelo de Duffie y Lando (2001) es muy similar a los modelos de primer paso (el tiempo predeterminado es fijado por los gerentes para la maximización del valor de las acciones de las empresas). Los inversores están recibiendo informes contables periódicos e imperfectos, y hacen inferencias acerca de la evolución de la empresa sobre la base de estos informes, y añadiendo, obviamente, sus creencias (ruido). El punto crucial del artículo es que, a pesar de que comience más como un modelo estructural, se puede asegurar un el incumplimiento financiero impredecible, al igual que los modelos de forma reducida. La explicación de este importante cambio del tiempo de parada de predecible a impredecible es que entre los tiempos de observación, el inversor no puede ver la evolución de los activos. Giesecke y Goldberg (2004) asumen, de acuerdo con los modelos estructurales, un valor del activo continuamente observado. En su enfoque también se introduce el ruido, pero la forma es bastante diferente: el punto de quiebra es una curva al azar, más exactamente con distribución beta, con la altura expresada en términos de apalancamiento de la empresa. La explicación detrás de esto es bastante intuitiva: cuando el ratio de apalancamiento es en una reciente alta, entonces la incertidumbre a corto plazo es alta también. El modelador no puede ver la curva, que es independiente del modelo estructural subyacente, por lo que el tiempo predeterminado depende de una curva no observable, por lo tanto es inaccesible. De esta forma se resuelve el problema de la previsibilidad de los modelos estructurales. El modelo de Cetin et al. (2004) se puede ver como un enfoque alternativo de los dos anteriores. Los autores, en lugar de añadir ruido a la información **oscura** (*obscure information*) como en Duffie y Lando (2001), comienzan también con un modelo estructural, pero con el modelador de filtración G para ser una sub-filtración estricta de la que disponen los gerentes (incompleta, pero información correcta). El tiempo predeterminado es la primera vez, después de que los flujos de caja están por debajo de cero, cuando el flujo de caja se mantiene por debajo de cero durante un cierto tiempo y luego se duplica en magnitud absoluta. De esta forma se obtiene un tiempo predeterminado totalmente inaccesible, y el proceso (*point process*) tiene intensidad, por lo que este es un modelo de riesgo basado en la intensidad. Cualquier modelo estructural con información incompleta admite una tendencia de precios, pero no todos admiten intensidad. La imprevisibilidad de incumplimiento es una condición necesaria, pero no suficiente, para que la tendencia de precios admita intensidad. Además, el nivel de información determina si el modelo admite intensidad: cuando no hay certeza, no hay intensidad. Entonces, partiendo de un modelo estructural, y relajando en diferentes grados los supuestos de información, Giesecke y Goldberg (2004), considerando el modelo estructural con la información completa acerca de los activos y punto de quiebra como punto de partida el más "completo" en términos de cantidad y calidad de la información disponible, nótese que si se supone una información incompleta acerca de una barrera, se puede calcular la tendencia de precios en términos de la función de distribución para la barrera y el valor histórico de los activos observable. Aquí tendencia de precios no admite una intensidad de defecto. Si se supone información incompleta tanto de activo como barrera, la tendencia de precios admite una representación de la intensidad. Un hallazgo importante es que, en algunas condiciones particulares, sin tener en cuenta que la "*default barrier*" es observable o no, un modelo estructural con información incompleta de activos admite una representación de la intensidad. Así que no hay ganancia para un modelador de la reconciliación por relajar en la parte de la barrera; también la existencia de intensidad, si es necesario en un modelo en particular, está asegurada. Otro marco para desarrollar un modelo de reconciliación deseable podría ser empezar al revés, es decir, con el enfoque de modelos reducidos. Aquí el modelador sólo conoce partes discretas y perfectas de información (informes de contabilidad en el momento que son revelados, información pública acerca de la empresa, también en el corto plazo después de revelar). Él tiene que hacer inferencias para los períodos en los que no dispone de información. Por esa razón, el modelador puede utilizar la teoría y fórmulas desarrolladas para los modelos de forma reducida, pero debe ponderar de forma diferente la información, confiando más en lo que él sabe perfectamente (por un periodo corto de tiempo) y menos en la parte inferida, ya que es alterado con sus propias creencias e inferencias.

En el tercer capítulo se analizan los efectos de crédito comercial en las decisiones de inversión de una empresa restringida financieramente en un sector manufacturero, con particular referencia a un contexto de turbulencias financieras y racionamiento del crédito. La evidencia empírica muestra que la tasa de interés implícita en un contrato de crédito comercial es en general muy alta en comparación con las tasas de crédito bancario. A pesar de este aparente alto costo, el crédito comercial es ampliamente utilizado y representa una proporción importante de las finanzas de las empresas. El crédito comercial crea riesgo sistémico pero, al mismo tiempo, aumenta la resistencia de las empresas a los shocks de liquidez y, a través de la flexibilidad de los plazos de amortización, afecta positivamente a las decisiones de inversión de inventario y, en consecuencia, los niveles de producción e ingresos futuros esperados. Estas características del crédito comercial se investigan con el objetivo de caracterizar sus efectos sobre la financiación de la inversión en inventario de las empresas con restricciones de liquidez. Con este fin, hemos presentado un modelo multifactorial de una empresa que maximiza el beneficio sujeto a las restricciones de crédito bancarias y con tres fuentes de financiación: la auto-financiación, crédito bancario y crédito comercial. El modelo es capaz de captar el efecto del seguro de crédito comercial, es decir, la cobertura de seguro contra el riesgo de liquidez implícitos en los contratos de crédito comercial, gracias a la cual una empresa financieramente restringida que sufre de una escasez de liquidez puede mantener un nivel de inversión de inventario esperado (y, como en consecuencia, un futuro nivel de producción esperado) lo más cerca posible al nivel óptimo deseado. Medimos el efecto del seguro de crédito comercial mediante la caracterización de los efectos de un shock de liquidez exógeno en las decisiones de inversión y futuro nivel esperado de producción de una empresa financieramente restringido en una cadena de suministro manufacturera, bajo el supuesto de que su proveedor está dispuesto a conceder una prórroga de pago de una parte de la deuda comercial. La solución del problema de inversión óptima muestra que, bajo limitaciones financieros actuales y futuras previstas, el importe del crédito comercial actual y futura disponible afecta a las decisiones de inversión. Debido a este efecto de seguro, el crédito comercial es una fuente óptima de financiación para una empresa financieramente restringida bajo una futura escasez de liquidez esperada. Entonces mediante el uso de condiciones de optimización de la modalidad de análisis y derivando una ecuación de inventarios de forma reducida, proponemos un marco empírico con el fin de probar conjeturas e implicaciones del modelo analítico. Se lleva a cabo un conjunto de regresiones econométricas sobre una muestra de empresas manufactureras italianas. La muestra se obtiene del conjunto de datos AIDA (Empresa Italiana de Información e Inteligencia de Negocios), provista por la Bureau Van Dijk: está constituido por 1 millón de pequeñas y medianas empresas en Italia durante un período de diez años. Los resultados del análisis empírico se dibujarán en la próxima versión de este artículo.

El cuarto capítulo es un artículo que estudia los efectos que dos características de la topología de una red financiera, es decir sus grados de conectividad y de centralización, tienen en la respuesta de la red a los choques externos que pueden generar fenómenos de "*default contagion*". Presentamos algunas conjeturas acerca de tales efectos, conjeturas basadas en algunos resultados analíticos que arrojan algo de luz sobre la exposición al riesgo sistémico de tres clases muy estilizadas de redes: i) redes completas, que son las más conectadas; ii) redes circulares (también conocidas como "ruedas"), que son las redes menos conectadas y menos centralizadas; y iii) las redes en forma de estrella, que son las redes menos conectadas y más centralizadas. Se conjetura que cuanto más conectada es una red, más muestra un carácter robusto aunque frágil, en el sentido de que es completamente resistente a shocks relativamente pequeñas pero está expuesto al riesgo de una fusión total (el incumplimiento financiero de todos los agentes de la red) si es golpeado por shocks suficientemente grandes. También conjeturar que la centralización de una red financiera tiene los mismos efectos: cuanto más centralizada es una red, más sólida aunque frágil será. A la inversa, conjeturar que una red dispersa y descentralizada, como las redes circulares, tiene la característica opuesta: cuanto más dispersa y descentralizada es una red, tendrá una naturaleza más vulnerable aunque resistente, en el sentido de que está expuesta a episodios de contagio local debido a shocks relativamente pequeños, mientras que se a un riesgo pequeño de contagio completo. Más precisamente: i) la conectividad completa así como la máxima centralización hacen una red robusta y sin embargo frágil, en el sentido de que estos tipos de redes tienen un único umbral de contagio: para choques más pequeños que ese umbral, no hay defaults (incumplimientos financieros) secundarios, mientras que para los choques más grandes que ese umbral todos los miembros de la red entran en estado de default; y, a la inversa, ii) la red con conectividad mínima y centralización mínima se caracteriza por una gran diferencia entre los primeros y últimos umbrales de contagio, por lo tanto, muestra un comportamiento vulnerable sin embargo resistente: se expone a episodios de contagio local debido a shocks relativamente pequeños mientras que es resistente con respecto a grandes shocks. Por otra parte, conjeturamos que estos efectos de centralización y conectividad se aplican a la red genérica de una manera que es proporcional a su grado de densidad y centralización. Para probar estas conjeturas, corremos simulaciones numéricas en redes generadas al azar con diversos grados de conectividad y centralización. Los resultados obtenidos confirman nuestras conjeturas. Hemos probado las conjeturas de los efectos de la conectividad mediante la ejecución de simulaciones de la clase de redes regulares, donde todos los nodos tienen el mismo grado de centralidad, por lo tanto, la centralización se mantiene a cero. Obtenemos que, a medida que aumenta la densidad, las redes se vuelven progresivamente más robusto aunque frágiles: el primer y último umbrales de contagio convergen al umbral único de la configuración pura en forma de estrella de una manera cuasi-monotónica. Del mismo modo, hemos probado los efectos de la centralización en una clase de redes con conectividad constante y casi mínimo, pasando de las redes circulares hacia las redes en forma de estrella. Nos encontramos con que la brecha entre el primero y el último umbral de contagio disminuye a medida que nos movemos de redes dispersas y descentralizadas hacia las redes dispersas y altamente centralizados, lo cual demuestra que las características vulnerables pero elásticos de las redes circulares es reemplazado progresivamente por la naturaleza robusta sin embargo frágil de la red en forma de estrella muy centralizado. También en estas pruebas, la convergencia del primera y último umbral hacia el umbral único de la red en estrella es cuasi-monotónica. Curiosamente, nuestros resultados muestran que el patrón del primer umbral en la primera serie de experimentos es claramente convexo, mientras que el patrón del mismo umbral en el segundo conjunto de experimentos es claramente cóncavo. Todo lo contrario se aplica a la estructura del umbral contagio final: en las simulaciones destinadas a probar los efectos de la conectividad, el umbral final muestra un patrón cóncavo, mientras que, en las simulaciones que ponen a prueba los efectos de la centralización, el umbral final tiene un patrón cuidadosamente convexo. Este resultado indica que los efectos del aumento de la conectividad, en la prestación de una red robusta aunque frágil, se convirtió en notable a partir de niveles relativamente bajos de conectividad. Por el contrario, el aumento de la centralización, que también hace que una red sea cada vez más robusta y sin embargo frágil, produce efectos evidentes sólo para valores altos de centralización. En otras palabras, las pérdidas debidas a shocks exógenos se distribuyen entre los miembros de una red de una forma aún - generando el fenómeno robusto y sin embargo frágil - partiendo de niveles relativamente bajos de densidad. Por el contrario, el aumento de la centralización proporciona el mismo efecto pero sólo para altos niveles de centralización.

Chapter 1

Introduction and Motivation

Andrea TOTO*

february 2016

Abstract

This paper is the first introductory chapter of the PhD thesis "Three Essays on Liquidity and Contagion". It serves the aim to introduce the basic background concepts of the subsequent three chapters. More precisely, we explain how the essential literature defines the concepts of financial contagion, counterparty risk and systemic risk. Financial contagion is broadly defined as the transmission of financial distress across agents, sectors or regions of the economy. The literature has distinguished among different forms of financial contagion, corresponding to different possible channels of propagation. Two major kinds of financial distress can spread by contagion. One kind leads to insolvency, i.e. the kind of default that occurs when the value of assets drops below the value of liabilities or, equivalently, when capital becomes negative. The other kind of financial distress involves illiquidity. Then we introduce how the literature applies network theories and techniques developed in mathematics and theoretical physics to study contagion and systemic risk in financial systems. Modern financial systems exhibit a high degree of interdependence. There are different possible sources of connections between financial institutions, stemming from both the asset and the liability side of their balance sheet. Banks are directly connected through mutual exposures acquired on the interbank market. Likewise, holding similar portfolios or sharing the same mass of depositors creates indirect

*Università "G. d'Annunzio", Department of Business Economics, Pescara, Italy. I thank Mario Eboli, my PhD supervisor, for having given me suggestions and indications and for having read preliminary versions of this paper. I am the sole responsible for any mistake or shortcoming that can be found in these pages. I can be contacted via e-mail at: andrea.toto@unich.it

linkages between financial institutions. Thought as a collection of nodes and links between nodes, networks can be a useful representation of financial systems. By providing means to model economic interactions, network analysis can explain certain economic phenomena. The use of network theories can enrich the understanding of financial systems. We introduce some recent addresses in financial networks concerning the synergies created from applying network theory to answer financial questions about the issue of systemic risk. In this context, two questions arise: how resilient financial networks are to contagion, and how financial institutions form connections when exposed to the risk of contagion.

Keywords: financial contagion – credit risk – counterparty risk – credit contagion – credit risk models – financial networks - trade credit

1 Contagion

The word "contagion", applied in a financial context, suggests that financial contagion is analogous to the spread of disease, and that damaging financial crises may be better understood by bringing to bear ideas that have been developed to understand the breakdown of other complex systems in our world. It also suggests that the aim of systemic risk management is similar to a primary aim of epidemiology, namely to identify situations when contagion danger is high, and then make targeted interventions to damp out the risk. Contagion, meaning the transmission of a disease by direct or indirect contact, is an appropriate term for the damaging effects that can be transmitted through the interbank network.

Financial contagion is broadly defined as the transmission of financial distress across agents, sectors or regions of the economy. The literature has distinguished among three different forms of financial contagion, also known as systemic risk, corresponding to different possible channels of propagation[1]: 1) Informational contagion, that can occur in banking systems,where depositors'expectations about the possibility of a crisis can lead to bank runs, and in imperfectly informed financial markets, where "bad news" can affect the sentiments of the traders; 2) Direct contagion transmitted via networks of financial obligations. In banking and financial systems, such networks arises from three sources: i) loans and deposits in the interbank money market, ii) "over-the-counter" trading in assets and derivatives, and iii) payment systems; while, in the manufacturing sector, networks of

[1] See the review articles by Dow (2000) and by De Bandt-Hartmann (2000).

financial obligations arise from trade credit.[2] 3) Common exposure to losses in the value of assets, losses that can be exogenous or endogenous to a financial network, the latter being the case of fire sales of illiquid assets induced by liquidity shortages.

Modelling credit risk in a coherent yet applicable manner is an important yet challenging problem. The difficulties arise from the combination of a large, and co-dependent set of risk parameters such as default rates, recovery rates, or exposures, which are correlated and non-stationary in time. Credit contagion examines the role of counterparty risk in credit risk modelling. If a firm is in economic distress, or defaults, this will have implications for any firm which is economically influenced by this given firm, for example, a service provider to it, purchaser of its goods or a bank with a credit line to the firm. The direct correlations between firms caused by credit contagion lead to further complications in modelling the overall, either portfolio or economy wide, level of risk. One must also distinguish contagion from correlation, and counterparty contagion from other kinds of contagion.

The counterparty risk is an element of credit contagion: in other words, it produces systemic effects that must be taken into account in assessing credit risk. Credit risk is one of the most analyzed market risk components and it's very difficult to quantify it in a reliable way. Credit risk measurement and management models have become more important in recent years in the process of risk management for financial institutions. Traditionally the problem has been addressed through the application of actuarial methods based on historical data; but as recent empirical studies have demonstrated that counterparty risk is able to produce significant systemic effects on credit risk, they have highlighted the inadequacy of the traditional standard credit risk models, because they fail to consider and capture the credit contagion phenomenon.

The notion of counterparty risk has been introduced by Jarrow and Yu (2001) and afterwards has been considered both in reduced form models (see Giesecke, 2003, Frey and Backhaus, 2004) and in structural models (see Egloff, Leippold and Vanini, 2007, Giesecke, 2004). Moreover, the counterparty risk may be taken into consideration also in binomial-type models for the credit risk of portfolios of financial positions, as has been made by Davis and Lo (2001) and Giesecke and Weber (2004).

The literature on counterparty contagion is broad and includes mathematical treatments of random graphs, interacting particle systems, and Markov processes, but also financial theorizations about balance sheet constraints and haircuts in collateralized lending,

[2]see Kiyotaki-Moore (2001, 2002)

and also empirical studies of data provided by banking regulators. Counterparty contagion comprises a family of phenomena. There are different kinds of counterparty relationships, and multiple distresses that a firm can transmit to its counterparty.

Credit events tend to cluster in times of economic stress, forcing banks to recognize disproportionately many defaults in recessions. This is due to two reasons. First, the profitability of firms depends on common macro-economic variables, such as economic growth, leading to an increase in default rates when macroeconomic conditions are poor. Second, firms are directly linked with each other through business relations. The default of a large customer or supplier, for instance, will adversely affect the credit position of a firm, which may then default and in turn influence its customers and suppliers. Such a direct dependency of defaults is referred to as credit contagion. Both mechanisms underlying clustering of defaults have received considerable attention in the credit risk literature for more than a decade; see (Davis, Lo, 2001) and (Frey, Backhaus, 2003), (Rogge, Schonbucher, 2003), (Giesecke, Weber, 2004), (Neu,Kuhn, 2004), (Hatchett, Kühn, 2006, 2009), (Egloff, Leippold, Vanini, 2007). Evidence suggests that the dependence on common factors can by itself not explain observed levels of correlation, and that credit contagion, possibly in conjunction with the effect of further unobserved macro-economic covariates — so-called frailty — is important to explain the data .

Clustering of Defaults means that multiple defaults can occur at the same time. Clustering can occur within or across industries via common shocks to cash flows, or via counterparty effects, which arises from trade credit between industrial partners or from lending by financial institutions. The clustering in default correlations is sometimes called "credit contagion". Sources of default clustering are: i) common risk factors (correlation among firms): ii) contagion due to financial connection (trade credit between industrial partners or from lending by financial institutions). Unexplained default clustering is a major issue for traditional credit risk models because it generates greater dispersion, or fatter tails, in the distribution of credit losses. This implies a greater likelihood of large losses and an understatement of economic capital. This could lead to a greater number of bank failures in periods of stress, or losses on CDOs that exceed worst estimates.

Correlation in multiple factor effects, or industry factors, is a source of default clustering. When a firms defaults, other firms in the same industry could suffer from contagion effects, reflecting shocks to cash flows that are common to that industry. Examining firms within the same industry, Lang and Stulz (1992) and Jorion and Zhang (2007) present evidence that industry peers are negatively affected by a Chapter 11 bankruptcy, creating higher correlation within the industry.

As important channel of credit contagion, the counterparty credit risk is a completely different factor, compared to common risk factors. This arises when the default of one firm causes financial distress for its creditors. In an extreme case, this can push a creditor toward default as well. This in turn can lead to a cascade of other defaults. Such interactions are particularly worrisome for financial institutions, given their intricate web of relationships. This channel is very different from industry or factor effects. It requires detailed information about counterparty exposures. Counterparty risk affects the shape of the default distribution, thus providing a potential explanation for the observed default clustering.

For industrial firms, most exposures take the form of trade credit, defined as direct lending in a supplier-customer relationship. For financial firms, exposures take the form of loans or bonds and are generally larger in money amounts than for industrial creditors, but less so in relative terms, when considering the larger balance sheets of financial creditors[3]. Counterparty risk is likely to be smaller for financial firms as lenders or bondholders for a number of reasons. First, banks impose limits on the amount of lending to one borrower and are forced to diversify by regulators. Secondly, there are other mechanisms that can help mitigate risk. Financial institutions have the power to choose whom they lend to, in contrast to trade credit, which is generally involuntary. Thirdly, bank loans are generally secured, leading to higher recovery rates than unsecured debt. In contrast, the bankruptcy of a debtor subjects an industrial firm to a double penalty, loss of trade credit and loss of valuable customer relationship. Therefore, the direct counterparty effects should be stronger for an industrial counterparty than for a financial institution.

1.1 Contagion and Correlation

Contagion must be distinguished from correlation, and counterparty contagion from other kinds of contagion. More precisely, contagion is distinguished from correlation between firms that does not feature a causal link; counterparty contagion as mediated by various kinds of bilateral deals is distinguished from other forms of contagion that are intermediated by markets. It is a well established principle that correlation is not causation. Contagion is a form of causation: the default of one firm contributes to the default of another firm, via

[3]In Jorion, Zhang (2009) sample, the average exposure for financial institutions is 0.16% of equity.

contagion. Contagion is one of the phenomena that can yield correlation among defaults. Exposure to common risk factors is another such phenomenon. Common risk factors and contagion are both phenomena that can cause correlated defaults, i.e. multiple defaults in the absence of any direct causal link between those defaults.

Contagion between asset markets during financial crises is defined as the transmission of shocks via newly opened channels associated with crisis events. Contagion effects may be evidenced as increased correlation, or as lower correlation consistent with breaking linkages between financial institutions as proposed in network theory; see Allen and Babus (2008).

Contagion, as defined in Forbes and Rigobon (2002), means significant increases in cross-market comovements, while any continued market correlation at high levels is considered to be interdependence (Forbes and Rigobon, 2002): it implies that the existence of contagion must involve evidence of a dynamic increment in correlations.

Default correlations are the most important drivers of the tails of portfolio credit risk distributions. Empirically, default correlations are positive, and this positive correlation increases portfolio risk. Default correlations cannot be measured directly, and must be inferred from a model. Higher default correlations imply greater probabilities of extreme losses on the portfolio.

Correlation involves exposure to common risk factors. This type of correlation is source of contagion. But contagion can cause correlation in default events.

1.2 Types of contagion

Two major kinds of financial distress can spread by contagion. One kind leads to insolvency, which is the kind of default that occurs when the value of assets drops below the value of liabilities; equivalently, when capital, which is the difference between the values of assets and liabilities, becomes negative. The other kind of financial distress involves illiquidity. Firms must meet their currently due liabilities using their liquid assets, those that can be readily converted into currency. Default occurs when a firm's liquid assets are insufficient to meet its currently due liabilities. Either kind of financial distress can spread by affecting a firm's assets or its liabilities. There are multiple channels of contagion, some direct between counterparties, and some indirect, mediated by markets and information.

Contagion spreads directly between counterparties when a distressed firm imposes losses on its creditors, or when a distressed firm withdraws funding from its borrowers. Within the financial system, losses are transmitted from the liability side of a defaulting firm's balance sheet to the asset side of its creditors' balance sheets. This phenomenon can involve insolvency or illiquidity. Often, the liabilities considered were only interbank loans, but other securities, including derivative securities, have been considered too. A closely related line of research studies counterparty contagion of insolvency among industrial firms, including trade credit and contracts in the supply chain. The supply chain has some unique features: for example, Battiston et al. (2007) consider among the effects of counterparty contagion the decrease in profits due to losing a supplier or customer. Research on counterparty contagion of illiquidity in the payments system began some years ago. This type of contagion is similar to that which travels from obligor to creditor, except that the money that ought to be paid is primarily that of the banks' customers. A rather different phenomenon is when the distress of a lender is transmitted to the liability side of its borrowers' balance sheets. A lender may refuse to roll over a maturing loan for various reasons. For example, it may need to reduce the risk of its asset portfolio due to prudential considerations or regulatory constraints, or it may need to shrink its balance sheet or hoard liquidity due to present or anticipated demands on its liquidity. A lender may fail to perform on a line of credit due to a shortage of liquidity. Thus, in a market environment in which funding is scarce, financial distress can travel from the lender to a borrower that does not receive the funding it expected. Contagion arises when this distress makes the borrower in turn unwilling or unable to extend credit to its own borrowers.

There is a form of contagion between counterparties that is indirect, mediated by information. Market participants may believe that the distress of one firm is likely to cause as-yet-undisclosed distress to its creditors, and consequently become less willing to lend to the creditors of the distressed firm. In this case, the creditors suffer counterparty contagion, even though the distressed firm has transmitted no losses directly to them.

Indirect channels allow several other types of contagion to spread between firms that are not counterparties. One is a general phenomenon of information-driven contagion via funding liquidity, akin to the case of funding lost by counterparties of a distressed firm. Firms that are similar to a distressed firm can also be perceived as less creditworthy, on the grounds that they are likely to experience similar trouble soon. Another channel involves the imperfect liquidity of asset markets. The sale of assets by one firm can impair the ability of other firms to sell the same assets and cause their price to drop, as the pool of available buyers is depleted, and motivated sellers drive down prices to find more buyers. This makes possible market-mediated contagion via asset portfolios: a firm that responds to distress by selling some of its assets drives down their prices and thus causes losses to other firms that hold the same assets. Collateralized lending is a third channel. Contagion can spread among firms that use the same assets as collateral for borrowing, in a phenomenon that combines funding liquidity and asset-market liquidity.

These types of contagion exist when market participants have faith in probabilistic models and act rationally. When they do not, further types of contagion are possible. As described by Krishnamurthy (2010), market participants face Knightian uncertainty, meaning that there is not enough information to quantify risks confidently by probabilities. Because market participants' probabilistic models are unreliable, when new evidence suggests that the models are flawed, market participants may suddenly adopt a more risk-averse posture as they become unwilling to trust their models. This can create contagion through channels described above: market participants who become more risk-averse sell risky assets, causing abnormal declines in their prices, causing more market participants to doubt their models and become more risk-averse; firms experiencing financial distress contribute to a general opinion that the scope of financial distress may exceed what is likely under normal conditions, causing potential lenders to doubt their models and reduce the amount of funding they provide, causing more financial distress. Herd behavior creates another channel of contagion. When market participants lose trust in their own models, they may base their behavior more on what they observe that others are doing, leading to herd behavior. This is a mechanism for the contagion of panic, leading to such crisis phenomena as the collapse of markets for commercial paper or various asset-backed securities. Related to this, there is an econophysics literature devoted to analyzing phenomena such as bubbles and crashes (Kaizoji, 2000) and contagion (Kaizoji, 2001) using models of interacting particle systems, in which particles are financial agents and interactions take the form of emulation. Samanidou et al. (2007) review econophysics research on agent-based models.

1.3 Models of Counterparty Contagion

Models of counterparty contagion can be visualized in terms of a network of firms or banks. The interbank market is interpreted as a network where banks are nodes and the claims and liabilities between banks define the links. This allows us to apply methods from general network theory. In other terms, the main feature of the financial network is a graph whose nodes represent firms or banks and whose edges represent counterparty relationships. In most models, the edges are directed, representing the direction in which contagion travels, e.g., from obligor to creditor. The edges may also be weighted, representing the magnitude of the counterparty relationship and its propensity to transmit contagion.

Recent years have witnessed technical advances in portfolio credit risk models, in theory allowing financial institutions to measure their distribution of their potential credit losses at the top level of the institution. Such information can be used to infer economic capital, which is the amount of equity capital the institution should carry to absorb a large loss over a specified horizon with a high confidence level. These credit models have been in widespread use in the financial industry, and they constitute the basis for the Basel II regulatory capital charges for commercial banks[4].

The calibration of these models is notoriously difficult. This is in large part because default correlations cannot be directly measured for specific obligors. Instead, default correlations are modeled indirectly, typically using a reduced-form model of default intensity or a structural model of the value of the firm based on a Gaussian copula. Standard models typically assume a factor structure, where correlations are induced by a common factor that can be interpreted as the state of the economy, plus possibly other factors. More precisely, factor models: i) need to simplify the correlation matrix; ii) generate joint movements in defaults; iii) defaults are driven by common risk factors (common negative shocks to cash flows); iv) conditional on these common factors, defaults are independent The Basel II regulatory capital charges are based on such factor models. This common feature largely explains why recent comparative studies of industry portfolio models show remarkable similarities in their outputs, or measures of economic capital. As reported, in Das et al. (2007), however, such models do not fully capture the clustering in default correlations, sometimes called “credit contagion”.

Inside the factor models approach, while structural models generate correlations in asset values from equity data and infer default correlations from movements in asset value below thresholds, reduced-form models generate correlations between defaults by allowing hazard rates to be stochastic and correlated with macroeconomic variables. Structural models of counterparty contagion involve modeling some quantitative details of the nodes and flows on the network, e.g., firms’ balance sheets and losses incurred on loans between

[4]The Basel II rules impose minimum levels of capital that commercial banks have to hold to guard against credit and other risks. The credit risk charge roughly corresponds to the worst credit loss over a one-year horizon at the 99.9 percent level of confidence.

firms. They model the causes, mechanism, and consequences of defaults. Correlated defaults emerge from the mechanism when the financial system is subjected to stochastic shocks. Reduced-form models avoid these details and directly provide a stochastic model of correlated defaults.

Structural models vary according to the type and channel of contagion they represent. The most commonly studied case is contagion spreading from borrower to lender, transmitting insolvency. In the graph, directed edges represent loans, pointing from borrowers to lenders. A structural model of solvency compares assets to liabilities and determines the contagious defaults that follow as a consequence of the fundamental defaults. Inside this branch, we mention:

1. Cascade models. The aim is to capture the domino effect: the default of one firm can directly cause the default of its creditor, and thus indirectly cause the default of its creditor's creditor. In cascade models, defaulted firms are like fallen dominoes: once a firm defaults, what happens to it no longer depends on what happens to other firms. For this reason, cascade models do not suffice to model the severity of defaults. Among cascade models, we mention Amini et al. (2010), Battiston et al. (2009), Gai and Kapadia (2010), and Nier et al. (2007)[5]

2. Clearing models. Clearing models of solvency allow more verisimilitude in modeling the severity of defaults and magnitude of losses, and they allow for endogenous recovery rates. They deal with the cycle by computing "clearing" vectors of asset values and recovery rates, which are as high as possible, given the defaults that can not be avoided.

In reduced form-models the objects of study and modeling are the default and the hazard rate of default. In these models, contagion causes distress, a state in which a firm's hazard rate of default is elevated. Inside the reduced-form approach, we mention:

1. Default Contagion Models. In this models there is no cascade of contagion: contagious defaults can not spread any further contagion. In continuous time, it features default clustering, e.g. multiple defaults can occur at the same time because of common risk factors[6]. These models are not able to incorporate contagion as another source of default cluster clustering.

[5] See Chapter 2

[6] See Sun et al. (2011) and Davis and Lo (2001), in Chapter 2

2. Models in which Contagion causes distress. Many models in which contagion affects the hazard rate for default are formulated in terms of Markov chains or interacting particle systems. The state of the system includes the state of each of the firms, such as default, distress, or health. In some models, it also includes common risk factors. The transition probabilities (in discrete time) or transition rates (in continuous time) associated with changes in the state of firm i naturally depend on the state of firm i and on some of the common risk factors. The graph on which contagion travels contains an edge directed from node j to node i if transition probabilities or rates associated with changes in the state of node i also depend on the state of node j. The weight associated with such an edge is related to the strength of the dependence. Physical models of interacting particle systems are relevant because of an analogy between economics and physics: firms are like particles whose states are influenced by their interactions with each other, allowing distress to spread via contagion. These models are used because of the mathematical tractability of the asymptotic behavior of an interacting particle system in the limit as the number of particles becomes large. Sometimes an assumption of homogeneity among firms is imposed to achieve mathematical tractability in an interacting particle system model or computational tractability in a Markov chain model.

Second-generation models attempt to provide structural explanations for the default clustering phenomenon. For instance, Duffie et al. (2008) estimate a "frailty" model where defaults are driven by an unobserved time-varying latent variable, which partially explains the observed default clustering. In these second generation models, excess clustering could be explained by counterparty risk, which occurs when default of one firm causes financial distress on other firms with which it has close business ties. Davis and Lo (2001) is the first theoretical work inside this branch of research.

Davis and Lo (2001) considered a model in which defaults occur either directly, or through infection by another defaulted firm, with probabilities for direct default or infection taken uniform throughout the system (or throughout sectors, assuming independence across sectors). Defaults occurring due to both, endogenous or exogenous causes were not considered in their set-up. Jarrow and Yu (2001) introduced a framework of primary and secondary firms, the former would default depending on some background stochastic process while the latter were affected by a stochastic process and the performance of the primary firms. They argued that this was a reasonable level of detail for their purposes and it also simplifies matters as there are no feedback loops in the system. Secondary firms depend only on primary firms whose performance is independent of the secondary. firms. Rogge and Schonbucher (2003) use copula functions to quantify correlations in default dynamics, and in particular to determine the impacts a defaulting obligor will have on the hazard rates of other obligors in a portfolio — conditioned on a specification of the set of countdown levels of surviving obligors and on the set of defaults that have already occurred at the given time. Another approach for modelling credit contagion dynamics was provided by Giesecke and Weber 2006 who used the well known voter process, from the theory of interacting particle systems, to model interactions between firms. They assumed a regular structure for their firms (a regular infinite hyper-cubic lattice) and focussed on the equilibrium properties of the model. The model is highly idealised and both the regularity and the symmetry of the interaction pattern have to be abandoned in order to calibrate a model of this kind to represent realistic patterns of mutual dependencies. Egloff et al. (2007) model contagion using a linear coupling of asset returns between business counterparts to describe the micro-structure of mutual dependencies. This leads to a self-consistent description of mutual dependencies in equilibrium (though an autoregressive mechanism is mentioned to capture non-equilibrium situations), which allows analytic solutions even for the case of asymmetric and heterogeneous impacts. Frey and Backhaus (2003) and Kraft and Steffensen (2007) use continuous time Markov models to describe the dynamics of transitions of the indicator variables describing rating classes of the obligors in a portfolio. The major problem here is that the state space of the system grows exponentially in portfolio-size. Frey and Backhaus reduce this problem by using a mean-field approximation for large portfolios, assuming that these portfolios contain only a small number of different sectors, and that contagion effects are homogeneous within sectors, whereas Kraft and Steffensen (2007) concentrate on small portfolios (involving 2 or 3 firms), and so-called "n-to default baskets" with small n chosen such that the dimension of the state space remains small, allowing them, among other things, to derive explicit results for loss-distributions, and also to address pricing issues in some detail.

There are a variety of techniques for modelling the correlations between firms' default behaviour, which is a major complication in credit risk modelling. The binomial expansion technique assumes independence between firms so that the number of defaults in a portfolio is described by a binomial distribution. In order to capture the effects of correlations a binomial distribution with an "effective" number of firms is assumed which is smaller than the actual number in the portfolio, but the weight given to each firm scaled so as to keep the mean number of defaults constant, while the variance of the overall number of defaults is increased. The relationship between the true number of firms and the effective reduced number is a modelling choice that depends on the diversity of the firms in terms of sectors, geographic locations or any other identifiable trait that would lead to strong correlations in default behaviour. JP Morgans' CreditMetrics approach and Credit Suisse First Financial Products CreditRisk+ uses the correlations in equity values as a surrogate for the correlations in credit quality. The structural modelling approach goes back a long way to work by Merton (1974) which directly models the dynamics of a firm's assets, with default being triggered by the asset value hitting some predetermined value. Correlations between firms are due to correlations in the dynamics of different firms' assets. This approach is very general, as it is relatively transparent to identify different driving forces of asset levels and straightforward to include them in the model.. However, it suffers from the fact that the asset level is not an observable quantity. On the other hand, the reduced form approach gives default rates for a given firm without modelling the underlying default process. Correlations are then directly introduced between the default rates. There was some discussion in the literature about whether the reduced form model could describe the true level of default correlations seen empirically. Yu (2007) seems to have answered this question in the affirmative if a suitable structure between the default rates is taken into account, while the results of Das et al. (2007) seem to imply that the reduced form model is insufficient to fully account for observed default correlations and direct contagion would indeed be required for a full explanation.

2 Counterparty Contagion and Systemic Risk

Duffie and Singleton (2003) identify five categories of risk faced by financial institutions: i) market risk: the risk of unexpected changes in market prices; ii) credit risk: the risk of changes in value due to unexpected changes in credit quality, in particular if a counterparty defaults on one of their contractual obligations; iii) liquidity risk: the risk that costs of adjusting financial positions may increase substantially; iv) operational risk: the risk that fraud, errors or other operational failures lead to loss in value; v) systemic risk: the risk of market wide illiquidity or chain reaction defaults.

The concept of systemic risk must comprise at least three ingredients. First, a triggering event. Second, the propagation of shocks through the financial system. And third, significant impact of the crisis on the macroeconomy. Propagation of shocks may happen through direct linkages between banks or indirectly, such as through the impact on the asset holdings of many banks caused by the forced sales of a few banks or through a crisis of confidence.

Andrew G. Haldane (2009)[7], identifies two phenomenon causing system risk: increasing complexity and decreasing diversity. In real world networks these two trends are observed to lead to fragility, and ring alarm bells for ecologists, engineers, geologists. Highly connected, heterogeneous networks may be robust yet fragile, by which he means that they may be resistant to average or typical shocks, yet highly susceptible to an attack that targets a highly connected or dominant node. In such networks, connections that we think of as shock absorbers may turn out to act as shock amplifiers during a crisis. There may be a sharp tipping point that separates normal behaviour from a crisis regime. Thus, a network with a fat-tailed degree distribution (i.e. where there is a significant number of highly connected nodes) may be robust to random shocks while vulnerable to shocks that preferentially target these highly connected nodes.

Financial networks generate chains of claims and at times of stress, these chains can amplify uncertainties about true counterparty exposures. In good times, counterparty risk is known to be small, and thus we are in presence of uncertainty which describes modelling situations where probabilities cannot plausibly be assigned to outcomes; in such times we might expect that stability will improve with connectivity. In bad times, counterparty risk can be large and highly uncertain, due to the complicated web and the nature of the links; risk describes situations where uncertainty can be adequately captured in a probability distribution, and we would then expect stability to decline with connectivity.

Systemic contagion that causes the failure or impairment of a large number of banks will in reality always manifest itself through a multitude of different channels, with spillover or domino effects from one to another. In the language of network science, financial networks are multiplex, meaning there are interbank links of many different types, and a contagious event that starts with one type of link will likely quickly infect all other types of links. Nonetheless, it is important to identify the basic types of shock mechanisms that we expect to find activated during a financial crisis, either as the primary cause, or else as the result of spillover effects stemming from the initial shock:

[7] Executive Director of Financial Stability at the Bank of England

- Asset Correlation: Different banks tend to hold common assets in their portfolios.
- Default Contagion: Bank deposits held in other banks can be considered as a form of interbank lending, but banking in modern times has dramatically expanded the range of interbank exposures. There is a multitude of linkage types between bank counterparties that range well beyond traditional interbank lending, to include swaps, derivatives and other securitized assets. At any moment, banks can at least in principle identify their exposures to all other banks and they also work hard to identify their expected potential exposure over different future time horizons. When a bank becomes insolvent, if it is not bailed out by a government agency, it will be forced into bankruptcy. Its creditors, including other banks, will then experience severe losses given this default, possibly losing close to 100% of their total exposure in the short term aftermath. Such shocks to creditor banks' interbank assets at the time of default of a debtor bank are the channel for default contagion. If left unchecked by government intervention, such shocks can in principle chain together like dominos to create a default cascade. Default cascades can only happen when interbank exposures are a high fraction of lending banks' equity, and Upper (2011) provides evidence that this was the case in Europe before and during the crisis, when many banks' interbank exposures exceeded their capital by factors of 5 or more. In reality, few bank defaults seem to lead to this type of contagion, mostly because of bank bailouts.
- Liquidity Contagion: Funding illiquidity is the situation of a bank with insufficient access to short term borrowing. Such banks, being short of cash or other liquid assets, will adopt a variety of strategies that can be considered as shrinking their balance sheets. They will try to access the repo[8] markets for untapped sources of collateralized borrowing. They will refuse to rollover short term loans and repo lending to other counterparties. When banks respond to funding illiquidity by curtailing a large

[8] A repurchase agreement (repo) is a form of short-term borrowing for dealers in government securities. The dealer sells the government securities to investors, usually on an overnight basis, and buys them back the following day.

For the party selling the security (and agreeing to repurchase it in the future) it is a repo; for the party on the other end of the transaction, (buying the security and agreeing to sell in the future) it is a reverse repurchase agreement.

fraction of their interbank lending, the resulting funding shocks to other banks are the channel for liquidity contagion in the system.

- Market Illiquidity and Asset Fire Sales: As Adrian, Shin (2010) discussed, in good times banks tend to create upward asset price spirals by increasing their leverage through large scale asset purchasing. This pushes up prices, and creating the illusion of even better times, and further increases in leverage. As they also discuss, the reverse is true in bad times. This tendency for distressed banks to sell assets into a depressed market creates the contagion mechanism known as an asset fire sale. A fire sale cascade proceeds through a double step mechanism: first, asset sales by distressed banks decreases prices, then marking-to-market leads to losses by other banks holding these assets.

2.1 Financial Networks

Counterparty contagion cause systemic risk. Jorion and Zhang (2009) provide quantitative evidence about counterparty risk among industrial firms and its contribution to the distribution of the system-wide severity of defaults. Upper (2011) surveys studies of national banking systems using regulatory data, primarily on interbank lending. He summarizes, *"contagion due to interbank exposures is likely to be rare. However, if it does take place, it could destroy a sizable proportion of the banking system in terms of total assets."* That is, these systems show robust yet fragile behavior with respect to contagion . The conclusions of the surveyed studies depend on how contagion is modeled, which risk factors are considered, and which method is employed for dealing with incomplete data. For example, the impact of contagion is magnified in models that include bankruptcy costs. The maximum-entropy method of dealing with incomplete network data tends to imagine that there is a large number of small interbank loans, resulting in a different impact of contagion than in a real banking system, which has fewer, larger loans (Mistrulli, 2011). Since the time of Upper's survey, Cont et al. (2010) used Brazilian data and found that contagion causes a significant proportion of the expected number of defaults. They argue that for the true importance of contagion to be evident, one must analyze events in which correlated shocks deplete the capital of many banks and cause some to default: the depleted banks are more vulnerable to contagion from the defaulting banks. Drehmann and Tarashev (2011), working with data on 20 large banks in the global financial system, also found that contagion is a significant source of systemic risk: systemic risk increased by 30% when they modeled contagion spread by interbank loans.

The theoretical literature on contagion in networks takes two approaches. The first one considers contagious effects via direct linkages. A seminal paper in this field is Allen and Gale (2000). The paper studies how the banking system responds to contagion when banks are connected under different network structures. In a setting where consumers have the Diamond and Dybvig (1983) type of liquidity preferences, banks perfectly insure against liquidity shocks by exchanging interbank deposits. The connections created by swapping deposits expose the system to contagion. The authors show that incomplete networks are more prone to contagion than complete structures. Better connected networks are more resilient to contagion since the proportion of the losses in one bank's portfolio is transferred to more banks through interbank agreements. To show this, they take the case of an incomplete network where the failure of a bank may trigger the failure of the entire banking system. They prove that, for the same set of parameters, if banks are connected in a complete structure, then the system is resilient to contagious effects.

Dasgupta (2004) also discusses how linkages between banks represented by crossholding of deposits can be a source of contagious breakdowns. Fragility arises when depositors, that receive a private signal about banks' fundamentals, may wish to withdraw their deposits if they believe that enough other depositors will do the same. To eliminate the multiplicity of equilibria the author uses the concept of global games. A unique equilibrium is isolated and this depends on the value of the fundamentals.

It is not only in the banking industry that contagion can occur. Cummins et al. (2002) show how the structure of catastrophe insurance markets can also lead to contagion. They show how the network structure limits the capacity of the insurance industry to absorb the effects of a major catastrophic event to well below the total amount of equity capital in the industry.

Parallel to this literature, there is a number of papers that make use of network techniques developed in mathematics and theoretical physics to study contagion. For instance, Eisenberg and Noe (2001) take this kind of technical approach when investigating default by firms that are part of a single clearing mechanism. First the authors show the existence of a clearing payment vector that defines the level of connections between firms. Next, they develop an algorithm that allows them to evaluate the effects that small shocks have on the system. This algorithm produces a natural measure of systemic risk based on how many waves of defaults are required to induce a given firm in the system to fail.

Similarly, Minguez-Afonso and Shin (2007) use lattice-theoretic methods to study liquidity and systemic risk in high-value payment systems, such as for the settlement of accounts receivable and payable among industrial firms, and interbank payment systems. Gai and Kapadia (2007) develop a model of contagion in financial networks and use similar techniques as the epidemiological literature on spread of disease in networks to assess the fragility of the financial system depending on the banks'capital buffers, the degree of connectivity and the liquidity of the market for failed banking assets. They find that greater connectivity reduces the likelihood of widespread default. However, shocks may have a significantly larger impact on the financial system when they occur. Moreover, the resilience of the network to large shocks depends on shocks hitting particular fragile points associated with structural vulnerabilities.

The second approach focuses on indirect balance-sheet linkages. Lagunoff and Schreft (2001) construct a model where agents are linked in the sense that the return on an agent's portfolio depends on the portfolio allocations of other agents. In their model, agents who are subject to shocks reallocate their portfolios, thus breaking some linkages. Two related types of financial crisis can occur in response. One occurs gradually as losses spread, breaking more links. The other type occurs instantaneously when forward-looking agents preemptively shift to safer portfolios to avoid future losses from contagion. Similarly, de Vries (2005) shows that there is dependency between banks'portfolios, given the fat tail property of the underlying assets, and this carries the potential for systemic breakdown. Cifuentes et al. (2005) present a model where financial institutions are connected via portfolio holdings. The network is complete as everyone holds the same asset. Although the authors incorporate in their model direct linkages through mutual credit exposures as well, contagion is mainly driven by changes in asset prices. Fragility, not only arises exogenously, from financial institutions'exposure to macro risk factors, as is the case in de Vries (2005). It also evolves endogenously, through forced sales of assets by some banks that depress the market price inducing further distress to other institutions, as in Cifuentes et al. (2004).

Complementary to the literature on network effects, Babus (2007) considers a model where banks form links with each other in order to reduce the risk of contagion. The network is formed endogenously and serves as an insurance mechanism. At the base of the link formation process lies the same intuition developed in Allen and Gale (2000):

better connected networks are more resilient to contagion. The model predicts a connectivity threshold above which contagion does not occur, and banks form links to reach this threshold. However, an implicit cost associated to being involved in a link prevents banks from forming connections more than required by the connectivity threshold.

Besides the theoretical investigations, there is a substantial interest in looking for evidence of contagious failures of financial institutions resulting from the mutual claims they have on one another. These papers use balance sheet information to estimate bilateral credit relationships for different banking systems. Subsequently, the stability of the interbank market is tested by simulating the breakdown of a single bank. Upper and Worms (2004) analyze the German banking system. Sheldon and Maurer (1998) consider the Swiss system. Cocco et al. (2005) present empirical evidence for lending relationships existing on the Portuguese interbank market. Furfine (2003) studies the interlinkages between the US banks, while Wells (2004) looks at the UK interbank market. Boss et al. (2004) provide an empirical analysis of the network structure of the Austrian interbank market and discuss its stability when a node is eliminated. In the same manner, Degryse and Nguyen (2007) evaluate the risk that a chain reaction of bank failures would occur in the Belgian interbank market. These papers find that the banking systems demonstrate a high resilience, even to large shocks. Simulations of the worst case scenarios show that banks representing less than 5% of total balance sheet assets would be affected by contagion on the Belgian interbank market, while for the German system the failure of a single bank could lead to the breakdown of up to 15% of the banking sector in terms of assets. These results heavily depend on how the linkages between banks, represented by credit exposures in the interbank market, are estimated. For most countries, data is extracted from banks' balance sheets, which can provide information on the aggregate exposure of the reporting institution vis-a-vis all other banks. To estimate bank-to-bank exposures, it is generally assumed that banks spread their lendings as evenly as possible. In effect, this assumption requires that banks are connected in a complete network. Upper (2006) contains a survey of this literature.

2.2 Random Networks

Researchers have investigated the qualitative behavior of randomly constructed financial networks in order to determine what behaviors contagion causes in financial networks, and what characteristics of the network affect them. They vary the characteristics of the stochastic model used to generate financial networks and observe the effect on systemic risk in these networks. This enables conclusions about how systemic risk depends on the characteristics that are varied, such as the average number of edges per node ("degree") or the total amount of interbank lending. Early papers on this topic (Gai and Kapadia, 2010; Iori et al., 2006; Nier et al., 2007) used simulation of random networks and random shocks that affect the networks. Then May and Arinaminpathy (2010) presented mean-field approximations to the models of Gai and Kapadia (2010) and Nier et al. (2007). Amini et al.(2010), Gleeson et al. (2011), Hurd and Gleeson (2011) have derived tools for computing the expected size of a default cascade, based on the asymptotic behavior of large networks.

The most important distinction among papers on this topic is whether they consider the initial shock to the network to be a single shock causing the default of a single node (as in most papers), independent shocks simultaneously affecting all banks (Battiston et al., 2009; Iori et al., 2006), or correlated shocks (Amini et al., 2011; Georg, 2011; Ladley, 2011). The complex systems literature on contagion often considers an initial shock to the system in which a disease or an innovation is introduced at a single node. Following this approach, many papers on contagion in financial networks have assumed that the initial shock to the network is the default of a single bank[9]. These papers focus on the resilience of the network, meaning the propensity for a small shock to lead to a small number of defaults rather than a large cascade of defaults. This is similar to the study of contagion in epidemics. The asymptotic study of the resilience of large networks to small shocks leads to the study of contagious links and vulnerable nodes. The focus on a small shock is useful if one wants to investigate the systemic impact of idiosyncratic risks, such as operational risk involving error or fraud. However, if the aim is understanding scenarios such as the recent global financial crisis in which hundreds of banks experienced fundamental defaults, then the object of study must be correlated shocks, which have a substantial probability of causing many fundamental defaults.

2.3 The Robust Yet Fragile nature of Financial Networks

Several studies find that financial networks can be "robust yet fragile" (Gai and Kapadia, 2010; Gallegati et al., 2008). Of course, robustness and fragility depend on parameters of

[9]In the study of the asymptotic behavior of large networks, the equivalent assumption is that the initial shock is the default of a small proportion of the banking system.

the system: a well-capitalized banking system with only a few, low-risk interbank exposures will be robust and not fragile. Gallegati et al. (2008) follow Watts (2002), who says a robust-yet-fragile system ”may. . . withstand many external shocks (robust), then suddenly. . . exhibit a large cascade (fragile).” Gai and Kapadia (2010) describe a robust-yet-fragile financial system as follows: ”while the probability of contagion may be low, the effects can be extremely widespread when problems occur.” In a system that is robust in the sense that there is a high probability that a shock causes no contagion (or contagion whose extent falls below a low threshold), fragility arises because of a high conditional probability that the extent of contagion is very great given that contagion occurs (or exceeds a low threshold). Robust-yet-fragile behavior that has been found in financial networks comes from some kind of non-linearity in the system's response to shocks.

Network structure provides one kind of non-linearity leading to robust-yet-fragile behavior. First, consider models in which correlated shocks affect all banks' capital. Capital has a non-linear effect on contagion (Gai and Kapadia, 2010; Nier et al., 2007) because of its effects on the sets of vulnerable nodes and contagious links in the network. Together with the distribution of shocks, this can generate robust-yet-fragile behavior. Most shocks leave the network resilient to contagion. Rare, large shocks create contagious links as well as multiple fundamental defaults, so they often trigger large cascades of defaults. This kind of non-linearity is also at the heart of robust-yet-fragile behavior in models with a single idiosyncratic shock. The robust-yet-fragile system of Gai and Kapadia (2010) has two features. It has moderately few contagious links and vulnerable nodes, so that one default is unlikely to lead to more than a few more defaults. It has many nodes that are not vulnerable to the default of only one of their creditors, but that do default if a larger number of their creditors default. Thus, the threat of further contagion conditional on the event that a cascade contains at least n defaults can be more than n times the threat of contagion from one fundamental default. In the presence of $n-1$ defaults, more nodes become susceptible to a contagious default that would be caused by the default of just one more of their obligors. Therefore most shocks lead to very small default cascades that contain only vulnerable nodes, but a few shocks lead to very large default cascades that contain also nodes that are not vulnerable to the default of a single creditor.

Risk-sharing entails another kind of non-linearity that causes robust-yet-fragile behavior. In a very simple model, Gallegati et al. (2008) observe that risk-sharing decreases the variance of loss for each firm while increasing the correlation among losses. Thus, risk-sharing decreases the expected number of defaults, but it increases the variance of the number of defaults and it increases the number of defaults in bad scenarios. In the extreme case of complete risk-sharing, there are no defaults unless the aggregate shock exceeds the capacity of the system as a whole to absorb it, in which case every firm fails. Such a system can be extremely robust yet fragile, with a very low probability of a bad outcome (any defaults), but always the worst outcome (every firm defaults) if the outcome is bad. The same phenomenon is at work in more complicated models, such as the dynamic model of uncorrelated liquidity shocks to banks due to Iori et al. (2006). Lending evolves over time as banks without enough liquidity to meet demand seek to borrow from banks that connect to them in the network and that possess excess liquidity. Interbank lending constitutes risk-sharing: it enables banks to survive liquidity shocks by borrowing, but it drains liquidity from lenders, leaving them more exposed to future liquidity shocks. If there is enough connectivity, then default is rare because banks subjected to large shocks are probably sufficiently well-connected to draw upon liquidity that exists elsewhere in the network. However, default is usually part of a large cascade, because it usually occurs only when the system as a whole has been drained of liquidity, leaving many banks vulnerable to shocks and contagion. Such a system is robust yet fragile: it is characterized by rare but large cascades of defaults.

Recent analytic results (Acemoglu et al. 2013, 2015; Eboli 2013, 2016) have shown that complete networks, i.e. networks where everybody is connected to everybody else, confirm the conjecture by Haldane (2009) that highly dense interbank networks have a 'robust-yet-fragile' nature:

"In a nutshell, interconnected networks exhibit a knife-edge, or tipping point, property. Within a certain range, connections serve as a shock-absorber. The system acts as a mutual insurance device with disturbances dispersed and dissipated. Connectivity engenders robustness. Risk-sharing – diversification – prevails. But beyond a certain range, the system can flip the wrong side of the knife-edge. Interconnections serve as shock-amplifiers, not dampeners, as losses cascade. The system acts not as a mutual insurance device but as a mutual incendiary device. Risk-spreading – fragility - prevails." Eboli (2013) shows that star-shaped networks display the same feature: the bank at the center of such a kind of network acts as a hub, distributing losses evenly among the other members of the network, in case of a crisis. In summary, the above cited works resort to highly stylised examples of financial networks and these results suggest that: highly dense and sparse but highly centralised networks have a robust-yet-fragile nature.

2.4 The Effects of Interconnectedness on Network Fragility

The characteristics of the network affect the probability of large cascades and the expected number of defaults. The most frequently studied characteristic is interconnectedness. In discussing interconnectedness, we must distinguish between increasing the size of interbank loans while leaving the network of lending relationships fixed, and increasing the degree of connectivity, i.e., increasing the number of interbank loans while decreasing their size so as to leave unchanged the total amount of aggregate interbank loans, and thus leave the banks' balance sheets constant. Aside from interconnectedness, Caccioli et al. (2011) also explore the effects of disassortativity[10] and of heterogeneity in degree and size. In heterogeneous networks, contagion does not depend just on the average degree, but also depends on the distribution of degree (Amini et al., 2010; Gai et al., 2011; Georg, 2011).

The effect of increasing the degree of connectivity on the probability of a large default cascade may be monotonic or non-monotonic, depending on the model and even on the parameters in the model. Gai and Kapadia (2010), Gai et al. (2011), and Nier et al. (2007) find a non-monotonic relationship in their cascade models with a single shock. The number of contagious defaults tends to be low if connectivity is low. This is because a network with few links has few contagious links. Increasing connectivity increases contagion up to a point by providing more links and more contagious links. Further increasing connectivity decreases contagion by decreasing the proportion of links that are contagious. The proportion of links that are contagious drops because a node with sufficiently high degree has made many small interbank loans, all of which are smaller than its capital, so they are not contagious links. Indeed, both zero connectivity and sufficiently high connectivity prevent contagion altogether in this kind of model with a single idiosyncratic shock. Nier et al. (2007) also consider a version of their model with fire sales in which they find similar non-monotonic behavior if the financial system is well-capitalized. The same effect was found by Cifuentes et al. (2005) in their clearing model with fire sales. However, Nier et al. (2007) find a monotonic increasing effect of connectivity on contagion in their cascade model with fire sales if the financial system is under-capitalized: given the parameters they used, a network with sufficiently high connectivity is very likely to

[10]Disassortativity is the tendency of nodes with low degree, like single-branch retail banks, to be linked to nodes with high degree, like money center banks

experience default of all firms due to the amplification of losses in fire sales. In a more complicated model featuring independent shocks to all banks and amplification of losses, Battiston et al. (2009) find a non-monotonic effect opposite to the non-monotonic effect described above. In their model, adding edges to a graph with few edges reduces the frequency of large default cascades by increasing diversification, but adding too many edges eventually increases the frequency of large default cascades by promoting contagion.

The results about the effect of connectivity on expected number of defaults are mixed. In their dynamic model of liquidity, Iori et al. (2006) find that increasing the degree of connectivity in a homogeneous system reduces defaults, but increasing degree in a heterogeneous system has a non-monotonic effect on the expected number of defaults. With very low connectivity, banks are nearly isolated and likely to fail, but to fail without causing contagion. In the heterogeneous system, as connectivity increases, banks become more able to withstand liquidity shocks they suffer, but become more exposed to contagion originating from shocks at other banks. When connectivity is too large in the heterogeneous system, contagion outweighs risk-sharing, so more connectivity increases the expected number of defaults. Ladley (2011) finds that the relationship between connectivity and expected number of defaults depends on the size of the shocks. If the shocks are small enough, increased connectivity decreases the expected number of defaults by dispersing the shocks so that the losses felt by nodes are too small to cause default. If the shocks are large enough, increased connectivity increases the expected number of defaults by spreading large losses more thoroughly. In either case, higher connectivity leads to more extreme robust-yet-fragile behavior: a lower probability of contagion, but a higher conditional expectation of the number of defaults given that contagion occurs.

This literature does not support a simple conclusion about the effect of interconnectedness. Interconnectedness can produce a range of good and bad effects, including diversification, risk-sharing, larger potential default cascades via counterparty contagion, and greater potential to trigger severe episodes of other kinds of contagion. The net impact of these effects on defaults and contagion depends on the model's parameters. However, one may summarize the literature by saying that it is often found that more highly interconnected financial networks are more robust yet fragile. Eboli (2016) puts forward a novel approach to the analysis of direct contagion in financial networks in which financial systems are represented as flow networks[11]. The model shows that complete networks, i.e. networks

[11]Flow Networks are directed and weighted graphs ar endowed with source nodes and sink nodes and

where everybody is connected to everybody else, confirm the conjecture by Haldane (2009): highly dense interbank networks have a robust-yet-fragile nature[12] In the fourth chapter we present the results of several numerical simulations that confirm the predictions of the model: highly dense and sparse but highly centralised networks have a robust-yet-fragile nature; sparse and decentralised networks display a vulnerable-yet-resilient behaviour.

3 Trade Credit and its role in Supply Chains: systemic risk vs insurance effect

For industrial firms, most exposures take the form of trade credit, defined as direct lending in a supplier-customer relationship. From the viewpoint of debtors, trade credit is important. Indeed, it constitutes the single most important source of external finance, representing about 20% of debtors' assets[13]. In case of default, the trade creditor will lose part of the unsecured exposure. Depending on the size of this exposure, this loss may create financial distress for the for the creditor[14]. There is a major difference between bank lending and trade credit. The on-going business of the trade creditor can be impaired by the bankruptcy of its borrower because this is often a major customer. Thus, in addition to the loss on the current credit exposure, which represents a balance-sheet measure, a client bankruptcy will affect future earnings, which is a flow, if the client cannot be replaced quickly.

The use of trade credit in supply chains creates systemic risk and puts the grounds for financial contagion, but, at the same time, provides funding and the sharing of liquidity

the propagation of losses and defaults, originated by an exogenous shock, is represented as a flow that crosses such a network.

[12] "In a nutshell, interconnected networks exhibit a knife-edge, or tipping point, property. Within a certain range, connections serve as a shock-absorber. The system acts as a mutual insurance device with disturbances dispersed and dissipated. Connectivity engenders robustness. Risk-sharing – diversification – prevails. But beyond a certain range, the system can flip the wrong side of the knife-edge. Interconnections serve as shock-amplifiers, not dampeners, as losses cascade. The system acts not as a mutual insurance device but as a mutual incendiary device. Risk-spreading – fragility – prevails. . . "

[13] See Cunat (2007). Boissay (2006) reports that the average trade debt of S&P 500 firms is around 30% to 40% of quarterly sales.

[14] In Jorion, Zhang (2009) sample, the average exposure ratio is small, at 0.32% of market value of the creditor's equity; the median is only 0.01%. Some firms have large and undiversified exposures, however, reaching 37% of equity.

risk. Empirical evidence and economic theory show that the default of a buyer does not imply the end of the commercial relation between the supplier and the defaulting client, cause is more convenient for a supplier to concede a deferral of payment to a defaulting client, rather than to push for the liquidation of its assets, i.e. its bankruptcy. For this reason, in spite of its implicit high cost, trade credit is widely used and represents an important proportion of firms finance. In other terms, the use of trade credit as a source of funding improves the resilience of a liquidity-constrained firm to unexpected financial shortages, i.e. it helps an illiquid firm to stay solvent and to maintain a level of investments which is as well as possible close to the optimal level. Trade credit contracts embed an insurance coverage against liquidity risk. The cost of this insurance is incorporated in the pricing policies set by suppliers (who often use trade credit terms to discriminate among their clients) and, as a consequence, affect the allocation of earnings among the firms that belong to a supply chain.

In the third chapter we put forward an analytical framework which models a trade credit supplier-buyer relationship with the aim to capture the insurance effect of trade credit. In specific terms:

- Trade Credit yields effects on the investment decisions of a financially constrained firms in manufacturing supply chains, with particular reference to a context of financial turmoil and credit rationing levels.

- Trade Credit enhances the resilience of firms to liquidity shocks and, through the flexibility of repayment terms, affects positively the inventory investment decisions and, consequently, the future expected levels of output and revenues.

- An insurance coverage against liquidity risk is embedded in trade credit contracts, thanks to which a financially-constrained firm suffering a liquidity shortage can maintain a level of expected inventory investment (and, as a consequence, a future expected output level) as close as possible to the optimal desired level.

The thesis is organised as follows. In chapter two, we review credit risk models and models of counterparty risk and contagion and their application in credit risk management, and compare the two primary types of models in the literature that attempt to describe default processes for debt obligations and other defaultable financial instruments, usually referred to as structural and reduced-form models. We put emphasis on their principal differences and possible conjunctions and common aspects. Furthermore, the chapter discusses challenges and possible progress to be made in closing the distance between structural and reduced-form models in modeling counterparty and credit risk, mainly inside an information based perspective. In chapter three we put forward a multi-factor model of trade credit connections in supply chains which is able to capture the insurance effect of trade credit, i.e. the insurance coverage against liquidity risk embedded in trade credit contracts, thanks to which a financially-constrained firm suffering a liquidity shortage can maintain a level of expected inventory investment (and, as a consequence, a future expected output level) as close as possible to the optimal desired level. Finally in chapter four, we introduce a network model that describes analytically the behaviour of highly stylised examples of financial networks in a stress scenario: the complete network, i.e. the most densely connected network, the star, i.e. the most centralised and sparse network, the circle, i.e. the most sparse and decentralised network. Futhermore, We present the results of several numerical simulations that we run to test the model conjectures. Such results show the effects that connectivity and centralization have on such stylised financial networks: the more a financial network is a) densely connected or b) sparse and highly centralised, the more the network has a robust-yet-fragile nature, likewise the complete and the star-shaped networks; while the more a network is sparse and decentralised – likewise the circular networks – the more it displays a vulnerable-yet-resilient behaviour.

References

[1] Acemoglu, D., A. Ozdaglar and A. Tahbaz-Salehi, 2013), Systemic Risk and Stability in Financial Networks, NBER Working Paper, No. 18727.

[2] Acemoglu, D., Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2015, Networks, shocks, and systemic risk, No. w20931. National Bureau of Economic Research.

[3] Allen, F. and Babus, A., 2008, Networks in Finance,Wharton Financial Institutions Center Working Paper 08-07. http://ssrn.com/abstract=1094883.

[4] Allen, F. and Gale, D.,2000, Financial Contagion, *Journal of Political Economy*, 108(1): 1-34.

[5] Amini, H., Cont, R., Minca, A., 2012, Stress-testing the resilience of financial networks, *International Journal of Theoretical and Applied Finance*, Volume 15, Issue 01.

[6] Amini, H., Cont, R., Minca, A., December 2010, Resilience to contagion in financial networks, *Mathematical Finance*, http://onlinelibrary.wiley.com/doi/10.1111/mafi.12051/full.

[7] Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C., Stiglitz, J. E., 2007, Credit chains and bankruptcy propagation in production networks, *Journal of Economic Dynamics and Control* 31, 2061-2084.

[8] Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C., Stiglitz, J. E., 2012, Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk, *Journal of Economic Dynamics and Control*, Volume 36, Issue 8.

[9] Boissay, Frédéric, 2006, Credit Chains and the Propagation of Financial Distress, Working Paper, European Central Bank.

[10] Boss, M., H. Elsinger, S. Thurner, and M. Summer, 2004, Network Topology of the Interbank Market, *Quantitative Finance*,Volume 4, Issue 6.

[11] Caccioli, F., Catanach, T. A., Farmer, J. D., 2012, Heterogeneity, correlations and financial contagion, *Advances in Complex Systems*, Volume 15.

[12] Cifuentes, R., Ferrucci, G., Shin, H. S., 2005, Liquidity risk and contagion. *Journal of the European Economic Association* 3 (2-3), 556-566.

[13] Cocco, J., F. Gomes, and N. Martins, 2009, Lending Relationships in the Interbank Market, *Journal of Financial Intermediation,* Volume 18, Issue 1.

[14] Cont, R., Moussa, A., Bastos e Santos, E., December 2010, Network structure and systemic risk in banking systems. URL ssrn.com/abstract=1733528.

[15] Credit Suisse First Boston 1997 Credit Risk+: A Credit Risk Management Framework Technical Document (New York) (www.csfb.com/creditrisk).

[16] Cummins, J. D., N. Doherty, and A. Lo, 2002, Can Insurers Pay for the "Big One"? Measuring the Capacity of the Insurance Market to Catastrophic Losses, *Journal of Banking and Finance,* Volume 26, Issues 2–3.

[17] Cunat, Vicente, 2007, Trade Credit: Suppliers as Debt Collectors and Insurance Providers, *Review of Financial Studies* 20(2), 491-527.

[18] Das, S.R., Duffie, D., Kapadia, N. Saita, L., 2007, Common failings: How corporate defaults are correlated, *The Journal of Finance*, Volume 62, Issue 1.

[19] Dasgupta, A., 2004, Financial Contagion Through Capital Connections: A Model of the Origin and Spread of Bank Panics, *Journal of European Economic Association*, Volume 2, Issue 6.

[20] Davis, M., Lo, V., 2001, Infectious defaults, *Quantitative Finance* 4 (1), 382-387.

[21] De Bandt, O. Hartmann, P., 2000, Systemic risk: a survey, BCE working paper n. 35

[22] De Bandt, O., Hartmann, P., Peydro , J.L., 2009, Systemic risk in banking: an update, In A.N. Berger, P. Molyneux, and et al. Wilson, J., editors, *Oxford Handbook of Banking*, Oxford University Press, Oxford, 2009.

[23] de Vries, C., 2005, The Simple Economics of Bank Fragility, *Journal of Banking and Finance*, Volume 29, Issue 4.

[24] Degryse, H., and G. Nguyen, 2007, Interbank Exposures: An Empirical Examination of Systemic Risk in the Belgian Banking System, *International Journal of Central Banking*, vol. 3 nr. 2.

[25] Dow, James, 2000, What is systemic risk? Moral hazard, initial shocks and propagation, *Monetary and Economic Studies*, December.

[26] Drehmann, M., Tarashev, N., March 2013, Measuring the systemic importance of interconnected banks, *Journal of Financial Intermediation*, Volume 22, pp 586–607.

[27] Duffie D., Singleton K, 2003. Credit risk: pricing, measurement and management. Princeton University Press.

[28] Eboli, M., 2013, A flow network analysis of direct balance-sheet contagion in financial networks, Technical report 1862, Kiel Institute for the World Economy.

[29] Eboli, M., 2016, 'Robust-yet-fragile' and 'vulnerable-yet-robust' financial networks: A flow network analysis of systemic risk, Working Paper.

[30] Egloff, D., Leippold, M., Vanini, P., 2007, A simple model of credit contagion, *Journal of Banking and Finance* 31, 2475-2492.

[31] Eisenberg, L., and T. Noe, 2001, Systemic Risk in Financial Systems, *Management Science*, Volume 47, pp 236–249

[32] Frey R., Backhaus J., 2004, Portfolio credit risk models with interacting default intensities: a markovian approach, Preprint, Department of Mathematics, University of Leipzig.

[33] Frey, R., and Backhaus, J., 2003, Interacting Defaults and Counterparty Risk: A Markovian Approach, Working Paper, Department of Mathematics, University of Leipzig.

[34] Furfine, C., 2003, Interbank Exposures: Quantifying the Risk of Contagion, Journal of Money, *Credit and Banking*, Vol. 35, No. 1 (Feb., 2003), pp. 111-128

[35] Gai, P., Haldane, A., Kapadia, S., 2011, Complexity, concentration and contagion, *Journal of Monetary Economics* 58, 453-470.

[36] Gai, P., Kapadia, S., 2010, Contagion in financial networks, Proceedings of the Royal Society A 466, 2401-2423.

[37] Gallegati, M., Greenwald, B., Richiardi, M. G., Stiglitz, J. E., 2008, The asymmetric effect of diffusion processes: risk sharing and contagion. *Global Economy Journal* 8 (3).

[38] Georg, C.-P., 2013, The effect of the interbank network structure on contagion and common shocks, *Journal of Banking & Finance*, Volume 37, Issue 7.

[39] Giesecke K., 2003, A simple exponential model for dependent defaults, *Journal of Fixed Income*, 13, 74-83.

[40] Giesecke, K., 2004, Credit RiskModelling and Valuation: An Introduction, URL: http://ssrn.com/abstract=479323

[41] Giesecke, K., Weber, S., 2004. Cyclical correlation, credit contagion, and portfolio losses. *Journal of Banking and Finance* 28, 3009-3036.

[42] Giesecke, K., Weber, S., 2006, Credit contagion and aggregate loss, *Journal of Economic Dynamics and Control* 30, 741-761.

[43] Haldane A. G., Rethinking the financial network, April 2009. Speech delivered at the Financial Student Association, Amsterdam.

[44] Hatchett, Kühn, 2006, *Effects of Economic Interactions on Credit Risk, Journal of Physics A: Mathematical and General*, Volume 39, Number 10 pp 2231–2251.

[45] Hatchett , J.P.N., Kühn, R., 2009, Credit Contagion and Credit Risk, *Quantitative Finance*, Volume 9, pp 373-382

[46] Hurd, T. R., 2015, *Contagion! The Spread of Systemic Risk in Financial Networks.* Springer

[47] Hurd, T. R., Gleeson, J. P., October 2011. A framework for analyzing contagion in banking networks, URL: http://ssrn.com/abstract=1945748

[48] Iori, G., Jafarey, S., Padilla, F. G., 2006. Systemic risk on the interbank market. *Journal of Economic Behavior and Organization* 61, 525-542.

[49] J P Morgan Global Research 1997 CreditMetricsTM: The Benchmark for Understanding Credit Risk Technical Document (New York) (www.creditmetrics.com)

[50] Jarrow, R. A., Yu, F., 2001. Counterparty risk and the pricing of defaultable securities. *Journal of Finance* 56 (5), 1765-1799.

[51] Jorion, P., Zhang, G., 2009. Credit contagion from counterparty risk. *Journal of Finance* 64 (5), 2053-2087.

[52] Jorion, P., Zhang, G., 2007, Good and Bad Credit Contagion: Evidence from Credit Default Swaps, *Journal of Financial Economics* 84(3), 860-883.

[53] Kaizoji, T., 2000, Speculative bubbles and crashes in stock markets: an interacting-agent model of speculative activity, *Physica A* 287 (3-4), 493-506.

[54] Kaizoji, T., 2001, A model of international financial crises, *Physica A* 299 (1-2), 279-293.

[55] Kiyotaki, N, and Moore, 2001, Credit Chains", Clarendon Lectures series, University of Oxford.

[56] Kiyotaki, N, and Moore, 2002, Balance Sheet Contagion, *American Economic Review*, May, vol. 92, n 2.

[57] Kraft, H., Steffensen, M., 2007, Bankruptcy, counterparty risk, and contagion. *Review of Finance* 11, 209-252

[58] Krishnamurthy, A., 2010a. Amplifcation mechanisms in liquidity crises. *American Economic Journal: Macroeconomics* 2 (3), 1-30.

[59] Ladley, D., February 2013, Contagion and risk-sharing on the inter-bank market, *Journal of Economic Dynamics and Control*, Volume 37, Issue 7.

[60] Lagunoff, R., and L. Schreft, 2001, A Model of Financial Fragility, *Journal of Economic Theory*, Volume 99, pp 220–264

[61] Lang, Larry, and René M. Stulz, 1992, Contagion and Competitive Intra-Industry Effects of Bankruptcy Announcements, *Journal of Financial Economics* 32(1), 45-60.

[62] Minguez-Afonso, G., and H. Shin, 2007, Systemic Risk and Liquidity in Payment Systems, Working Paper, London School of Economics. http://ssrn.com/abstract=1297363

[63] Mistrulli, P. E., 2011. Assessing financial contagion in the interbank market: maximum entropy versus observed interbank lending patterns. *Journal of Banking and Finance* 35, 1114-1127.

[64] Neu, Kühn, 2004, Credit Risk Enhancement in a Network of Interdependent Firms, *Physica A*, 342:639–655.

[65] Nier, E., Yang, J., Yorulmazer, T., Alentorn, A., 2007, Network models and financial stability, *Journal of Economic Dynamics and Control* 31, 2033-2060.

[66] Rogge, E., and Schonbucher, P. J., 2003, Modelling Dynamic Portfolio Credit Risk, Working Paper, Department of Statistics, Bonn University.

[67] Samanidou, E., Zschischang, E., Stauffer, D., Lux, T., 2007, Agent-based models of financial markets, *Reports on Progress in Physics* 70, 409-450.

[68] Sheldon, G., and M. Maurer, 1998, Interbank Lending and Systemic Risk: An Empirical Analysis for Switzerland, *Swiss Journal of Economics and Statistics*.

[69] Staum, J., 2012, Systemic risk components and deposit insurance premia, *Quantitative Finance*, 12

[70] Staum, J., 2012. Counterparty Contagion in Context: Contributions to Systemic Risk URL http://ssrn.com/abstract=1963459

[71] Tobias A., and Hyun S. Shin, July 2010, Liquidity and Leverage, *Journal of Financial Intermediation*, 19(3):418–437, .

[72] Upper, C. and A. Worms, 2004, Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?, *European Economic Review*, Volume 48, pp 827–849

[73] Upper, C., 2006, Contagion Due to Interbank Credit Exposures: What Do We Know, Why Do We Know It, and What Should We Know?, Working Paper, Bank for International Settlements.

[74] Upper, C., 2011, Simulation methods to assess the danger of contagion in interbank markets, *Journal of Financial Stability* 7, 111-125.

[75] Watts, D. J., 2002. A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences of the United States of America 99 (9), 5766-5771.

[76] Wells, S., 2002, U.K. Interbank Exposures: Systemic Risk Implications, *Bank of England Financial Stability Review*, December, pages 175-82.

[77] Yu, F., 2007. Correlated defaults in intensity-based models. *Mathematical Finance* 17 (2), 155-173.

Chapter 2

Reduced-form and Structural models of correlated defaults: two tales of the same phenomenon

Andrea TOTO*

february 2016

Abstract

This article reviews credit risk models and models of counterparty risk and contagion and their application in credit risk management, and compare the two primary types of models in the literature that attempt to describe default processes for debt obligations and other defaultable financial instruments, usually referred to as *structural* and *reduced-form* (or *intensity*) models. In doing this, we put emphasis on their principal differences and possible conjunctions and common aspects. Furhermore, we discuss challenges and possible progresses to be made in closing the distance between structural and reduced-form models in modeling counterparty and credit risk, mainly inside an information based perspective.

Keywords: default clustering – correlated defaults – credit risk – counterparty risk – credit contagion – reduced form models – structural models – intensity approach

*Università "G. d'Annunzio", Department of Business Economics, Pescara, Italy. I thank Mario Eboli, my PhD supervisor, for having given me suggestions and indications and for having read preliminary versions of this paper. I am the sole responsible for any mistake or shortcoming that can be found in these pages. I can be contacted via e-mail at: andrea.toto@unich.it

1 Introduction

The aim of this article is to review credit risk models and models of counterparty risk and contagion and their application in credit risk management, and to compare the two primary types of models in the literature that attempt to describe default processes for debt obligations and other defaultable financial instruments, usually referred to as *structural* and *reduced-form* (or *intensity*) models. In doing this, we put emphasis on their principal differences and possible conjunctions and common aspects. The article discusses challenges and possible progress to be made in closing the distance between *structural* and *reduced-form* models in modeling counterparty network and risks, under an information based perspective.

In the literature we can find basically two different approaches to measuring credit risk: i) model-based approaches (market models) and ii) traditional approaches (or non-model-based), based on historical data of defaults. Among market models, we can distinguish structural models and reduced-form models, which are considered the primary classes of models for modeling credit risk.

We can distinguish structural models in:

- firm-value models, based on the evolution of the issuing firms' asset values and capital structure and based on the theory of financial option pricing (Black and Scholes, 1973, and Merton, 1974) for the determination of default probability and recovery rates in the event of insolvency; the main difficulty associated with the use of these models concerns the estimation of the parameters that describes the evolution of firm value, because their magnitudes are not directly observable. Structural models have been applied in the approaches of CreditMetrics (developed by JPMorgan, based on the concept of credit migration) and KMV (consulting firm specialized in the analysis of credit risk). Firm value models explain the defaultable term structure of interest rate; they are not applicable for large portfolio of corporate bonds; the defaults are endogenous;

- first-passage time models (introduced by Black and Cox, 1976, see also Longstaff and Schwartz, 1995), who consider the possibility of default before debt maturity, if assets values fall below a certain threshold level.

Structural models originated with Black and Scholes (1973), Merton (1974) and reduced form models originated with Jarrow and Turnbull (1992), and subsequently, they were studied by Jarrow and Turnbull (1995), Duffie and Singleton (1999) among others. These models are viewed as competing (see Bielecki and Rutkowski, 2002; Rogers, 1999; Lando, 2009; Duffie, 2003), and there is a very close debate in the professional and academic literature directed to establish which class of models is best (see Jarrow et al., 2003). This debate usually revolves around default prediction and/or hedging performance.

Structural models use the evolution of firms' structural variables, such as asset and debt values, to determine the time of default. Merton's model (1974) was the first modern model of default and is considered the first structural model. In Merton's model, a firm's default occurs if, at the time of reimbursing the debt, its assets are below its outstanding debt. A second approach, within the structural framework, was introduced by Black and Cox (1976). In this approach defaults occur as soon as firm's asset value falls below a certain threshold. In contrast to the Merton approach, default can occur at any time. Structural default models provide a link between the credit quality of a firm and the firm's economic and financial conditions. Thus, defaults are endogenously generated within the model. Furthermore, the value of the firm's assets and liabilities at default will determine recovery rates.

In reduced-form models, default is treated as an unexpected event whose probability is governed by a default-intensity process. Some authors (see for example Jarrow and Protter, 2004) stress that the key aspect that distinguishes reduced-form models from structural models is that the former use only public (market) information which is fully observable by everybody. Indeed, reduced-form models do not consider the relation between default and (the true) firm value in an explicit manner. Therefore, it is argued that reduced-form models are much more useful for investors who use them for pricing and hedging, whereas the classical structural models are more appropriate for managers and for regulatory needs. The reduced form approach relies on market prices of defaultable instruments as the only source of information about the firms' credit risk. Inside the reduced form framework, intensity models represent the most extended type of models[1]. The intensity based model is designed for large portfolios of corporate bonds, it does not explain defaultable term

[1] Brody, Hughston and Macrina (2005) present an alternative reduced form model, based on the amount and precision of the information received by market participants about the firm's credit risk. Such model does not require the use of default intensities; it belongs to the reduced form approach because (as intensity models) it relies on market prices of defaultable instruments as the only source of information about the firms' credit risk.

structure of interest rate, it fits term structure of interest rate into market data and the defaults are exogenous. In contrast to structural models, the time of default in intensity models is not determined via the value of the firm, but it is the first jump of an exogenously given jump process; in other terms, differently from the structural approach, defaults are exogenously given instead of being endogenously determined. The parameters governing the default hazard rate are inferred from market data. Another difference between the two approaches refers to the treatment of recovery rates: as well as in the case of the defaults treatment, reduced models exogenously specify recovery rates, while in structural models the value of the firm's assets and liabilities at default defines recovery rates. Among the reduced form models, we can mention the CreditRisk+ approach (proposed by Credit Suisse Financial Products, it only considers default risk) in which the default event is not related to firm's capital structure and there are no assumptions about the causes that can leading to defaults. Reduced-form models ignores the knowledge of the assets values and liabilities of the company: the available information is the same available for the market. These models are widely and commonly used in the IRB approach for credit risk measurement and determination of capital requirements . However, they have important limitations and some weaknesses. For example, there is lack of data for a reliable implementation of these models and it is necessary to perform a process of model validation under Basel rules. But the main problem, which is more relevant to a reliable estimate of the credit risk, is the fact that these models fail to consider the contagion phenomenon, so they do not capture the systemic effects produced by counterparty risk, with negative consequences in terms of underestimation of the credit risk components and in terms of inadequate determination of Regulatory Capital.

Structural models of counterparty contagion involve modeling quantitative details of the nodes and flows on the network: among them, firms' and banks' balance sheet entries and losses incurred on loans between firms, or losses inside the interbank deposits structure. They model the causes, mechanism, and consequences of defaults. In the structural approach, when the financial system is subjected to stochastic shocks, the correlated defaults phenomenon emerges. Reduced-form models skip these details and directly provide a stochastic model of correlated defaults. Some models make it easy to distinguish between fundamental defaults, i.e. the defaults which would occur even in the absence of contagion, and contagious defaults, which happen because of contagion, as a direct consequence of fundamental defaults. In some reduced-form models, this distinction is not easy to make.

However, the impact of contagion can still be quantified through a comparison between two different outputs of the same investigated model. In other terms it is possible by comparing the behavior of the model to that of a model in which the parameters have been changed so that there is no contagion.

All of the models provide output that shows which firms have defaulted by a given time horizon. Some of the models also provide default times. Clearing models (a subset of the structural models set) also provide the severity of defaults, e.g., the recovery rate on loans to a defaulting firm. These models could also be considered useful under a public policy point of view, linking the outcomes in the financial system to public welfare[2].

One of the main problems in the evaluation and management of credit risk and counterparty risk is to be able to assign a correct estimate of the default probability to a counterparty or to each financial asset in a portfolio of risks. Moreover, a crucial role in the evaluation of the counterparty risks or the risks of a credit portfolio is played by the evaluation of the probabilities that many joint defaults occur in a given time interval. This phenomenon, sometimes described as "credit contagion", is called "default clustering" and means that multiple defaults can occur at the same time or that happen close to each other in time[3]. Such a problem is especially relevant in the risk management of portfolios of bank loans. Therefore, given the relevance of this kind of activities, held by all banks, this problem turns out to be crucial for the stability of the whole credit system of a country. In order to model joint defaults in a portfolio of risks, it is necessary to consider models which introduce a structural dependence among the portfolio positions. Different approaches can be used in order to obtain dependence in a large portfolio (see e.g. Lando, 2009):

1. A first approach introduces dependence among the portfolio positions by considering

[2] From a public policy perspective, it is valuable to go beyond the number of defaults, or the magnitude of the associated losses, to link the outcomes in the financial system to public welfare. A way to do this is to focus on the losses to a deposit insurer or a hypothetical government bailout fund that result from financial system outcomes. Another way is to model net new investment by banks (Aikman et al., 2009) or bank lending capacity (Pokutta and Schmaltz, 2011) as a consequence of financial system outcomes, because of the social cost of a credit crunch in which a shortage of credit has effects such as diminished investment and growth. Gai et al. (2011) mention a possible extension of their model to include a credit crunch caused by liquidity hoarding.

[3] Unexplained default clustering is a major issue for traditional credit risk models because it generates greater dispersion, or fatter tails, in the distribution of credit losses. This implies a greater likelihood of large losses and an understatement of economic capital. This could lead to a greater number of bank failures in periods of stress, or losses on CDOs that exceed worst estimates.

common or correlated risk factors which affect the default probabilities of the positions.

2. A second approach gives rise to dependence by modeling a direct contagion effect.

3. A third way to bring dependence into default events is to take into account a learning effect.

The basic idea underlying the factor models (a subset of the structural models set), which apply the first approach, is that all portfolio positions are affected by the value of some state variables which represent the fundamentals of the economy and are connected to the business cycle. Such an approach may be implemented both in reduced form models and in structural models. An interesting feature of the factor models is that, conditional on the value of the macroeconomic factors, defaults become independent events. In reduced form models the default intensities depend (usually linearly) on the value of some macroeconomic factors. The dependence of the default intensity of each position on common factors naturally induces dependence in the default probabilities. Among these models we may cite Giesecke (2003). On the other hand, in structural models it is the fundamental value of the assets that depends (usually linearly) on the value of some macroeconomic factors. Among these models we may cite Schonbucher (2000) and the widely used Creditmetrics and KMV models.

The paper is organized as follows. In the next section, we analyze the reduced form approach. In doing this, we review the most important models in this stream of literature, we briefly describe the informational and probabilistic building blocks of this approach and, finally, we analyze how default correlation and contagion mechanism are considered inside the reduced form framework. In section three we analyze the structural approach. In doing this we review the most important models in the literature, starting from Merton's model (1974), we briefly describe three subsets inside the structural approach, i.e. First Passage Models, Liquidation Process Models and State Dependent Models; finally we analyze how the default correlation is taken account in structural models. In section four we discuss the reconciliation of the two approach, under an informational perspective, starting from the vision of Jarrow, Protter (2004). Conclusions are drawn in section five.

2 Reduced-Form Models

Reduced-form models or intensity-based models (see, among others, Jarrow and Turnbull, 1995, Jarrow, Lando and Turnbull, 1997, and Duffie and Singleton, 1999), represent a recent approach to credit risk which consists in the development of models that deal with the insolvency event as a completely exogenous event, not dependent on the firm's capital structure. These models are based on the specification of an exogenous process that governs the default event: typically it is assumed as a Poisson process and often it's assumed that the recovery rate is exogenous to the model.

Reduced-form credit risk models generate default probabilities directly from market credit spreads with no assumptions regarding the source or causes of obligors' credit risk premiums. In that sense, reduced-form models might be considered as a statistical approach to credit risk. These models are rooted firmly in financial theory, as extensions to risky assets of the no-arbitrage theory that has proved so successful for interest rate modelling. Furthermore, reduced-form credit models have proven to be the most useful models for trading and hedging risky securities by broker and dealers.

The crucial aspect of the reduced-form approach is that it doesn't condition default on the value of the firm, and parameters related to the firm's value need not be estimated to implement them. In addition to that, reduced-form models introduce separate explicit assumptions on the dynamic default probability, which is modeled independently from the structural features of the firm, its asset volatility and leverage. Reduced-form models fundamentally differ from typical structural-form models in the degree of predictability of the default. A typical reduced-form model assumes that an exogenous random variable drives default and that the probability of default over any time interval is non-zero. Default occurs when the random variable undergoes a discrete shift in its level. These models treat defaults as unpredictable Poisson events.

The reduced form approach does not consider endogenous cause of defaults; rather, they rely on exogenous specifications for credit default and debt recovery. Generally speaking, reduced-form models assume an exogenous recovery rate that is independent from the default probability and the dynamics of a firm's assets, and take as basics the behavior of default-free interest rates, the recovery rate of defaultable bonds at default, as well as a stochastic process for default intensity. At each instant, there is some probability that a firm defaults on its obligations. Both this probability and the recovery rate in case of a default event may vary stochastically through time. Those stochastic processes determine the price of credit risk. Although these processes are not formally linked to the firm's asset value, there is presumably some underlying relation. The exogenous specification for credit default and debt recovery is both a strength and a weakness: while these models suffer from the lack of economic insights about default occurrence, they offer more degrees of freedom resulting in more flexibility and functionality. Such flexibility contributes to analytical tractability and ease of implementation and calibration, compared to structural models. However, reduced form models' dependence on historical data may result in good in-sample fitting properties but limited out-of-sample predictive power. Reduced form models are widely used on credit security trading floors where traders require fast computation tools to help them react to market movements quickly.

The reduced-form approach was developed precisely with the aim to avoid to model directly the firm's unobservable asset value process[4]. This was instead accomplished by modelling the price process of the firm's liabilities, for example, a zero-coupon bond issued by the firm. This approach was originated by Jarrow and Turnbull (1992, 1995), Artzner and Delbaen (1995), and Duffie and Singleton (1999). Typically, reduced-form models characterize default as the first jump time of a point process, often assumed to follow a Cox process (i.e., a doubly stochastic Poisson process). As such, the default time is usually a totally inaccessible stopping time, implying non-zero credit spreads for short maturity corporate debt. A review of the credit risk literature can be found in many books, including Ammann (2001), Bielecki and Rutkowski (2002), Duffie and Singleton (2003), and Lando (2009). A systematic study of the mathematical tools in reduced-form models is available in Elliott, Jeanblanc, and Yor (2000) and Jeanblanc and Rutkowski (2002).

The reduced form approach (inside which the "intensity approach" is the most extended one), assumes that a firms default time is inaccessible or unpredictable and driven by a default intensity that is a function of latent state variables. Jarrow, Lando, and Turnbull (1995), Duffie and Singleton (1999), and Hull and White (2001) present detailed explanations of several well known reduced-form modeling approaches. Many professionals and experts in the credit rating environment have shown the tendency to gravitate toward this modeling approach, given its mathematical tractability. They can be made consistent with the risk-neutral probabilities of default backed out from corporate bond prices or credit default swap (CDS) spreads/premia. Jarrow and Protter (2004) argue further that reduced-form models are more appropriate in an information theoretic context given that we are unlikely to have complete information about the default point and expected recovery.

[4] The unobservability of the firm's asset value process is one of the strongest limitations of the Structural Approach (see Section 3).

Central to the reduced-form models is the assumption that multiple defaults are independent conditional on the state of the economy. In reality, the default of one party might affect the default probabilities of other parties. Collin-Dufresne et al. (2010) and Jorion and Zhang (2009) find that a major credit event at one firm is associated with significant increases in the credit spreads of other firms. Giesecke (2004), Das et al. (2007), and Lando and Nielsen (2010) find that a defaulting firm can make the firms less resilient in its network of business links. These findings have important implications for the management of credit risk portfolios, where default relationships need to be explicitly modeled. The main drawback of the conditionally independent assumption of the reduced-form models is that the range of default correlations that can be achieved is typically too low when compared with empirical default correlations (see Das et al., 2007). The countermeasures to correct this weakness can be generally classified into two categories: endogenous default relationship approaches and exogenous default relationship approaches.

The endogenous approaches include the contagion (or infectious) models and frailty models. The frailty models (see Duffie et al. ,2009; Koopman et al. 2011) describe default clustering based on some unobservable explanatory variables. In variations of contagion or infectious type models (see Davis and Lo, 2001; Jarrow and Yu, 2001), the assumption of conditional independence is relaxed and default intensities are implemented depending on default events of other entities. Contagion and frailty models fill an important gap but at the cost of analytic tractability[5]. They can be especially difficult to implement for large

[5] To get a realistic assessment of the impact of counterparty contagion on systemic risk, one must consider other phenomena that cause correlation or contagion. In reduced-form modeling, one must separate counterparty contagion from other sources of correlated defaults, such as common risk factors. Systematic risk factors, such as the overall health of the economy, produce correlated defaults when used in reduced-form models. In these models, the challenge is to calibrate the model to data without confounding three sources of correlation among defaults: systematic risk, frailty, and contagion. In a model without contagion, Duffie et al. (2009) show how to distinguish the effects of frailty from the effects of systematic risk factors. Their empirical study shows that there is a significant amount of default correlation that can not be explained by the systematic risk factors that have been identified so far. It is difficult to distinguish between frailty, in which an unobserved risk factor can contribute to the default of firms A and B, and contagion, in which the default of firm A can contribute to the default of firm B. In either case, the default of firm A increases the default hazard of firm B. In the case of contagion, the default of firm A directly causes an increased default hazard of rm B. In the case of frailty, by Bayes' rule, the default of firm A implies that a frailty risk factor that also a ects rm B has probably taken on a dangerously high level. For this reason, Giesecke and Kim (2011) refrain from estimating the effects of frailty and contagion separately, and class them together into "spillover hazard." Failure to account properly for systematic and frailty risk

portfolios. The exogenous approaches (see Li , 2000; Laurent and Gregory, 2005; Hull and White; 2004; Brigo et al., 2011) attempt to link marginal default probability distributions to the joint default probability distribution through some external functions. Due to their simplicity in use, the exogenous approaches become very popular in practice.

The reduced form approach provides a stochastic model of correlated defaults. In some reduced-form models, default is the sole object of study while, in others, the hazard rate of default is also modeled. In these models, contagion causes a distress state, a state in which a firm's hazard rate of default is elevated (Davis, 2011). Davis and Lo (2001) have a single-period reduced-form model of contagious default.

In their model, let D_i be the default indicator of firm i, which is 1 if firm i defaults and 0 otherwise. Let D_i' be the fundamental default indicator of firm i. First, the fundamental default of each firm i happens with probability p_i. Second, given that firm i does not have a fundamental default, then its contagious default happens with probability $1 - \Pi_j D_i' q_{ji}$, where q_{ji} is the probability that the fundamental default of firm j causes a contagious default of firm i. The graph on which contagion travels contains an edge directed from node j to node i if the weight q_{ji} is positive. In this model, unlike many other models, there is no cascade of contagion: contagious defaults can not spread any further contagion.

Such a model can be extended to continuous time. The extended version of the model consider the default clustering phenomenon, meaning that multiple defaults can occur at the same time. Sun et al. (2011) produced an efficient simulation of a continuous-time model in which default clustering occurs because of common risk factors. Anyway, their model would need to extend in order to incorporate contagion as another source of default clustering. In continuous-time models, there are no simultaneous defaults, and default clustering refers to defaults that happen close to each other in time.

In many reduced form models contagion affects the hazard rate for default; they are formulated in terms of Markov chains (see, e.g., Nelson, 2002) or interacting particle sys-

factors contributes to over-estimation of contagious effects, because one relies on contagion to explain all of the correlation among defaults, some of which is really due to other causes. Azizpour et al. (2015) show how to distinguish between contagion, frailty, and systematic risk. The history of defaults and systematic risk factors, which are observable, can be used to forecast defaults. Because the modeler does not observe the frailty risk factors, they can not be used to forecast defaults. Within a model of default hazards, this allows for distinction among the e ects of systematic risk, the effects of default contagion as the component of default correlation that is explained by history but not by systematic risk, and the effects of frailty as the component of default correlation that is not explained by history.

tems (see, e.g., Liggett, 1985). The state of the system includes the state of each of the firms, such as default, distress, or health. In some models, it also includes common risk factors. The transition probabilities (in discrete time) or transition rates (in continuous time) associated with changes in the state of firm i naturally depend on the state of firm i and on some of the common risk factors. The graph on which contagion spreads contains an edge directed from node j to node i if transition probabilities or rates associated with changes in the state of node i also depend on the state of node j. The weight associated with such an edge is related to the strength of the dependence. Physical models of interacting particle systems are relevant because they suppose analogy between economics and physics: firms are like particles whose states are influenced by their interactions with each other, allowing distress to spread via contagion[6]. These models are used because of the mathematical tractability of the asymptotic behavior of an interacting particle system in the limit as the number of particles becomes large. In some models an assumption of homogeneity among firms is considered necessary to achieve mathematical tractability in an interacting particle system model or computational tractability in a Markov chain model.

In the discrete-time Markov chain model of Egloff et al. (2007), the state of a firm is its credit rating, default being the lowest rating. The probability of a firm's transition to a lower credit rating can be increased by bad levels of common risk factors and by low credit ratings of those firms that can transmit contagion to it. These firms can be thought of as "neighbors" in a graph. This model is related to interacting particle systems with local interactions characterized by a graph. For mathematical tractability in such models, it is necessary to assume that the graph has a simple structure.

As a matter of fact, firms seem to show correlation in the default probabilities due not only to a dependence on common risk factors, but also to some firm-specific risks. The empirical evidence shows the existence of clustering of default events in recession periods which cannot be explained only by the dependence induced by common factors; on this subject see for example Jarrow and Yu (2001) and Giesecke and Weber (2004). Giesecke and Weber (2004, 2006) use an infinite lattice as an undirected graph of contagion, and assume that if one firm can receive contagion from another firm, it can also transmit contagion to that firm. Their model is based on the long-run behavior of a continuous-time Markov

[6]There is an econophysics literature devoted to analyzing phenomena such as bubbles and crashes (Kaizoji, 2000) and contagion (Kaizoji, 2001) using models of interacting particle systems, in which particles are nancial agents and interactions take the form of emulation. Samanidou et al. (2007) review econophysics research on agent-based models.

process in which the state of a firm is healthy or distressed. Similar to the model of Egloff et al. (2007), transitions depend on common risk factors and distress among neighboring firms.

Das, Duffie and Kapadia (2007) develop statistical tests of the standard doubly stochastic model of default, under which the default times are correlated only through the factors determining their default intensities. The results of the application of these tests to data on U.S. corporations in the period 1987-2000 suggest that either there is a mis-specification of the default intensities or such a model is not able to account for all the dependence exhibited by the data.

On the whole, the outcomes of the empirical studies carried out suggest that the idiosyncratic or firm-specific risk should be described not only by a residual noise term but it should also include a term which takes into account the business connections among different firms. This is what direct contagion models do. Following Jarrow and Yu (2001), a counterparty risk can be modeled explicitly[7].

Jarrow and Yu (2001), Giesecke and Weber (2004) and Frey and Backhaus (2004), develop mixed models, all in a reduced form framework, that consider the contagion effect due to the counterparty risk jointly with the macroeconomic effect induced by the dependence on common factors: in other terms these models consider both a macroeconomic term and a microeconomic contagion effect. Similar to the model of Egloff et al. (2007), transitions depend on common risk factors and distress among neighboring firms. Instead, the microeconomic contagion effect is used alone in Davis and Lo (2001) model to account for the dependence in the default probabilities.

Yu (2007), extends the work of Jarrow and Yu (2001); it has a framework in which the hazard rate for default of each node can depend on common risk factors and on the default history, i.e., the times of previous defaults and identities of the nodes that defaulted. Frey and Backhaus (2008), Herbertsson and Rootzèn (2008), and Kraft and Steffensen (2007) set up the contagion model in terms of a Markov chain. The Markov property means that hazard rates depend only on which firms have defaulted, not on the time of default. In their numerical investigation, Herbertsson and Rootzèn (2008) find that a non-homogeneous model of contagion could be well approximated by a homogeneous model,

[7] The counterparty risk can be defined as the risk that the default of a firm's counterparty might affect its own default probability. By explicitly introducing the counterparty risk in the model, which can either be a reduced form or a structural model, an additional source of dependence is taken into consideration.

for which computation is much easier. Davis (2011) contains further references on this type of model. In a different kind of model, the time of default matters because a default raises the hazard rate of default for other firms, but only temporarily: the effect of a default decays over time. This promotes stronger default clustering. Giesecke et al. (2011) provide efficient simulation algorithms for such models.

The empirical testing of reduced-form models is still nascent. The reason relates back to the lack of theoretical guidance on characterizing the default intensity process. Duffee (1999) found that the parameter estimates using a square-root process of intensity can be fairly unstable. Another reason is that the bond data, on which these models are usually calibrated, are typically indicative in their nature, creating data problems as information slowly leaks into the price; this may produce misleading results. Sources of bond data continue to be plagued by missing and mistaken data. A final reason involves the difficulty in empirically separating the merits of the modeling framework and the quality of the underlying data given that bond data are typically used to fit the model as well as test the model. Structural models based on equity price data will not suffer from this difficulty when they are then tested on bond data. The recent availability of credit default swap data provides a new opportunity to understand the power of both the structural and reduced-form modeling frameworks.

2.1 Default Correlation in Reduced-Form Models

The dependence between defaults caused by common factors is captured by the standard reduced form credit risk models such as Lando (1998) or Duffie and Singleton (1999). Counterparty risk on the other hand has only recently attracted attention in the credit risk literature. Jarrow and Yu (2001) is the first paper, where the impact of defaults on the default probabilities of surviving firms is explicitly modelled; see also Davis and Lo (2001) for a related approach. Mathematical aspects of the Jarrow-Yu model were discussed among others in Bielecki and Rutkowski (2002). Finally, Giesecke and Weber (2006) use the voter model, which is well-known in the literature on interacting particle systems (see for instance Liggett (1985)), to model interaction between defaults. They come up with a model for the loss distribution of a given portfolio, which is constructed as a mixture of the equilibrium distributions of the voter model.

Modeling the default dependence between firms in the reduced-form approach is a different problem with respect to calculate the survival or default probability of a given firm in a given time interval. The problem under investigation concerns the default or survival probability of more than one firm. If we are currently at time t ($0 \leq t \leq T$) and no default has occurred so far: what is the probability that $n \geq 1$ different firms default before time T?, or, what is the probability that they all survive until time T?

Schönbucher (2003), again, points out some properties that any good approach to model dependent defaults should verify.

1. the model must be able to produce default correlations of a realistic magnitude.
2. it has to do it by keeping the number of parameters introduced to describe the dependence structure under control, without growing dramatically with the number of firms.
3. it should be a dynamic model, able to model the number of defaults as well as the timing of defaults.
4. since it is clear from the default history that there are periods in which defaults may cluster, the model should be capable of reproducing these periods.
5. the easier the calibration and implementation of the model, the better.

We can distinguish three different approaches to model default correlation in the literature of intensity credit risk modeling.

1. The first approach introduces correlation in the firms' default intensities making them dependent on a set of common variables X_t and on a firm specific factor. These models are known as conditionally independent defaults (CID) models, because they are conditioned to the realization of the state variables X_t the firm's default intensities are independent as are the default times that they generate. The main drawback of these models is that they do not generate sufficiently high default correlations. However, Yu (2002a) indicates that this is not a problem of the model itself, but rather a signal of the lack of sophistication in the choice of the state variables. Two direct extensions of the CID approach try to introduce more default correlation in the models. One is the possibility of joint jumps in the default intensities (Duffie and Singleton 1999) and the other is the possibility of default-event triggers that cause joint defaults (Duffie and Singleton 1999, Kijima 2000, and Kijima and Muromachi 2000).

2. The second approach to model default correlation, contagion models, relies on the works by Davis and Lo (1999) and Jarrow and Yu (2001). It is based on the idea of default contagion in which, when a firm defaults, the default intensities of related firms jump upwards. In these models default dependencies arises from direct links between firms. The default of one firm increases the default probabilities of related firms, which might even trigger the default of some of them.

3. The last approach to model default correlation makes use of copula functions[8].

[8]Copula is one of the possible ways to describe a dependence of random variables. Formally, copula is the joint multivariate distribution function for a multivariate distribution with standard uniform marginal distributions. Any joint multivariate distribution function can be expressed as a copula with arguments equal to univariate marginal distribution functions of respective variables. More precisely, a copula is a function that links univariate marginal distributions to the joint multivariate distribution with auxiliary correlating variables. To estimate a joint probability distribution of default times, we can start by estimating the marginal probability distributions of individual defaults, and then transform these marginal estimates into the joint distribution using a copula function. Copula functions take as inputs the individual probabilities and transform them into joint probabilities, such that the dependence structure is completely introduced by the copula. The copula approach separates individual default probabilities from the credit risk dependence structure. The copula function takes as inputs the marginal probabilities and introduce the dependence structure to generate joint probabilities. Copulas were introduced in 1959 and have been extensively applied to model, among others, survival data in areas such as actuarial science. A copula function transforms marginal probabilities into joint probabilities. The copula function takes as inputs the marginal probabilities without considering how we have derived them. Thus, the intensity approach is not the only framework with which we can use copula functions to model the default dependence structure between firms. Any other approach to model marginal default probabilities, such as the structural approach, can use copula theory to model joint probabilities. Within the reduced-form approach, we can distinguish two approaches to introduce default dependence using copulas. The first one, which we will refer to as Li's approach, was introduced by Li (1999) and represents one of the first attempts to use copula theory systematically in credit risk modelling. Li's approach takes as inputs the marginal default (survival) probabilities of each firm and derives the joint probabilities using a copula function. Li (1999) considers a copula that links individual survival probabilities to model the joint survival probability. The second approach was introduced by Schönbucher and Schubert (2001), in which the idea is to link the default thresholds with a copula. The simulation of the default times in this approach is exactly the same as in Li's approach. The only difference with the SS approach is that it allows to recover the dynamics of the "real" default intensities, which include the default contagion effects implicit in the default threshold copula. In contrast to the models of Jarrow and Yu (2001) and Davis and Lo (1999), the SS approach allows the contagion effects to arise endogenously through the use of the copula. Schönbucher (2003) calls the SS approach a dynamic approach in the sense that it considers the dynamics of the "real" default intensities, as opposed to Li's approach, which only considers the dynamics of the pseudo default intensities. Galiani

The complete specification of the default correlation will be given by the joint distribution of default times. Correlation coefficients, when estimated via a risk neutral intensity model, are based on the risk neutral measure. In order to calculate the correlation coefficients using empirical default events, the correlation coefficients are obtained under the physical measure. Jarrow, Lando and Yu (2003) and Yu (2002a) provide a procedure for computing physical default correlation through the use of risk neutral intensities. Duffee (1999), Zhang (2003), Driessen (2005), propose, and estimate, different CID models.

The literature on credit risk correlation has criticized the CID approach, arguing that it generates low levels of default correlation when compared with empirical default correlations. However, Yu (2002a) suggests that this apparent low correlation is not a problem of the approach itself but a problem of the choice of state or latent variables: this is due to the inability of a limited set of state variables to fully capture the dynamics of changes in default intensities. In order to achieve the level of correlation seen in empirical data, a CID model must include among the state variables, the evolution of the stock market, corporate and default-free bond markets, as well as various industry factors. Yu (2005) argues that the default correlation in reduced-form models can be quite sensitive to the common factor structure imposed on individual default intensities. According to Yu, the problem of low correlation in may arise because of the insufficient specification of the common factor structure, which may not capture all the sources of common variation in the model, leaving them to the idiosyncratic component, which in turn would not be independent across firms.

Driessen (2005) proposes a model in which the firms' hazard rate is a linear function of two common factors, two factors derived from the term structure of interest rates, a firm idiosyncratic factor, and a liquidity factor. Yu also examines the model of Driessen (2005), finding that the inclusion of two new common factors elevates the default correlation.

Duffie and Singleton (1999) propose two ways in order to solve the low correlation problem. First, they introduce correlation to the firm's jump processes, keeping unchanged the characteristics of the individual intensities. They postulate that each firm's jump component consists of two kinds of jumps, joint jumps and idiosyncratic jumps. The joint jump process has Poisson intensity and an exponentially distributed size. The idiosyncratic jump (independent across firms) is set to have an exponentially distributed size and intensity. The second alternative considers the possibility of simultaneous defaults triggered by com-

(2003) provides a detailed analysis of the use of copula functions to price multiname credit derivatives using both a normal and t-student copula.

mon credit events, at which several obligors can default with positive probability. If given the occurrence of a common shock, the firm's default probability is less than one. This common shock is called non-fatal shock, whereas if this probability is one, the common shock is called fatal shock. In addition to the common credit events, each entity can experience default through an idiosyncratic Poisson process, which is independent across firms. Duffie and Singleton (1999) also propose algorithms to simulate default times within these two frameworks. The criticisms that the joint credit event approach has received stem from the fact that it is unrealistic that several firms default at exactly the same time, and also from the fact that after a common credit event that makes some obligors default, the intensity of other related obligors that do not default does not change at all. Duffie and Singleton (1999) model is theoretically appealing, but it present a drawback: there are not papers in the literature which carries out an empirical calibration and implementation of a model.

2.2 Contagion mechanisms in Reduced-Form Models

In CID and contagion models the specification of the individual intensities includes all the default dependence structure between firms. Contagion models take CID models one step further, introducing into the model two empirical facts: the first one is that the default of one firm can trigger the default of other related firms; the second one is that default times tend to concentrate in certain periods of time, in which the default probability of all firms is increased. The model of "joint credit events" differs from contagion mechanisms in that if a debtor does not experience a default, its intensity does not change due to the default of any related debtor. The literature of default contagion includes two approaches: the infectious defaults model of Davis and Lo (1999), and the model proposed by Jarrow and Yu (2001). The main issues to be resolved concerning these two models are associated with difficulties in their calibration to market prices.

The Davis and Lo model (1999) has two versions, a static version that only considers the number of defaults in a given time period, and a dynamic version in which the timing of default is also incorporated.[9]

In the dynamic version of the model, each firm has an initial hazard rate of $\lambda_{i,t}$, for $i = 1, ..., I$, which can be constant, time dependent or follow a CID model. When a default occurs, the default intensity of all remaining firms is increased by a factor $a > 1$,

[9] This dynamic version is introduced in Davis and Lo (2001).

called *enhancement factor*, to $a\lambda_{i,t}$. This augmented intensity remains for an exponentially distributed period of time, after which the enhancement factor disappears ($a = 1$). During the period of augmented intensity, the default probabilities of all firms increase, reflecting the risk of default contagion.

With the aim of incorporating the clustering of default in specific periods, Jarrow and Yu (2001) extend CID models to account for counterparty risk, i.e. the risk that the default of a firm may increase the default probability of other firms with which it has commercial or financial relationships. This allows them to introduce extra-default dependence in CID models to account for default clustering. In a first attempt, Jarrow and Yu assume that the default intensity of a firm depends on the status (default/not default) of the rest of the firms, i.e. symmetric dependence. However, symmetric dependence introduces circularity in the model, which they refer to as looping defaults, which makes it extremely difficult and troublesome to construct and derive the joint distribution of default times.

Jarrow and Yu restrict the structure of the model to avoid the problem of looping defaults. They distinguish between primary firms and secondary firms. First, they derive the default intensity of primary firms, using a CID model. If a primary firm defaults, this increases the default intensities of secondary firms, but not the other way around (asymmetric dependence). This model introduces a new source of default correlation between secondary firms, and also between primary and secondary firms, but it does not solve the drawbacks of low correlation between primary firms, which CID models apparently imply. Yu (2002a) and Frey and Backhaus (2003) offer a further extension of Jarrow and Yu (2001) model.

The mean-field approximation is an alternative way to introduce contagion phenomenon in reduced form models[10]. The characteristic of a mean-field approximation is that the neighborhood structure is ignored. In physical models, a force field is approximated as a constant equal to the mean taken over the space in which the particles are located. That is, particles are treated as though their behavior produce the same force everywhere. In the financial context, defaults are treated as though the identities of the firms did not matter; any default (or any default of the same size) has the same impact on a firm. That is, counterparty contagion as a local phenomenon is approximated by a global phenomenon of contagion. Mean-field approximations to contagion are used by Giesecke et al. (2011) with time-decaying impact of default and by Cvitanic et al. (2012), Dai Pra et al. (2009),

[10]Interacting particle systems may admit tractable mean-field approximation.

and Dai Pra and Tolotti (2009) with permanent impact of default.

Going one step further, it may be possible to assume that all firms contribute to and are affected by contagion in the same way. All that matters is the number, size, and timing of defaults. This makes it possible to create a "top-down" model of default, meaning that the model directly specifies the hazard rate for the next default in the system. Only the aggregate effect of contagion is visible in a top-down model. In contrast, the models discussed previously are "bottom-up," meaning that the hazard rate for default of each firm is specified, and the effect of contagion on each firm is visible. Giesecke (2008) make an introduction to top-down vs. bottom-up modeling of default

3 Structural Models

This models are called "structural" because a firm's probability of default is estimated by an examination of its capital structure as inferred from financial statements and equity market information. They provide a concise interpretive framework for understanding the factors that influence credit quality and have served as the catalyst for a great deal of academic research on credit[11]. Although structural models are popular among investors for avoiding potential defaults, their use has proved to be problematic for the valuation and hedging of credit portfolios.

Under structural models, all the relevant credit risk elements, including default probabilities and recovery at default, are a function of the structural characteristics of the firm: asset levels, asset volatility (business risk) and leverage (financial risk). The recovery rate is therefore an endogenous variable, as the creditors' payoff is a function of the residual value of the defaulted company's assets[12]. In structural models of risky debt default is trig-

[11] An extension of the structural approach is the so called Hybrid Approach. Hybrid credit models are extensions of structural models that incorporate other financial and market factors in an effort to more accurately quantify default risk. The term hybrid model is used to indicate that the model is a combination of a structural model with the statistical approach.

[12] Under the 'second generation' structural models, the recovery rate in the event of default is exogenous and independent from the firm's asset value. It is generally defined as a fixed ratio of the outstanding debt value and is therefore independent from the default probability. For example, Longstaff and Schwartz (1995) argue that, by looking at the history of defaults and the recovery rates for various classes of debt of comparable firms, one can form a reliable estimate of the recovery rate. In their model, they allow for a stochastic term structure of interest rates and for some correlation between defaults and interest rates. They find that this correlation has a significant effect on the properties of the credit spread. This approach

gered when the market value of the firm's assets falls below a certain solvency boundary. They link the default of an entity to the value of the firm through its equity price.

These models treat equity as an option to buy the company's assets, and use option pricing formulas to link the equity price, which is used as a proxy of the (generally unobservable) firm's asset value, to probability of default. The benefit of such models is that they can use the latest market prices to provide a "marked to market" probability of default for individual companies. The major shortfall of structural models is that they deliberately simplify the capital structure of a firm, meaning that these models are hardly suitable for analyzing assets that have unusual capital structures or unusual payoffs.

Structural models view a firm's liabilities as complex put options on the firm's assets. Therefore, their aim of this approach is to model the firm's liability structure and the firm's asset value process. In these models, the default time is usually characterized as the first hitting time of a firm's asset value to a given boundary, the boundary being determined by the firm's liabilities. As such, if the firm's asset value process follows a diffusion, then the default time is usually a predictable stopping time.

While the reduced form approach models credit defaults as exogenous events driven by a stochastic process (such as a Poisson jump process), the structural approach aims to provide an explicit relationship between default risk and capital structure. In a structural model, the probability of the firm's default over any horizon is derived from the model given the capital structure and the assumptions concerning the firm's value process and conditions determining default. Reduced-form models take the default process as the model's "primitive": a process is directly posited for default probability that is then calibrated to the prices of securities issued by the firm or to the prices of derivatives based on those securities. Consequently, while reduced-form models are commonly implemented using debt-market (usually bond-price) or credit derivative (credit-default swap) data, structural model implementation is typically undertaken using equity market information. Under structural models, a default event is deemed to occur for a firm when its assets reach a sufficiently low level compared to its liabilities. These models require strong assumptions on the dynamics of the firm's asset, its debt and how its capital is structured. The main advantage of structural models is that they provide an intuitive picture, as well as an endogenous explanation for default.

---

simplifies the first class of models by both exogenously specifying the cash flows to risky debt in the event of bankruptcy and simplifying the bankruptcy process.

The difficulty of using the structural approach is twofold:

1. the firm's asset value process is not directly observable and it makes empirical implementation difficult;

2. a predictable default time implies credit spreads should be near zero on short maturity corporate debt. This second implication is well known to be inconsistent with historical market credit spread data.

The basis of the structural model approach is the observation that the value of the liabilities (debt and equity) of a firm at a point in time depends on the value of the firm's assets at that point as well as the outlook concerning that value. Debt and equity are contingent claims on the firm's assets and the value of the firm's assets acts as the central driving variable in structural models. In the typical structural model, the firm's debt and equity structure is taken as given, a process is posited for the evolution of the firm's asset value, conditions that constitute "default" are specified, and debt and equity are priced off the posited process. Since the firm's value process is unobserved, implementation of structural models is commonly performed in an indirect manner using the characteristics of the firm's equity. That is, given that equity is a contingent claim on the firm's assets whose value and other properties are observed, the implied value and other properties of the firm's assets may be backed out from this information. From this implied value, then it is possible to calculate the desired output such as probability of the firm's default over any chosen horizon. One of the most successful commercial implementations of the structural model approach is that developed by KMV Corporation (now Moody's KMV) in the late 1980s. KMV's model derives default correlations from a structural model that links correlations with fundamental factors. Finally, KMV provides an analytical derivation of the asymptotic loss distribution of the portfolio at a given time horizon, assuming the bank's loan portfolio is infinitely fine grained and that all instruments in the portfolio mature within this time horizon. This distribution is characterized by high skew and leptokurtosis.

Structural models was pioneered by Black, Scholes and Merton: they employ modern option pricing theory in corporate debt valuation. Merton model was the first structural model and has served as the cornerstone for all other structural models. Credit pricing models changed forever with the insights of Black and Scholes (1973) and Merton (1974); more precisely, the structural literature on credit risk starts with the paper by Merton

(1974), who applies the option pricing theory developed by Black and Scholes (1973) to the modelling of a firm's debt. In Merton's model, the firm's capital structure is assumed to be composed by equity and a zero-coupon bond with maturity T and face value of D. The firm's equity is simply a European call option with maturity T and strike price D on the asset value and, therefore, the firm's debt value is just the asset value minus the equity value. This approach assumes a very simple and unrealistic capital structure and implies that default can only happen at the maturity of the zero-coupon bond.

The works of Black, Scholes and Merton started the literature of structural credit risk modeling. Starting from that point, many researchers have proposed extensions to Merton model, cause it has been criticized for being based on a number of simplifying assumptions. The extended structural models represent important improvements for Merton's original framework as they are more realistic and able to better align with market data (e.g., CDS spreads). We can summarize these improvements as follows:

- In Merton's framework, a company could only default at its debt maturity date. The model can be modified to allow for early defaults by specifying a threshold level such that a default event occurs when asset value A_t falls below this critical level. The methods for pricing barrier/threshold options can be applied in this setting. Such threshold level sometimes results from shareholders' optimal default strategy to maximize equity value. Extensions to Merton model along this direction were pioneered by Black and Cox, and this group of models is often referred to as First Passage Time models.

- The constant interest rate assumption is not reliable, and a stochastic interest rate model can be incorporated into Merton model or its extended versions. In this case, correlation between asset and interest rate processes can also be introduced if needed.

- Mapping all debts into a single zero-coupon bond is not always feasible. It has been shown that multiple debts with different characteristics can also be modeled using a structural approach. The Geske Compound Option model developed by Robert Geske was the first structural model of this category.

- Several more sophisticated structural models involving stochastic volatility, jump diffusion and even regime-switching methods have also been proposed. These applications can help explain market observations with higher accuracy, but they often involve a high level of analytical complexity.

Jones, Mason and Rosenfeld (1984) criticized the promise of these structural models of default by showing how these types of models systematically underestimated observed spreads. Their research reflected a sample of firms with simple capital structures observed during the period 1977 to 1981. Ogden (1987) confirmed this result finding that the Merton model under-predicted spreads over U.S. treasuries by an average of 104 basis points. KMV (Moody's KMV or MKMV) revived the practical applicability of structural models by implementing a modified structural model called the Vasicek-Kealhofer (VK) model (see Crosbie and Boh, 2003; Kealhofer, 2003a; Kealhofer, 2003b; Vasicek, 1984). This VK model is combined with an empirical distribution of distance-to-default to generate the commercially available Expected Default Frequency — or EDF — credit measure. The VK model builds on insights obtained from modifications to the classical structural model suggested by other researchers. Black and Cox (1976) model the default-point as an absorbing barrier. Geske (1977) treats the liability claims as compound options. In this framework, Geske assumes the firm has the option to issue new equity to service debt. Longstaff and Schwartz (1995) introduce stochastic interest rates into the structural model framework to create a two-factor specification. Leland and Toft (1996) consider the impact of bankruptcy costs and taxes on the structural model output. In their framework, they assume the firm issues a constant amount of debt continuously with fixed maturity and continuous coupon payments. Collin-Dufresne and Goldstein (2001) extend the Longstaff and Schwartz model by introducing a stationary leverage ratio, allowing firms to deviate from their target leverage ratio in the short run, only.

A few empirical researchers have begun to test these model extensions. Lyden and Saraniti (2000) compare the Merton and the Longstaff-Schwartz models and find that both models under-predicted spreads; the assumption of stochastic interest rates did not seem to change the qualitative nature of the finding. Eom, Helwege, and Huang (2003) find evidence contradicting conventional wisdom on the bias of structural model spreads. They find structural models that depart from the Merton framework tend to over-predict spreads for the debt of firms with high volatility or high leverage. For safer bonds, these models, with the exception of Leland-Toft, under-predict spreads.

On the commercial side MKMV offers a version of the VK model applied to valuing corporate securities, which is built on a specification of the default-risk-free rate, the market risk premium, liquidity premium, and expected recovery in the context of a structural model. The VK model framework is used to produce default probabilities defined as EDF credit measures and then extended to produce a full characterization of the value of a credit risky security. This model appears to produce unbiased, robust predictions of corporate bond credit spreads. (see Bohn, 2000 and Agrawal, Arora, and Bohn, 2004 for more details.) Some important modifications to the typical structural framework include estimation of an implicit corporate-risk-free reference curve instead of using the U.S. treasury curve. Some of the under-prediction found in the standard testing of the Merton model likely results from choosing the wrong benchmark curve in the sense that the spread over U.S. treasuries includes more than compensation for just corporate credit-risk. The assumption here is that the appropriate corporate default risk-free curve is closer to the U.S. swap curve (typical estimates are 10 to 20 basis points less than the U.S. swap curve.) The MKMV implementation of the VK model allows for a time-varying market risk premium, which materially improves the performance of the model. Other important modifications to the framework include the specification of a liquidity premium that may be associated with the firm's access to capital markets and the assumption of a time-varying expected recovery amount. All these modifications contribute to producing a more usable structural model.

The paper by Black and Cox (1976) is the first of the so-called First Passage Models (FPM). First passage models specify default as the first time the firm's asset value hits a lower barrier, allowing default to take place at any time. When the default barrier is exogenously fixed, as in Black and Cox (1976) and Longstaff and Schwartz (1995), it acts as a safety covenant to protect bondholders. Alternatively it can be endogenously fixed as a result of the stockholders' attempt to choose the default threshold which maximizes the value of the firm (cf. Leland 1994 and Leland and Toft 1996.).

Concerning interest rates, structural models have considered them both as non-stochastic processes (Black and Cox 1976, Geske 1977, Leland 1994 and Leland and Toft 1996) and as stochastic processes (Ronn and Verma 1986, Kim, Ramaswamy and Sundaresan 1993, Nielsen et al. 1993, Longstaff and Schwartz 1995, Briys and de Varenne 1997 and Hsu, Saá-Requejo and Santa-Clara 2004).

In First Passage Models, by definition, default occurs the first time the asset value goes below a certain lower threshold, i.e. the firm is liquidated immediately after the default event. In contrast with First Passage Models, in another set of models, named Liquidation Process Models (LPM), supported by more recent theoretical and empirical research, a default event does not immediately cause liquidation but it represents the beginning of a process, the liquidation process, which might or might not cause liquidation after it is completed. This practice is consistent, for example, with Chapter 11 of the US Bankruptcy Law, where firms filling for bankruptcy are granted a court-supervised favor period (up to several years) aimed at sorting out their financial problems in order to, if possible, avoid liquidation.

State Dependent Models (SDM) represent, together with LPM, two recent efforts to incorporate into structural models different real-life phenomena. Although theoretically they make good sense, there is a lack of empirical research testing its performance. SDM assume that some of the parameters governing the firm's ability to generate cash flows or its funding costs are state dependent, where states can represent the business cycle (recession vs. expansion) or the firm's external rating.

Considering the contagion mechanism, structural models vary according to the type and channel of contagion they represent. Most of them focus on the most commonly studied case: contagion spreading from borrower to lender, transmitting insolvency. In the graph, directed edges represent loans, pointing from borrower to lender. The directed edge pointing from node i to node j has weight Lij equal to the size of the loan made by node j to node i. There are also structural models of liquidity (Gai et al., 2011; Iori et al., 2006) and of liquidity and solvency simultaneously. The latter may feature a single graph of lending relationships (Georg, 2011; Pokutta and Schmaltz, 2011) or distinct graphs of loans to model solvency and credit lines to model liquidity (Muller, 2006).

A structural model of solvency compares assets to liabilities. For example, we can consider a single-period model in which there are no external liabilities, i.e., liabilities owed to entities outside the system. The total liabilities of firm i are $\overline{p_i} = \sum_j L_{ij}$. The fraction of the liabilities of firm i owned by firm j is Dij . Let e_i be the value of external assets of firm i, i.e., of liabilities owed to firm i by entities outside the system. The vector e may be random, representing shocks to the system inflicted by risk factors. Let rec_j be the recovery rate, i.e., fraction of its liabilities that firm j is able to pay: it is 1 if firm j does not default, and less than 1 if firm j defaults. Taking default into account, the total asset value of firm i is $w_i = ei + \sum_j rec_j L_{ij}$. Firm i defaults, denoted $DEF_i = 1$, if its total asset value falls short of its liabilities, $w_i < \overline{p}_i$. Its default is fundamental if it would default even in the absence of any other defaults, $ei + \sum_j L_{ij}$; otherwise, the default is contagious.

Structural models of solvency determine the contagious defaults that follow as a consequence of the fundamental defaults. The difficulty is that the vector w of asset values both determines and depends on the vectors rec of recovery rates and DEF of default indicators. Cascade models deal with this issue by breaking the cycle. Typically, the recovery rate rec_i in case of default is taken to be an exogenously specified constant, and not allowed to depend on the asset value w_i.

Cascade models capture the domino effect: the default of one firm can directly cause the default of its creditor, and thus indirectly cause the default of its creditor's creditor. In cascade models, defaulted firms are like fallen dominoes: once a firm defaults, what happens to it no longer depends on what happens to other firms. For this reason, cascade models do not suffice to model the severity of defaults. Amini et al. (2010), Battiston et al. (2012), Gai and Kapadia (2010), and Nier et al. (2007) use cascade models.

Clearing models allow for endogenous recovery rates: $rec_i = \min\left\{1; \frac{w_i}{\bar{p}_i}\right\}$, meaning that in case of default, a firm's creditors receive its entire asset value. They deal with the recovery problem by computing "clearing" vectors of asset values and recovery rates, which are as high as possible, given the defaults that can not be avoided. Clearing models of solvency allow more verisimilitude in modeling the severity of defaults and magnitude of losses.

Giesecke (2004) also implements an approach which brings dependence into default events by taking into account a learning effect. More precisely it proposes a structural model of correlated multi-firm defaults in which investors update their information set on the health status of firms as new information on the defaults arrives over time. Such an approach may be able to represent the changes in the investors' beliefs consequent to sudden default events such as the recent accounting scandals at Enron, WorldCom and Parmalat and, unfortunately, several others.

The Eisenberg and Noe (2001) model involves only liabilities of equal seniority. Elsinger (2007) extends the analysis to include multiple levels of seniority and equity. Muller (2006) adds credit lines and liquidity.

In summary, structural approach, led by Merton model, has the highly appealing feature of connecting credit risk to underlying structural variables. It provides both an intuitive economic interpretation and an endogenous explanation of credit defaults, and allows for applications of option pricing methods. As a result, structural models not only facilitate security valuation, but also address the choice of financial structure. The main disadvantage of structural models lies in the difficulty of implementation. For example, the assumption that corporate assets are continuously tradable is unrealistic, and calibrating stochastic asset processes using publicly available information is sometimes more difficult than anticipated. Furthermore, although improved structural models have addressed several limitations of earlier models, they tend to be analytically complex and computationally intensive. In general, structural models are particularly useful in areas such as counterparty credit risk analysis, portfolio/security analysis and capital structure monitoring, while the difficulty in calibration limits their presence in front-office environments. More precisely, since its intuitive economic interpretation of the model facilitates consistent discussion regarding a variety of credit risk exposures, the structural model is particularly useful for practitioners in the credit portfolio and credit risk management fields. Corporate transaction analysis is also possible with the structural model. If an analyst wants to understand the impact on credit quality of increased borrowing, share repurchases, or the acquisition of another firm, the structural model naturally lends itself to understanding the transaction's implications. In general, the ability to diagnose the input and output of the structural model in terms of understandable economic variables (e.g. asset volatility as a proxy for business risk, the market's assessment of an enterprise's value, and the market leverage) facilitates better communication among loan originators, credit analysts, and credit portfolio managers.

3.1 First Passage Models and Liquidation Process Models

First Passage Models (FPM) were introduced by Black and Cox (1976) extending the Merton model to the case when the firm may default at any time, not only at the maturity date of the debt.

Consider that the dynamics of the firm's asset value under the risk neutral probability measure P are given by the diffusion process

$$dV_t = rV_t dt + \sigma_V V_t dW_t \tag{1}$$

and that there exists a lower level of the asset value such that the firm defaults once it reaches this level. Black and Cox (1976) considered a time dependent default threshold.

FPM have been extended to account for stochastic interest rates, bankruptcy costs, taxes, debt subordination, strategic default, time dependent and stochastic default barrier, jumps in the asset value process, etc. Although these extensions introduce more realism into the model, they increment its analytical complexity. For an extensive review of FPM see Bielecki and Rutkowski (2002, Chapter 3) and references therein.

The default threshold, always positive, can be interpreted as a safety covenant of the firm's debt which allows the bondholders to take control of the company once its asset value has reached this level. The safety covenant would act as a protection mechanism for the bondholders against an unsatisfactory corporate performance. In this case, the default threshold would be deterministic, or possibly time dependent, and exogenously fixed when the firm's debt is issued. Kim, Ramaswamy and Sundaresan (1993) and Longstaff and Schwartz (1995) assume an exogenously given constant default threshold K. Longstaff and Schwartz (1995) choose a constant default threshold and point out that (p. 793) *"since it is the ratio of V_t to K, rather than the actual value of K, that plays the major role in our analysis, allowing a more general specification for K simply makes the model more complex without providing additional insight into the valuation of risky debt."*

Hsu, Saá-Requejo and Santa-Clara (2004) suggest that V_t and K do not matter directly to the valuation of default risky bonds but only through their ratio, which is a measure of the solvency of the firm. They model the default threshold as a stochastic process, which together with the stochastic process assumed for the firm's asset value, allow them to obtain the stochastic process of the ratio $\frac{V_t}{K}$. The dynamics of the ratio $\frac{V_t}{K}$ are used to price corporate bonds.

The default threshold can also be chosen endogenously by the stockholders to maximize the value of the equity. See for example Mello and Parsons (1992), Nielsen et al. (1993), Leland (1994), Anderson and Sundaresan (1996), Leland and Toft (1996) and François and Morellec (2004). The literature has also considered the possibility of negotiation processes between stockholders and bondholders when the firm goes near the point of financial distress, from which the default threshold is determined. Similarly as how we described the choice of the face-value of the zero-coupon in the Merton model, in FPM the default threshold can be calculated as a weighted average of short and long-term debts. Interest rates can be considered either as a constant or as a stochastic process.[13]

The stochasticity of interest rates allows the model to introduce correlation between asset value and interest rates, and to make the default threshold stochastic, in the cases it

[13] See Black and Cox (1976), Leland (1994), and Leland and Toft (1996) for models with constant interest rates, and see Kim, Ramaswamy and Sundaresan (1993), Nielsen et al. (1993), Longstaff and Schwartz (1995), Bryis and de Varenne (1997), Collin-Dufresne and Goldstein (2001), and Hsu, Saá-Requejo and Santa-Clara (2004) for models with stochastic interest processes.

is specified as the discounted value of the face value of the debt. Nielsen et al. (1993) and Longstaff and Schwartz (1995) consider a Vasicek process for the interest rate, correlated with the firms' asset value.

Hsu, Saá-Requejo and Santa-Clara (2004) consider both the case of independence between risk-free interest rates and the default generating mechanism (given by the dynamics of the ratio $\frac{V_t}{K}$) and the case of correlation between both processes, specifying the risk-free rate as a CIR process. They present an interesting empirical illustration of the model, covering the calibration of the risk-free rate process and the estimation of the model's parameter through the Generalized Method of Moments.

The principal drawback of FPM is the analytical complexity that they introduce, which is increased if we consider stochastic interest rates or endogenous default thresholds. This mathematical complexity makes difficult to obtain closed form expressions for the value of the firm's equity and debt, or even for the default probability, forcing us to make use of numerical procedures. In the literature we can find that the empirical testing of FPM and structural models in general has not been very successful. Zhou (1997, Abstract) indicates that *"the empirical application of a diffusion approach has yielded very disappointing results."*

Another drawback of the structural models presented before is the so-called predictability of defaults. Generally, structural models consider continuous diffusion processes for the firm's asset value and complete information about asset value and default threshold. In this setting, the actual distance from the asset value to the default threshold tells us the nearness of default, in such a way that if we are far away from default the probability of default in the short-term is close to zero, because the asset value process needs time to reach the default point. The knowledge of the distance of default and the fact that the asset value follows a continuous diffusion process makes default a predictable event, i.e. default does not come as a surprise. This predictability of defaults makes the models generate short-term credit spreads close to zero. In contrast, it is observed in the market that even short-term credit spreads are bounded from below, incorporating the possibility of an unexpected default or deterioration in the firm's credit quality. The same characteristics of the structural models that imply the predictability of default also imply predictability of recovery. In models which do not consider strategic defaults, the bondholders get the remaining value of the firm in case of default, which is precisely the value of the default threshold at default. Thus, if we assume complete information about asset value and default threshold, the recovery rate is also a predictable quantity.

The predictability of default comes from the assumption of investors' perfect knowledge of the firm's asset value and default threshold. In practice, it is not possible to deduce from the capital structure of the firm neither the value of the firm V_t, its volatility σ_V , nor the level of the default threshold. If we consider incomplete information about either the firm value process, the default threshold (or both), investors can only infer a distribution function for these processes, which makes defaults impossible to predict. These considerations can be found, among others, in Duffie and Lando (2001), Giesecke (2004) and Jarrow and Protter (2004).

While in FPM default occurs the first time the asset value goes below a certain lower threshold, i.e. the firm is liquidated immediately after the default event[14], in Liquidation Process Models (LPM) the default event does not immediately cause liquidation but it represents the beginning of a process, the liquidation process, which might or might not cause liquidation after it is completed.

There is a clear distinction between the terms "default event" and "liquidation" and in this distiction lies the difference between LPM and FPM. A default event takes place when the firm's asset value V_t goes below the lower threshold K (which can be exogenous, constant, time dependent, stochastic or endogenously derived). A default event signals the beginning of a financially distressed period, which will not necessarily lead to liquidation. Liquidation takes place when the firm is actually liquidated, its activity stops and its remainings distributed among its claimholders.

In FPM described above the default event does coincide with liquidation.

If we consider that the liquidation process can take quite a while, it implies that past information shows itself as a significant explanatory variable, together with contemporaneous information, because it comprises information about the liquidation process. Information here comprises the firms' financial variables as well as financial markets, business cycle, credit markets and default cycle indicators. Couderc and Renault (2005) use a database containing the rating history of over ten thousand firms for the period 1981-2003 and analyze, using duration models, whether past values of several financial markets (business cycle, credit markets and default cycle) are relevant in explaining default probabilities in addition to their contemporaneous values. Their results show the critical importance of past information in default probabilities.

[14] The default event corresponds to the crossing of the asset value through the lower barrier.

LPM extend FPM to account for the fact that the liquidation time takes place after (sometimes quite a lot after) the occurrence of a default event. François and Morellec (2004), Moraux (2004), and Galai, Raviv and Wiener (2005) put forward theoretical LPM.

François and Morellec (2004) consider that, after a default event, i.e. after the asset value V_t goes below the lower threshold K, a firm is liquidated if and only if V_t remains below K consecutively during a period of time of a given length d (which in their numerical simulations they take to be 2 years). If a default event happens and the asset value remains under the lower threshold for a period lower than d the liquidation process finishes and the firm continues business activities as usual[15]. The authors provide closed-form solutions for corporate debt and equity values and analyze the implications of the model for optimal leverage and credit spreads. Numerical simulations show that credit spreads are an increasing function of the length d. Financial distress refers to the situation in which $V_t < K$. The author derives closed form solutions for different claims such as equity, different seniority debts and convertible debt. In particular, the value of equity is derived as a Down and Out Parisian option written on the firm assets and a Down and Out cumulative call option. Numerical simulations show that the value of equity is an increasing function of d, and that, unlike in François and Morellec (2004), credit spreads increase or decrease with d depending on the seniority of the debt.

Galai, Raviv and Wiener (2005) represent a step forward in the refinement of LPM, proposing a model extending and including the two previous ones. Galai, Raviv and Wiener argue that in the two previous models, the only thing that matters for a firm to be liquidated is the amount of time it spends in financial distress (either successively or cumulatively), but they fail to (p. 5) *"capture the following two common features of bankruptcy procedures: (i) Recent distress events may have a greater effect on the decision to liquidate a firm's assets then old distress events. ... (ii) Severe distress events may have greater effect on the decision to liquidate a firm than mild distress events."* To account for such two stylized facts, the authors propose a structural model in which a firm is liquidated when a state variable representing the cumulative weighted time period spent by the firm in distress exceeds d. At each time, the cumulative weighted time period is computed as a weighted average of the total time spent by the firm in distress, weighted by (i) how far away in the past such distress occurred and (ii) how severe was such a distress, where

[15] The number of successfully managed past default events and liquidation periods the firms has experienced does not affect the maximum length d of future liquidation periods.

distress severity is measured as an increasing function of max $\{0, K - V\}$. The authors solve the model numerically using Monte Carlo simulation based on Parisian options.

3.2 State Dependent Models

Another branch within the structural approach consists of extending standard models with regime switching: some of the model parameters are state-contingent. States can represent the state of the business cycle or simply the firm's external rating. Cash-flows, bankruptcy costs and funding costs might be state-dependent. This branch of structural models is able to reduce the problems of predictability of defaults (and recovery) suffered by standard models because the firm is subject to exogenous changes of parameters which affect its ability to generate cash flows or its funding costs, which are the main drivers of default probabilities.

Hackbarth, Miao and Morellec (2006) put forward two different models illustrating the previous ideas. In both cases the authors provide closed form expressions for the value of equity and debt, whose solution imply solving systems of ordinary differential equations. In Hackbarth, Miao and Morellec (2006) cash flows and recovery rates depend on the state of the business cycle. Cash flows follow a geometric Brownian motion and are scaled by a business cycle scalar factor: they are higher in expansions than in recessions. In the same way, bankruptcy costs are expressed as a state-dependent fraction of the firm's assets; again, the recovery rate in expansions, α_H is higher than in recessions α_L, $\alpha_H > \alpha_L$. At each point in time, there is an exogenous probability of switching between recession and expansion. The default threshold is endogenously chosen by shareholders to maximize the value of equity, and it turns out to be higher in recessions: the firm defaults earlier in recessions than in expansions. Numerical examples illustrate the implications of the model for default thresholds, default clustering, optimal leverage (countercyclical) and credit spreads.

As described by Duffie (2005, p. 2772), *"It has become increasingly common for bond issuers to link the size of the coupon rate on their debt with their credit rating, offering a higher coupon rate at lower ratings, perhaps in an attempt to appeal to investors based on some degree of hedging against a decline in credit quality. This embedded derivative is called a 'ratings-based step-up."'* The author illustrates an example of a ratings-based step-up bond issued by Deutsche Telecom in 2002 with coupon payments linked to the firm's rating.

As well as LPM, State Dependent Models (SDM) have only been developed theoretically and their future success in credit risk modelling (if any) lies in their empirical applicability and their ability to replicate and predict credit spreads and default probabilities.

3.3 Default Correlation in Structural Models

The most natural way to introduce default dependences between firms in structural models is by correlating the firms' asset processes. Describing a firm's asset value process by geometrical Brownian motion means that defaults are perfectly predictable. This approach is used in Vasicek formula, also called the asymptotic single risk factor approach, which forms the heart of the IRB[16] formula of Basel II. Vasicek (1987) assumes the portfolio is comprised of similar borrowers with the same default probabilities. Given correlation between returns on the assets of the borrowers in the portfolio and given the level of confidence, his formula specifies the level of capital that is required to prevent the bank from going bankrupt in one year, assuming no recovery.

Huang et al (2007) suggest a higher order saddlepoint approximation for estimation of asset returns correlation in a concentrated credit portfolio. This technique allows for a more adequate estimate of obligors' returns distributions, and thus, for an enhanced estimate of portfolio credit risk. Huang et al (2007) also demonstrate that this approach can be extended for a more general multi-factor returns correlation model with stochastic loss given default[17].

Vasicek (1987 and 1991) model and its derivatives are often criticized on the grounds that these models accommodate for predictable defaults. One way of leaving out the problem of the default predictability would be to introduce jump components in the firms' asset processes. Those jump components could be either correlated or uncorrelated across

[16] Internal Rating Based approach, proposed by Basle Committee, permits the lending institutions to use their own internal measures for key drivers of credit risk as primary inputs to the capital calculation, subject to meeting certain conditions and to explicit supervisory approval. All institutions using the IRB approach will be allowed to determine the borrowers' probabilities of default while those using the advanced IRB approach will also be permitted to rely on own estimates of loss given default and exposure at default on an exposure-by-exposure basis. These risk measures are converted into risk weights and regulatory capital requirements by means of risk weight formulas specified by the Basel Committee

[17] Basel II Capital Adequacy accord and most of the industrial credit risk models use loss given default (LGD), Or recovery rate (RR), with RR = (1-LGD), as one of the building blocks for estimation of the expected loss of a credit portfolio, defining loss given default as the ratio of losses to exposure at default. For measurement and estimation approaches of the LGD, see Shuermann (2004).

firms. Correlated jump components, besides making defaults unpredictable, would also account for credit risk contagion effects. The main problem lies in the calibration of those jump components. Giesecke (2004) proposes a model in which the default thresholds are constant and known, and in which the distribution of the historical lows of firms' asset value processes are linked through a copula function.

Giesecke (2004) and Giesecke and Goldberg (2004) suggest that the default correlation implied by the use of correlated firms' asset processes accounts for the dependence of the firms' credit quality on common macroeconomic factors, what they call cyclical default correlation, but it does not account for credit risk contagion across firms and periods of default clustering. In order to introduce the contagion correlation in the model, Giesecke (2004) and Giesecke and Goldberg (2004) propose a model in which the firms' default thresholds are dependent one to each other and are unknown to investors.

Cyclical default correlation does not account for all the credit risk dependence between firms. Giesecke (2004) extends structural models for default correlation to incorporate credit risk contagion effects under incomplete information. The default of one firm can trigger the default of other related firms. Furthermore, default times tend to concentrate in some periods of time in which the probability of default of all firms is increased and which can not be totally, or even partially, explained by the firms' common dependence on some macroeconomic factors. Contagion effects can arise in this setting by direct links between firms in terms of, for example, commercial or financial relationships. The news about the default of one firm have a big impact in the credit quality of other related firms which is immediately reflected in their default probabilities.

In structural first passage models it is assumed that investors have complete information about both asset processes and default thresholds, so they always know the distance of default for each firm, i.e. the distance between the actual level of the firm's assets and its default threshold. Giesecke (2004) introduces contagion effects in the model by relaxing the assumption that investors have complete information about the default thresholds of the firms, while maintaining the assumption of complete information about the diffusion process governing the dynamics of the firms' asset process. Bond holders do not have perfect information neither about such thresholds nor about their joint distribution. However, they form a prior distribution which is updated at any time one of such thresholds is revealed, which only happens when the corresponding firm defaults. In other terms, in Giesecke (2004) investors have incomplete information about the firms' default thresholds but complete information about their asset processes. Giesecke and Goldberg (2011) extend that framework to one in which investors do not have information neither about the firms' asset values nor about their default thresholds. In this case, default correlation is introduced through correlated asset processes and, again, investors receive information about the firms' asset and default barrier only when they default. Such information is used to update their priors about the distribution of the remaining firms' asset values and default thresholds. The incomplete information about the level of the default thresholds and the fact that those levels are dependent between firms represent the source of credit risk contagion. Investors form a belief, in terms of both individual and joint distribution functions, about the level of the firms' default thresholds. Each time one of the firms default, the true level of its default threshold is revealed, and investors use this new information to update their beliefs about the default thresholds of the rest of the firms. The update of the investors' perceptions about the default thresholds of the firm, and thus about the distance of default for each firm, introduces the default contagion effects in the models. This model allows the introduction of default correlation both through dependences between firms' asset values, cyclical default correlations, and through dependences between firms' default barriers, i.e. contagion effect. The major problem of this approach is to calibrate and estimate the default threshold.

The uncertainty of the default point is discussed by Galai, Raviv and Wiener (2005). The authors argue that in practice default does not necessarily lead to immediate change of control or to liquidation of the firm's assets by its debtholders. To consider the impact of this on the valuation of corporate securities, they develop a model in which liquidation is driven by a state variable that accumulates with time and severity of distress. This model can be viewed as a generalization of the Merton's framework, in which liquidation occurs only upon debt maturity, and the Black-Cox model, in which reorganization of the firm's assets is invoked when a minimum threshold is violated during the lifetime of the debt.

4 Structural vs Reduced-Form Models: an information based perspective

Under an information based perspective, the difference between these two model types can be characterized in terms of the information assumed known by the modeler. Structural models assume that the modeler has the same information set as the firm's manager, i.e. complete knowledge of all the firm's assets and liabilities. In most situations, this informational assumption implies that a firm's default time is predictable[18]. In contrast, reduced form models assume knowledge of a less detailed information set, i.e., that the modeler has the same information set as the market, and incomplete knowledge of the firm's condition. In most cases, this informational assumption implies that the firm's default time is inaccessible[19]. The reconciliation between structural and reduced form models is based on the information structure the models assume, and different information frameworks will have different implications.

The informational perspective, which leads to this reconciliation of the two approaches, is the hypothesis of Jarrow, Potter (2004). Jarrow and Protter (2004) argue that the key distinction between structural and reduced form models is not whether the default time is predictable or inaccessible, but whether the information set is observed by the market or not. According to Jarrow, Potter (2004), the reduced form approach and the structural approach are not disconnected and disjoint model types, as is commonly supposed, but rather they are really the same model containing different informational assumptions. Given this insight, one sees that the key distinction between structural and reduced form models is not in the characteristic of the default time, i.e. if it is predictable or inaccessible, but in the information set available to the modeler. Indeed, structural models can be transformed into reduced form models as the information set changes and becomes less refined, from that observable by the firm's management to that which is observed by the market.

As a consequence of this observation, the debate in the credit risk literature about these two model types is misleading. Rather than debating which model type should be used in force of forecasting performance, the debate should be focused on what type of information set the model should assume, i.e. whether the informational assumptions should comprise the information observed by the market or not. For pricing and hedging credit risk, the information set observed by the market is the relevant one, cause this information is the one used by the market, in equilibrium, to determine prices. This belief should suggest that the reduced form approach is the most suitable one and that it should be employed. Futhermore, inside this information structure, the characteristic of the firm's default time is determined as a corollary, i.e. whether it is a predictable or totally inaccessible stopping time.

[18]Complete knowledge leads to a predictable default time. But this is not necessarily the case: an example would be where the firm's asset value follows a continuous time jump diffusion process

[19]Imperfect knowledge leads to an inaccessible default time.

In the credit risk literature, there appears to be no disagreement that the asset value process is unobservable by the market (see especially: Duan, 1994; Ericsson and Reneby, 2002, 2005 in this regard). This consensus gives support to the use of reduced form models. Futhermore, without assuming the firm's asset value as continuously observed, the available information set implies that the firm's default time is inaccessible, and that a hazard rate model should be applied.

Most structural models assume complete information. Jarrow and Protter's claim rests on the premise that a modeler only has as much information as the market; this makes the reduced-form approach more realistic. In practice, however, the complete information assumption in structural models is an approximation designed to facilitate a simpler way of capturing the various economic nuances of how a firm operates. The strength or weakness of a model should be evaluated on its usefulness in real world applications. A reduced-form model, while not compromising on the theoretical issue of complete information, suffers from other weaknesses including lack of clear economic rationale for defining the nature of the default process.

Reduced-form models are characterized by flexibility in their functional form. This flexibility is both a strength and a weakness. Given the flexible structure in the functional form for reduced-form models, fitting a narrow collection of credit spreads is straightforward. But this flexibility in functional form may result in a model with strong in-sample fitting properties, but poor out-of-sample predictive ability. Since this type of model represents a generally non-theoretic[20] characterization of default risk, diagnosing how to improve performance of these models can be challenging. Difficulties in interpretation of results are particularly acute when modeling large cross-sections of debt instruments – particularly when there is a high degree of heterogeneity in terms of credit quality.

Structural and reduced-form models are viewed as competing paradigms. However, recent work by Duffie and Lando (2001), Collin-Dufresne, Goldstein, and Helwege (2010), Cetin et al. (2004) and Jarrow and Protter (2004) point out an intrinsic connection between these two approaches.

According to Guo, Jarrow and Zeng (2005, p. 2):

"Reduced-form models can be viewed as structural models that are analyzed under different filtrations. Structural models are based on the information set available to the firm's management, which includes continuous-time observations of both asset values and liabili-

[20]Less grounded in the economics driving default than in mathematical tractability.

ties. Reduced-form models are based on the information set available to the market, typically including only partial observations of both the firm's asset values and liabilities. As shown in examples by Duffie and Lando (2001), Collin-Dufresne, Goldstein, and Helwege (2003), Cetin et al. (2004), and Jarrow and Protter (2004), it is possible to transform a structural model with a predictable default time into a reduced-form model, with a totally inaccessible default time, by formulating the so called "incomplete information". For instance, Duffie and Lando (2001) used noisy and discretely observed asset value in a continuous-time model, Collin-Dufresne, Goldstein, and Helwege (2003) used a simple form of delayed information in a Brownian motion type model."

The aim of Guo, Jarrow and Zeng (2005, p. 2) is to address the issue of "incomplete information"; the notion of "incomplete information" has not been mathematically and systematically studied. Indeed, it is unclear if incomplete information such as the "noisy information" and the "delayed information" can be unified under a proper mathematical framework. They define the notion of "delayed information", for both discrete and continuous type. They generalize the work by Duffie and Lando (2001), Collin-Dufresne et al. (2003), and characterize the existence of an intensity process for any Markov models, with and without jumps. They show that delayed information transforms a predictable default time into a totally inaccessible stopping time.

In recent years, some papers have tried to bridge the gap between the two main approaches in credit risk modelling: structural and reduced form models. They consider standard structural models under a different informational perspective, i.e. modifying informational assumptions: more precisely they suppose incomplete information. The aim is to obtain reduced form models in which the intensity of default is not given exogenously but determined endogenously within the model; the intensity of default would be a function of the firm's characteristics and the level of information that investors posses. The main distinguishing characteristic of structural models with respect to reduced form models is that the structural approach provide a link between the probability of default and the firms' fundamental financial variables: assets and liabilities. Reduced form models use market prices of the firms' defaultable instruments (such as bonds or credit default swaps) to extract both their default probabilities and their credit risk dependencies, relying on the market as the only source of information regarding the firms' credit risk structure. Reduced form models are easy to be calibrated, but are characterized by lack of link between credit risk and the information regarding the firms' financial situation incorporated in their assets and liabilities structure.

The crucial element to close the distance between the two approaches lies in the model's information assumptions. Using a specification of a structural model where investors do not have complete information about the dynamics of the processes which trigger the firm's default, it would be possible to derive a cumulative rate of default consistent with a reduced form model. Generally, in structural models, there is the assumption of complete information; due to this, investors are able to predict the arrival of default. This predictability of default implies zero short-term credit spreads for the firm's debt, which is not consistent with the short-term spreads seen in practice[21]. The assumption of complete information about asset and default processes is not realistic from the point of view of investors, since it assumes that at each moment in time investors know the true value of the firm's assets and the true value of the default threshold. Relaxing the assumption of complete information makes the default time an unpredictable event.

Reduced form models overcome the limitation related to the lack of consistency of the complete knowledge assumption specifying an exogenous default intensity which makes default an unpredictable event. Duffie and Singleton (2003) point out that *"with imperfect information, default occurs at some intensity, so one may view this structural model with imperfect information as formally equivalent to a reduced-form model that has the endogenously determined default intensity with first passage."* In contrast to classical models, the time of default in intensity models is not determined via the value of the firm, but it is the first jump of a point process. Intensity models assume the existence of an exogenously given process, the intensity of default, which represents the intensity of arrival of the default time. These models solve the problem of default predictability but they lack of an endogenous specification of default based on the firm's economic fundamentals.

The main problem with reduced form models is that the arrival of default is not modeled on any characteristic of the firm's underlying credit quality. Relaxing the complete information assumption, Duffie and Lando (2001), Çetin et al. (2004), Giesecke (2004, 2006) and Giesecke and Goldberg (2004), and Guo, Jarrow and Zeng (2005) arrive, through different although equivalent ways, to a framework which links both credit risk modelling approaches.

Duffie and Lando (2001) consider a model in which the default time is fixed by the firm's managers in order to maximize the value of the equity, as in Leland and Toft (1996),

[21] In practice it is observed that they are bounded from below away from zero.

considering a geometric Brownian motion for the asset process. Investors cannot observe the issuer's assets directly, and receive only periodic and imperfect accounting reports. Duffie and Lando derive the distribution of the firm's asset value conditional to investors' information and from it the intensity of default in terms of the conditional asset distribution and the default threshold.

Giesecke (2006) represents the case of a structural model in which investors have complete information about both the level of the firm's asset value and the default threshold. Considering a continuous process for the asset value, it poses itself inside a standard first passage model which implies predictable defaults. After that, he deals with the case of complete information about the asset value but incomplete information about the default threshold. The default threshold is a constant value, but it is not known by the investors, who are forced to work under a distribution function for the default threshold. The impossibility of observing the default threshold makes the default time an unpredictable event. Investors can calculate the pricing trend in terms of the distribution function for the threshold and the observable historical asset value. After that, Giesecke studies the cases of incomplete information for the asset value and complete information about the default threshold, and finally, incomplete information for both the asset value and the default threshold. In contrast with the previous case in which investors have incomplete information about the default threshold but complete information about the asset value process, with imperfect asset information the pricing trend, calculated in terms of the threshold distribution and the distribution for the minimum historical asset level, admits an intensity representation.

Giesecke and Goldberg (2004) consider the case in which the default barrier is random and unobserved. Investors don't know the default threshold, so they use a priori distribution for its value.

Giesecke (2004) takes the incomplete information assumption in structural model and use it to model the default correlation. He provides a structural model in which the firms' default probabilities are linked via a joint distribution for their default thresholds. There is no perfect information about neither such thresholds nor about their joint distribution. They form a prior distribution which is updated at any time one of such thresholds is revealed, which only happens when one of the firms defaults. In Giesecke (2004) investors have incomplete information about the firms' default thresholds but complete information about their asset processes. Giesecke and Goldberg (2004) extend that framework to one in which investors do not have information neither about the firms' asset values nor about their default thresholds. In this case, default correlation is introduced through correlated asset processes and investors receive information about the firms' asset and default barrier only when they default.

Çetin et al. (2004) propose another approach to link structural and reduced form models assuming that investors receive only a reduced version of the information that is available for firm's managers. They claim that the default time is a predictable event for firm's managers, since they have enough information about the firm's fundamentals. But public investors do not have access to that information, since they observe a reduced version of this information. The firm's Cash Flow is the variable which triggers default, after reaching some minimum levels during a given period of time. While firm's managers can observe the CF levels, investors can only receive information about the sign of the CF and, due to this, the default time is unpredictable for them. They derive the default intensity as seen by the market.

Resuming, examining the two approaches under an informational perspective implies that, in order to distinguish which credit risk model is applicable, structural or reduced form, one needs to understand what information set is available to the modeler. Structural models assume that the information available is that held by the firm's managers, while reduced form models assume that it is the information observable to the market. Given this perspective, the defining characteristics of these models is not the property of the default time, i.e. if it is predictable or inaccessible - but rather the information structure of the model itself.

5 Conclusions

In this paper we have reviewed the two primary types of credit risk models in the literature, referred to as *structural* and *reduced-form* (or *intensity*) models, and we have analyzed them from an information based perspective, in the style of Jarrow and Protter (2004). This perspective allows to make easier the comparison between the two approaches and it is also a good source of creation for a new stream of literature that is able to incorporate and reconcile both the approaches inside hybrid models, in order to get improvements in assessing counterparty and credit risk. In both separated streams of literature, the main objective is the default prediction; hence the debate amongst models should naturally be concentrated on it. A unitary vision of the two approaches should consider the two streams of models as being actually the same, but using different assumptions about information available to the modeler.

From the informational point of view, structural models assume complete knowledge (very detailed and continuous information), closed to that of the managers visibility. The main implication is that this type of model is very appropriate for internal scopes and also for regulatory ones (in case of commercial banks)[22]. On the other hand, the reduced form models assume knowledge of a less detailed information set, like that observed by the market (available for the investors). The main implication is that default time is unpredictable. Market information should be taken into account, and, consequently, reduced models are important in force of realistic reasons. When prices - hence firms' evolution - are determined, in equilibrium, in the market, they are drawn by the actors knowing only public info. For example, assets value process is not observable to the outsiders, more exactly is discontinuously observed, when firm discloses accounting info and other relevant issues. This view makes the reduced form models very appropriate for pricing and hedging credit risk. For a correct analysis of these models one should also take into account a big problem that they induce, comparing with the first stream: now the default arises exogenously, not endogenously, like in the structural models (called, for this characteristic, cause-effect models). This is also contradicted by reality, because reduced form models assumed that default is not linked at all with firms' characteristics, which is clearly a shortcoming of these models.

The intuition behind the reconciliation view arises as a way to improve the previous separated streams, and also to take into account their weak points and to find some way to deal with them. Basically, the target is to implement models which should be more realistic from the general outsider point of view, that deal with incomplete information or at least with relaxed assumptions about complete information, where default should be unpredictable, but also endogenously influenced. The very recent development of the topic

[22]To clarify this, we consider that it is clear that for internal needs the most appropriate models should use as much as possible the information available for the managers; the question remains "Why the regulators could (and should) use structural models?". The answer is "Because they have the right, by law, (and access) to almost the same information as insiders (managers), so they should benefit from this situation". Also, when analyzing structural models, one should also take into account a weak point: if no jumps are allowed, default time is predictable, or in other words, they theoretically suggest a zero-short term spread, which contradicts the empirically observed behavior.

can be an explanation of the complexity and importance of these models, but also can suggest that many things about these models were not revealed up to now. Reconciliation or hybrid models should be aimed to relax complete information knowledge assumption, such that starting from structural models with predictable defaults to obtain hazard rate models with inaccessible default.

Duffie and Lando (2001), Giesecke and Goldberg (2004) and Cetin et al. (2004) represent three different approaches in order to reconcile the two main approaches.

Duffie and Lando (2001) model is very similar with first passage models (the default time is fixed by the managers for maximization of firms' equity value). The investors are receiving periodic and imperfect accounting reports, and they make inferences about the firm's evolution based on these reports, and adding, obviously, their beliefs (noise). The crucial point of the paper is that, even it starts more as a structural model, it can assure an unpredictable default, like the reduced form models. The explanation of this major changing of the stopping time from predictable to unpredictable is that between the observation times, the investor cannot see the evolution of assets.

Giesecke and Goldberg (2004) assume, in accordance with structural models, continuously observed asset value. In their approach the noise is also introduced, but the manner is quite different: default barrier is a random curve, more exactly beta distributed, with height expressed in terms of firm leverage. The explanation behind it is quite intuitive: when the leverage ratio is at a recent high, then the short-term uncertainty is high also. The modeler cannot see the curve, which is independent of the underlying structural model, so the default time depends on an unobservable curve, hence is inaccessible. That way it is solved the predictability problem from the structural models.

Cetin et al. (2004) model can be viewed as an alternative approach to previous two. The authors, instead of adding noise to obscure information as in Duffie and Lando (2001), they start also with a structural model but with a modeled filtration G that is a strict sub-filtration of that available to the managers (incomplete, but correct information). The default time is the first time, after the cash flows are below zero, when cash flow remains below zero for a certain time and then doubles in absolute magnitude. In this way a totally inaccessible default time is obtained, and the point process has intensity, so this is an intensity based hazard model.

Any structural model with incomplete information admits a pricing trend, but not all admit intensity. Unpredictability of default is a necessary, but not sufficient, condition for the pricing trend to admit intensity. Also, information level determines whether the model admits intensity: when there is certainty, there is no intensity. Then, by starting with a structural model (i.e. by considering the structural model with complete info about assets and default barrier as starting point, the most "complete" in terms of quantity and quality of available information), and relaxing in different degrees the information assumptions, Giesecke and Goldberg (2004) notice that, if it is assumed incomplete information about barrier, it can be calculated the pricing trend in terms of distribution function for the barrier and the observable historical asset value. Here pricing trend does not admit an intensity of default. If it is assumed incomplete info for both assets and barrier, the pricing trend admits an intensity representation. An important finding is that, in some particular conditions, regardless that the default barrier is observable or not, a structural model with incomplete asset info admits an intensity representation. So there is not gain for a reconciliation modeler to relax on barrier part; also the existence of intensity, if it is needed in a particular model, is assured.

Another framework to develop a desirable reconciliation model could be to start the other way around, i.e. with reduced models approach. Here the modeler only knows discrete and perfect pieces of information (accounting reports at the moment when revealed, public information about the firm, also in short term after revealing). He has to make inferences for the periods he does not have information. For that reason, the modeler can use the theory and formulas developed for reduced form models, but must weight differently information, relying more on what he perfectly knows (for short period of time) and less on the inferred part, because it is altered with its own beliefs and inferences.

References

[1] Acharya, V., and Carpenter, J., 2002, Corporate Bond Valuation and Hedging with Stochastic Interest Rates and Endogenous Bankruptcy, *Review of Financial Studies* 15, 1355-1383.

[2] Agrawal, D., N. Arora, and J. Bohn, 2004, Parsimony in Practice: An EDF-based Model of Credit Spreads, White Paper, Moody's KMV.

[3] Aikman, D., Alessandri, P., Eklund, B., Gai, P., Kapadia, S., Martin, E., Mora, N., Sterne, G., Willison, M., June 2009. Funding liquidity risk in a quantitative model

of systemic stability. Working paper 372, Bank of England.

[4] Amini, H., Cont, R., Minca, A., December 2010, Resilience to contagion in financial networks, *Mathematical Finance*, http://onlinelibrary.wiley.com/doi/10.1111/mafi.12051/full.

[5] Ammann, M., 2001, *Credit Risk Valuation. Methods, Models and Applications*, second edition. Springer

[6] Anderson, R., and Sundaresan, S., 1996, Design and Valuation of Debt Contracts, *Review of Financial Studies* 9, 37-68.

[7] Anderson, R., and Sundaresan, S., 2000, A Comparative Study of Structural Models of Corporate Bond Yields, *Journal of Banking and Finance* 24, 255-269.

[8] Artzner, P. and Delbaen, F., 1995, Default risk insurance and incomplete markets, *Mathematical Finance*, Volume 5, pp 187–195.

[9] Azizpour, S., Giesecke, K., Schwenkler, G., 2015, Exploring the sources of default clustering, URL http://ssrn.com/abstract=1127792

[10] Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C., Stiglitz, J. E., 2012, Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk, *Journal of Economic Dynamics and Control*, Volume 36, Issue 8.

[11] Benzschawel, T., 2012, *Credit Risk Modelling Facts, Theory and Applications*, Risk Books

[12] Bielecki, T. R., and Rutkowski, M., 2002, *Credit Risk: Modeling, Valuation and Hedging*, Springer Finance.

[13] Black, F., and J. C. Cox, 1976, Valuing Corporate Securities: Some Effects of Bond Indenture Provisions, *Journal of Finance* 31, 351-367.

[14] Black, F., and Scholes, M., 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637-654.

[15] Bohn, J., 2000, A Survey of Contingent-Claims Approaches to Risky Debt Valuation, *Journal of Risk Finance*, Vol. 1, No. 3 (Summer), 53-78.

[16] Bohn, J., 2000, An Empirical Assessment of a Simple Contingent-Claims Model for the Valuation of Risky Debt, *Journal of Risk Finance*, Vol. 1, No. 4 (Spring), 55-77.

[17] Brigo, D., Pallavicini, A., Papatheodorou, V., 2011, Arbitrage-free valuation of bilateral counterparty risk for interest-rate products: impact of volatilities and correlations, *International Journal of Theoretical and Applied Finance*, Volume 14.

[18] Briys, E., and de Varenne, F., 1997, Valuing Risky Fixed Rate Debt: An Extension, *Journal of Financial and Quantitative Analysis* 31, 239-248.

[19] Brody, D., Hughston, L., Macrina, A., 2005, Beyond Hazard Rates: a New Framework for Credit-Risk Modelling, in *Advances in Mathematical Finance*, part of the series *Applied and Numerical Harmonic Analysis* pp 231-257, Springer

[20] Çetin, U., Jarrow, R., Protter, P., and Yildirim, Y., 2004, Modelling Credit Risk with Partial Information, *The Annals of Applied Probability*, 14, 1167-1178.

[21] Collin-Dufresne, P., and R. Goldstein, 2001, Do Credit Spreads Reflect Stationary Leverage Ratios?, *Journal of Finance*, 56, 1929-57.

[22] Collin-Dufresne, P.,Goldstein R., Helwege, J. 2010, Is credit event risk priced? Modeling contagion via the updating of beliefs, NBER Working Paper 15733, http://www.nber.org/papers/w15733

[23] Collin-Dufresne, P.,Goldstein R., Helwege, J. 2003, Are jumps in corporate bond yields priced? Modeling contagion via the updating of beliefs, University of California.

[24] Couderc, F., and Renault, O., 2005, Times-to-Default: Life Cycle, Global and Industry Cycle Impacts, FAME Research Paper No. 142.

[25] Crosbie, P., and J. Bohn, 2003, Modeling Default Risk, White Paper, Moody's KMV.

[26] Cvitanic, J., Ma, J., Zhang, J., 2012, Law of large numbers for self-exciting correlated defaults, *Stochastic Processes and their Applications*, 122.

[27] Dai Pra, P., Ruggaldier, W., Sartori, E., Tolotti, M., 2009, Large portfolio losses: a dynamic contagion model, *Annals of Applied Probability* 19, 347-394.

[28] Dai Pra, P., Tolotti, M., 2009, Heterogeneous credit portfolios and the dynamics of the aggregate losses, *Stochastics Processes and their Applications* 119, 2913-2944.

[29] Das, S.R., Duffie, D., Kapadia, N. Saita, L., 2007, Common failings: How corporate defaults are correlated, *The Journal of Finance*, Volume 62, pp 93–117.

[30] Davis, M. H. A., 2011, Contagion models in credit risk. In: Lipton, A., Rennie, A. (Eds.), *Oxford Handbook of Credit Derivatives.* Oxford University Press, New York, Ch. 9.

[31] Davis, M., and Lo, V., 1999, Modelling Default Correlation in Bond Portfolios, Working Paper, Tokyo-Mitsubitshi International.

[32] Davis, M., Lo, V., 2001, Infectious defaults, *Quantitative Finance* 4 (1), 382-387.

[33] Driessen, J., 2005, Is Default Event Risk Priced in Corporate Bonds?, *Review of Financial Studies* 18, 165-195.

[34] Duan, J. C., 1994, Maximum Likelihood Estimation using Price Data of the Derivative Contract, *Mathematical Finance* 5, 155-167.

[35] Duan, J. C., and Fulop, A., 2009, Estimating the structural credit risk model when equity prices are contaminated by trading noises, *Journal of Econometrics*, Volume pp 288–296.

[36] Duffee, G., 1999, Estimating the Price of Default Risk, *Review of Financial Studies* Vol. 12, No. 1, 1997-2026.

[37] Duffie, D., 2005, Credit risk modelling with affine processes, *Journal of Banking and Finance* 29, 2751-2802.

[38] Duffie, D., 2003, *Dynamic Asset Pricing Theory*, 3rd ed., Princeton U. Press.

[39] Duffie, D., and Lando, D., 2001, Term Structure of Credit Spreads with Incomplete Accounting Information, *Econometrica* 69, 633-664.

[40] Duffie, D., and Singleton, K. J., 1999, Modeling Term Structures of Defaultable Bonds, *Review of Financial studies*, 12 (4) pp 687-720.

[41] Duffie, D., and Singleton, K. J., 1999, Simulating Correlated Defaults, Working Paper, Graduate School of Business, Stanford University.

[42] Duffie, D., and Singleton, K. J., 2003, *Credit Risk: Pricing, Measurement and Management*, Princeton Series in Finance.

[43] Duffie, D., Eckner, A., Horel, G., Saita, L., 2009, Frailty correlated default, *Journal of Finance* 64 (5), 2089-2123.

[44] Egloff, D., Leippold, M., Vanini, P., 2007, A simple model of credit contagion, *Journal of Banking and Finance* 31, 2475-2492.

[45] Eisenberg, L., Noe, T. H., 2001, Systemic risk in financial systems, *Management Science* 47 (2), 236-249.

[46] Elliott, R. J., Jeanblanc, M., Yor, M., 2000, On Models of Default Risk, *Mathematical Finance*, 10

[47] Elsinger, H., 2007, Financial networks, cross holdings, and limited liability, URL ssrn.com/abstract=916763

[48] Embrechts, P., McNeal, A., and Straumann, D., 2002, Correlation and Dependence in Risk Management: Properties and Pitfalls, in *Risk Management. Value at Risk and Beyond*, Cambridge Univerity Press.

[49] Eom, Y. H., Helwege, J., and Huang, J. Z., 2003, Structural Models of Corporate Bond Pricing: An Empirical Analysis, *Review of Financial Studies* 17, 499-544.

[50] Ericsson, J., and J. Reneby, 2004, An Empirical Study of Structural Credit Risk Models using Stock and Bond Prices, *The Journal of Fixed Income*, Vol. 13, No. 4: pp. 38-49.

[51] Ericsson, J., and Reneby, J., 2002, The Valuation of Corporate Liabilities: Theory and Tests, Working Paper, SSE/EFI No. 445.

[52] Ericsson, J., and Reneby, J., 2005, Estimating Structural Bond Pricing Models, Journal of Business 78, 707-736.

[53] François, P., and Morellec, E., 2004, Capital Structure and Asset Prices: Some Effects of Bankruptcy Procedures, *Journal of Business* 77, 387-411.

[54] Frey, R., and Backhaus, J., 2003, Interacting Defaults and Counterparty Risk: A Markovian Approach, Working Paper, Department of Mathematics, University of Leipzig.

[55] Frey R., Backhaus J., 2004, Portfolio credit risk models with interacting default intensities: a markovian approach, Preprint, Department of Mathematics, University of Leipzig.

[56] Frey, R., Backhaus, J., 2008, Pricing and hedging of portfolio credit derivatives with interacting default intensities, *International Journal of Theoretical and Applied Finance* 11 (6), 611-634.

[57] Frey, R., McNeil, A. J., and Nyfeler, M. A., 2001, Modelling Dependent Defaults: Asset Correlations Are Not Enough!, Working Paper, Department of Mathematics, ETHZ, Zurich.

[58] Gai, P., Haldane, A., Kapadia, S., 2011, Complexity, concentration and contagion, *Journal of Monetary Economics* 58, 453-470.

[59] Gai, P., Kapadia, S., 2010. Contagion in financial networks. Proceedings of the Royal Society A 466, 2401-2423.

[60] Galai, D., Raviv, A., and Wiener, Z., 2005, Liquidation Triggers and The Valuation of Equity and Debt, Working Paper, School of Business Administration, The Hebrew University of Jerusalem.

[61] Galiani, S., 2003, Copula functions and their application in pricing and risk managing multivariate credit derivative products, Master Thesis, King's College London.

[62] Georg, C. P., 2011, The effect of the interbank network structure on contagion and common shocks, Discussion Paper (Series 2: Banking and Financial Studies) 12/2011, Deutsche Bundesbank.

[63] Geske, R., 1977, The Valuation of Corporate Liabilities as Compound Options, *Journal of Financial and Quantitative Analysis*, 541-552.

[64] Giesecke K., 2003, A simple exponential model for dependent defaults, *Journal of Fixed Income*, 13, 74-83.

[65] Giesecke, K., 2004, Credit RiskModelling and Valuation: An Introduction, URL: http://ssrn.com/abstract=479323

[66] Giesecke, K., 2004, Correlated default with incomplete information, *Journal of Banking and Finance* 28, 1521-1545.

[67] Giesecke, K., 2006, Default and Information, *Journal of Economic Dynamics and Control*, 30.

[68] Giesecke, K., 2008, Portfolio credit risk: top down vs. bottom up approaches. In: Cont, R. (Ed.), *Frontiers in Quantitative Finance: Credit Risk and Volatility Modeling*. Wiley, New York.

[69] Giesecke, K., and Goldberg, L., 2004, Sequential defaults and incomplete information, *Journal of Risk* 7, 1-26.

[70] Giesecke, K., and Goldberg, L., 2004, Forecasting Default in the Face of Uncertainty, *Journal of Derivatives*, Vol. 12, No. 1, pp. 14-25, 2004.

[71] Giesecke, K., and Goldberg, L., 2011, A Top Down Approach to Multi-Name Credit. *Operations Research*, vol.59, pp 283 - 300.

[72] Giesecke, K., Kim, B., 2011, Systemic risk: What defaults are telling us, *Management Science* 57 (8), 1387-1405.

[73] Giesecke, K., Spiliopoulos, K., Sowers, R. B., 2011. Default clustering in large portfolios: typical events. URL: http://ssrn.com/abstract=1736843

[74] Giesecke, K., Weber, S., 2004, Cyclical correlation, credit contagion, and portfolio losses, *Journal of Banking and Finance* 28, 3009-3036.

[75] Giesecke, K., Weber, S., 2006, Credit contagion and aggregate losses, *Journal of Economic Dynamics and Control* 30, 741-761.

[76] Giesecke, K., Kim, B., Zhu, S., 2011, Monte Carlo algorithms for default timing problems, *Management Science*, vol. 17, pp 2115 - 2129

[77] Guo, X., Jarrow, R., and Zeng, Y., 2005, Information Reduction in Credit Risk Models, Cornell University (2005a, preprint).

[78] Hackbarth, D., Miao, J., and Morellec, E., 2006, Capital Structure, Credit Risk, and Macroeconomic Conditions, *Journal of Financial Economics*, Volume 82, pp 519–550.

[79] Herbertsson, A., Rootzen, H., 2008, Pricing kth-to-default swaps under default contagion: the matrix analytic approach, *Journal of Computational Finance* 12 (1), 49-78.

[80] Hsu, J., Saá-Requejo, J., and Santa-Clara, P., 2004, Bond Pricing with Default Risk, UCLA Working Paper.

[81] Huang, J. Z., 1997, Default Risk, Renegotiation, and the Valuation of Corporate Claims, Ph.D. Dissertation, NYU.

[82] Huang X, Oosterlee, C.W., Van der Weide, J.A.M., 2007, Higher Order Saddlepoint Approximations in the Vasicek Portfolio Credit Loss Model, *Journal of Computational Finance*, (93–113).

[83] Hughston, L. P., and Turnbull, S., 2001, Credit Risk: Constructing the Basic Building Blocks, *Economic Notes* 30, 281-292.

[84] Hull, J., 2009, *Options, Futures and Other Derivatives*, Prentice Hall Publications, seventh Edition.

[85] Hull, J., and White, A., 2001, Valuing Credit Default Swaps II: Modelling Default Correlations, *Journal of Derivatives* 8, 12-22.

[86] Iori, G., Jafarey, S., Padilla, F. G., 2006, Systemic risk on the interbank market, *Journal of Economic Behavior and Organization* 61, 525-542.

[87] Jarrow, R. A., and Protter, P., 2004, Structural versus reduced form models: a new information based perspective, *Journal of Investment Management* 2, 1-10.

[88] Jarrow, R. A., Yu, F., 2001, Counterparty risk and the pricing of defaultable securities, *Journal of Finance* 56 (5), 1765-1799.

[89] Jarrow, R. and Turnbull, S., 1992, Credit Risk: Drawing the Analogy, *Risk Magazine* 5(9).

[90] Jarrow, R. and Turnbull, S., 1995, Pricing Derivatives on Financial Securities Subject to Default Risk, *Journal of Finance*, 50, 53-86.

[91] Jarrow, R., Lando, D., and Turnbull, S., 1997, A Markov Model for the Term Structure of Credit Risk Spreads, *Review of Financial Studie*s 10, 481-523.

[92] Jarrow, R., Lando, D., and Yu, F., 2003, Default Risk and Diversification: Theory and Applications, Working Paper, Report, Cornell University.

[93] Jarrow, R., van Deventer, D., and Wang, X., 2003, A Robust Test of Merton's Structural Model for Credit Risk. *Journal of Risk* 6(1), 39–58.

[94] Jeanblanc, M., and Rutkowski, M., 2000, Modelling of Default Risk: An Overview, *Mathematical Finance: Theory and Practice*, Higher Education Press, Beijing, 171-269.

[95] Jones, P., Mason, S., and Rosenfeld, E., 1984, Contingent Claim Analysis of Corporate Capital Structures: An Empirical Investigation, *Journal of Finance* 39, 611-625.

[96] Jones, S., Hensher, D.A., 2008, *Advances in Credit Risk Modelling and Corporate Bankruptcy Prediction*, Cambridge University Press

[97] Jorion, P., 2007, *Value at Risk*, third edition, McGraw-Hill

[98] Jorion, P., Zhang, G., 2009, Credit contagion from counterparty risk, *Journal of Finance* 64 (5), 2053-2087.

[99] JP Morgan Global Research, 1997, CreditMetricsTM: The Benchmark for Understanding Credit Risk Technical Document (New York) (www.creditmetrics.com).

[100] Kaizoji, T., 2001, A model of international financial crises, *Physica A: Statistical Mechanics and its Applications*, Volume 299, pp 279–293.

[101] Kaizoji, T., 2001, Speculative bubbles and crashes in stock markets: an interacting-agent model of speculative activity, Volume pp 493–506.

[102] Kealhofer, S., 2003a, Quantifying Credit Risk I: Default Prediction, *Financial Analysts Journal*, January/February.

[103] Kealhofer, S., 2003b, Quantifying Credit Risk II: Debt Valuation, *Financial Analysts Journal*, March/April.

[104] Kijima, M., 2000, Valuation of a Credit Swap of the Basket Type, *Review of Derivatives Research* 4, 81-97.

[105] Kijima, M., and Muromachi, Y., 2000, Credit Events and the Valuation of Credit Derivatives of Basket Type, *Review of Derivatives Research* 4, 55-79.

[106] Kim, J., Ramaswamy, K., and Sunderasan, S., 1993, Does Default Risk in Coupons Affect the Valuation of Corporate Bonds? A Contingent Claims Model, *Financial Management*, 117-131.

[107] Koopman, S. J., Lucas, A., Schwaab, B., 2011, Modeling frailty-correlated defaults using many macroeconomic covariates, *Journal of Econometrics*, Volume 162, pp 312–325.

[108] Kraft, H., Steffensen, M., 2007, Bankruptcy, counterparty risk, and contagion, *Review of Finance* 11, 209-252.

[109] Lando, D., 1998, On Cox Processes and Credit Risky Securities, *Review of Derivatives Research*, 2, 99-120.

[110] Lando, D., Nielsen, M.S., 2010, Correlation in corporate defaults: Contagion or conditional independence?, *Journal of Financial Intermediation*, 19

[111] Lando, D., 2009, *Credit Risk Modeling: Theory and Applications*, Princeton University Press

[112] Laurent, J.P., Gregory, J., Basket default swaps, CDOs and factor copulas, *Journal of Risk.*

[113] Leland, H. E., 1994, Risky Debt, Bond Covenants and Optimal Capital Structure, *Journal of Finance* 49, 1213-1252.

[114] Leland, H. E., 2004, "Predictions of Default Probabilities in Structural Models of Debt", Journal of Investment Management 2.

[115] Leland, H., and K. Toft, 1996, Optimal Capital Structure, Endogeneous Bankruptcy, and the Term Structure of Credit Spreads, *Journal of Finance*, 51, 987-1019.

[116] Li, D. X., 1999, On Default Correlation: A Copula Function Approach, URL: http://ssrn.com/abstract=187289

[117] Liggett, T. M., 1985, *Interacting Particle Systems*, Springer-Verlag, New York.

[118] Longstaff, F. A., and Schwartz, E. S., 1995, A Simple Approach to Valuing Risky Fixed and Floating Rate Debt, *Journal of Finance* 50, 789-819.

[119] Longstaff, F., and E. Schwartz, 1995, Valuing Risky Debt: A New Approach, *Journal of Finance*, 50, 789-820.

[120] Lyden, S., and D. Saraniti, 2000, An Empirical Examination of the Classical Theory of Corporate Security Valuation, Research Paper, Barclays Global Investors.

[121] Mello, S. A., Parsons, J. E., 1992, Measuring the Agency Cost of Debt, The Journal of Finance, Volume 47, pp 1887–1904.

[122] Merton, R., 1974, On the Pricing of Corporate Debt: the Risk Structure of Interest Rates, *Journal of Finance* 29, 449-470.

[123] Moraux, F., 2004, Valuing Corporate Liabilities When the Default Threshold is not an Absorbing Barrier, University of Rennes Working Paper.

[124] Muller, J., 2006, Interbank credit lines as a channel of contagion, *Journal of Financial Services Research* 29 (1), 37-60.

[125] Nelson, B. L., 2002, *Stochastic Models: Analysis and Simulation*, Dover, New York.

[126] Nielsen, L. T., Saá-Requejo, J., and Santa-Clara, P., 1993, Default Risk and Interest Rate Risk: The Term Structure of Default Spreads, Working Paper, INSEAD.

[127] Nier, E., Yang, J., Yorulmazer, T., Alentorn, A., 2007, Network models and financial stability, *Journal of Economic Dynamics and Control* 31, 2033-2060.

[128] Ogden, J. P., 1987, Determinants of the ratings and yields on corporate bonds: Tests of the contingent claims model, *Journal of Financial Research*, Volume 10, pp 329–340.

[129] Pokutta, S., Schmaltz, C., April 2011. A network model of bank lending capacity. URL http://ssrn.com/abstract=1773964

[130] Rogers, L. C. G., 1999, Modelling Credit Risk. Working Paper, University of Bath.

[131] Ronn, E. I., and Verma, A. K., 1986, Pricing Risk-Adjusted Deposit Insurance: An Option Based Model, *Journal of Finance* 41, 871-895.

[132] Schmid, B. 2004, *Credit Risk Pricing Models: Theory and Practice*, second edition. Springer

[133] Schönbucher, P. J., 2000, Factor Models for Portfolio Credit Risk, Working Paper, Department of Statistics, Bonn University.

[134] Schönbucher, P. J., 2003, *Credit Derivatives Pricing Models*, Wiley Finance.

[135] Schönbucher, P. J., and Schubert, D., 2001, Copula-Dependent Default Risk in Intensity Models, Working Paper, Department of Statistics, Bonn University.

[136] Schuermann , T., 2004, What do We Know about Loss Given Default?, URL: http://ssrn.com/abstract=525702

[137] Staum, J., 2012. Counterparty Contagion in Context: Contributions to Systemic Risk URL http://ssrn.com/abstract=1963459

[138] Sun, Y., Mendoza-Arriaga, R., Linetsky, V., 2011, Marshall-Olkin multivariate exponential distributions, multidimensional subordinators, effcient simulation, and applications to credit risk, URL: http://www.researchgate.net/publication/228232458.

[139] Vasicek, O., 1984, Credit Valuation, White Paper, Moody's KMV.

[140] Vasicek, O., 1987, Probability of Loss on Loan Portfolio, in *Finance, Economics and Mathematics*, John Wiley & Sons, 2016

[141] Vasicek, O., 1991, Limiting Loan Loss Probability Distribution, in *Finance, Economics and Mathematics*, John Wiley & Sons, 2016.

[142] Yu, F., 2002a, Correlated Default in Reduced-Form Models, Working Paper, University of California, Irvine.

[143] Yu, F., 2002b, "Modelling Expected Return on Defaultable Bonds," Journal of Fixed Income 12, 69-81.

[144] Yu, F., 2005, Accounting Transparency and the Term Structure of Credit Spreads, *Journal of Financial Economics*, 2005

[145] Yu, F., 2007, Correlated defaults in intensity-based models, *Mathematical Finance* 17 (2), 155-173.

[146] Zhang, F. X., 2003, What Did the Credit Market Expect of Argentina Default? Evidence from Default Swap Data, Working Paper, Federal Reserve Board, Division of Research and Statistics, Washington.

[147] Zhou, C., 1997, A Jump-Diffusion Approach to Modelling Credit Risk and Valuing Defaultable Securities, Federal Reserve Board, Washington.

Chapter 3

Financing Production with Liquidity Constraints: the role of Trade Credit

Andrea TOTO, Mario EBOLI*

february 2016

Abstract

Empirical evidence shows that the implicit interest rate in a trade credit agreement is generally very high compared to the rates of bank credit. In spite of this apparent high cost, trade credit is widely used and represents an important proportion of firms' finance. This paper analyses the effects of trade credit on the investment decisions of a financially constrained firm in a manufacturing supply chain, with particular reference to a context of financial turmoil and credit rationing. Trade credit creates systemic risk but, at the same time, enhances the resilience of firms to liquidity shocks and, through the flexibility of repayment terms, affects positively the inventory investment decisions and, consequently, the future expected levels of output and revenues. These feature of trade credit are investigated with the aim of pinning down their effects on the financing of inventory investments of liquidity-constrained firms. To this end, we put forward a multi-factor model of trade credit connections in supply chains which is able to capture the insurance effect of trade credit, i.e. the insurance coverage against liquidity risk embedded in trade credit contracts, thanks to which a financially-constrained firm suffering a liquidity shortage can maintain a level of expected inventory investment (and, as a consequence, a future expected output level) as close as possible to the optimal desired level. We measure the insurance

*University "G. d'Annunzio", Department of Business Economics. We can be contacted via e-mail at: andrea.toto@unich.it, m.eboli@unich.it

effect of trade credit by characterising the impact of an exogenous liquidity shock on the investment decisions and future expected output level of a financially constrained firm in manufacturing supply chain, under the assumption that its supplier is willing to concede a repayment deferral of a portion of the trade debt . Finally we use the optimality conditions to derive a reduced form inventories equation in order to test empirically the implications of the model.

Keywords: Financing Constraints – Supply Chains – Trade Credit – Credit Rationing – Inventories.

1 Introduction

In general, it is well established that liquidity constraints limit the investment decisions of firms and, consequently, their output. In the manufacturing sectors, given the relative rigidity of the production in the short run, liquidity shocks are bound to affect not only the levels of investment and output, but also the profitability of firms (or the implicit wages paid in family firms). Financing the production of firms operating in manufacturing sectors is an issue that constantly attracts the attention of both researchers and policy makers, especially so during periods when the institutional framework of the relations between banks and firms is undergoing substantial changes (e.g. Basel Agreements II and III). The issue becomes particularly relevant in periods of economic downturn associated with a financial crisis (like the current one), when the traditional concerns for the difficulties faced by manufacturing firms in accessing bank credit are reinforced by the emergence of liquidity shortages and the detrimental effects that such constraints can have on the investment decisions and, consequently, on the production and earnings of manufacturing firms.

An analysis of the coverage of the financial needs arising from investment in working capital requires a careful evaluation of the role played by the sources of funding, such as self-financing and trade credit, that do not come from banks or other financial intermediaries. Regardless of the actual severity of liquidity constraints due to the rationing of bank credit, such sources of funds are relevant in as much as they are intertwined with the contractual terms (timing of payments, discounts, pricing, etc.) of the commercial links that firms have with their suppliers and buyers. Moreover, trade credit is particularly relevant for liquidity-constrained firms in that the default of a buyer does not imply the end of the commercial relation between the supplier and the defaulting client. It is well established by empirical evidence, and explained by economic theory, that it is more convenient for a supplier to concede a deferral of payment to a defaulting client, rather than to push for the liquidation of its assets, i.e. its bankruptcy. Thus, the use of trade credit as a source of funding improves the resilience of a liquidity-constrained firm to unexpected financial shortages, i.e. it helps an illiquid firm to stay solvent. In other words, trade credit contracts embed an insurance coverage against liquidity risk. The cost of this insurance is incorporated in the pricing policies set by suppliers (who often use trade credit terms to discriminate among their clients) and, as a consequence, affect the allocation of earnings among the firms that belong to a supply chain. On the other hand, the fact that – in a trade credit contract – a supplier shares part of the liquidity risk run by its client, implies that a liquidity shock suffered by a defaulting firm is transmitted to its supplier and from the latter to its own suppliers and so forth, generating a systemic liquidity risk affecting most of the firms in a supply chain. In brief, the use of trade credit in supply chains provides funding and the sharing of liquidity risk but, at the same time, it creates the grounds for financial contagion.

Trade credit arises when a supplier allows a customer to delay the payment of goods already delivered. It is generally associated with the purchase of intermediate goods. Appearing in firms' balance sheets as accounts receivable (on the seller's side) and accounts payable (on the buyer's side), trade credit is a type of credit sellers extend to buyers, allowing the latter to purchase goods from the former without immediate payment. Common wisdom in operations suggests that demand uncertainty and early commitment to order quantities leave the buyer (downstream firm) with significant inventory risks, hence forcing the buyer to order less than would be optimal for the supply chain. As a type of inter-firm contracts, trade credit has unique advantages in mitigating this problem: first, the amount and timing of trade credit are closely associated with purchase and inventory decisions, second, the repayment of trade credit is contingent on the demand realization and possible liquidity shocks, similar to many risk-sharing mechanisms in channel coordination.

Empirical evidence shows that the implicit interest rate in a trade credit agreement is generally very high compared to the rates of bank credit. The reason why trade credit appears to be an expensive form of finance lies within the structure of a standard trade credit contract. A typical deal normally involves three elements: a discount on the price agreed if the buyer pays early; the number of days that qualify for early payment; and the maximum number of days for payment. For example, a common contract called "2-10 net 30" means that if customers pay within ten days of delivery they qualify for a 2% discount. Otherwise they can pay up to 30 days after delivery. The discount for early payment implies an interest rate that the customer pays for the credit received. In the case of the "2-10 net 30" contract, the customer is effectively receiving credit at a 2% rate for 20 days. Thus the equivalent one year interest rate of this deal is about 44%. This is an extremely high rate compared with the market rate that a bank would charge for a similar type of loan. Other common deals also have very high interest rates. For example, one of the most common contracts in the US, according to Ng et al. (1999), is the "8-30 net 50", analogous to the previous one, but corresponding with an annual interest rate of 358%. In spite of this apparent high cost (i.e. the high interest rates implicit in delayed payment prices), trade credit is widely used and represents an important and considerable proportion of firms' finance, as will be extensively illustrated in the section 2. It is therefore surprising that banks do not take over this potentially profitable business, offering more credit lines to finance commercial transactions.

There are a number of empirical findings suggesting that firms suffering from credit rationing use trade credit. As noted by Mian and Smith (1992, 1994) credit extended by a seller who allows delayed payment for his products, or trade credit, represents a substantial fraction of corporate liabilities, especially for middle-market companies. The authors report that "for the 3,550 non financial Nasdaq firms covered by COMPUSTAT, accounts payable were 26% of corporate liabilities, at the end of 1992." Atanasova, Wilson (2003) find that in a Federal Reserve Board study, Elliehausen and Wolken (1993) note that accounts payable constituted 20% of all non-bank, non-farm small businesses' liabilities and 15% of all large firms' liabilities in 1987. In addition, more than 80% of all firms used trade credit. On the other hand, accounts receivable is one of the main assets on most corporate balance sheets, representing on average up to 30-35% of total assets. Petersen and Rajan (1994,1995) find that firms which are less likely to be bank credit constrained tend to rely less on trade credit. They show that firms without relationships with banks resort more to trade credit, and sellers with greater ability to generate cashflows provide more trade credit. Empirical results of Nilsen (1994) show that firms react to monetary contractions by using trade credit: when the interest rates increase (because the monetary policy is tightened), small firms react by using trade credit to avoid credit rationing. During monetary contractions, small firms, which Gertler and Gilchrist (1993a, 1993b) suggest are likely to be particularly bank credit rationed, react by borrowing more from their suppliers. Also, trade credit tends to be used less in economies where relationships between banks and firms are strong, such as Germany [see Breig (1994)], or where financial markets play an information transmission and monitoring role, such as the United States, and more in economies where financial markets are less developed and bank firm relationships are more distant, such as France.

These empirical findings raise a puzzle. i) Why is trade credit so expensive? ii) Why is trade financed by suppliers instead of banks? iii) Why is trade credit available when bank credit is rationed? and iv) Why do companies rely on their suppliers to obtain financing, rather than specialized financial intermediaries such as banks or financial markets? Suppliers are themselves more likely to be credit constrained and to have high cost of funds than banks. Hence when banks cannot lend, suppliers should not be able to lend either. In the section 2 we illustrate most of the answers that researchers gave to the trade credit puzzle.

This paper focuses on the role of trade credit in manufacturing supply chains, with the aim of putting forward an analytical model of trade credit connections in supply chains which is able to measure the degree of exposure of such investment decisions to unexpected liquidity constraints arising from liquidity risk and systemic risk, and to capture the insurance effect of trade credit, i.e. the insurance coverage against liquidity risk embedded in trade credit contracts, thanks to which the firm, under liquidity shortage, can maintain a level of inventory investments (and, as a consequence, a future expected output level) as close as possible to the optimal desired level. By enhancing the resilience of firms to liquidity shocks, the insurance effect incorporated in the trade credit, under an aggregated supply chain perspective, is able to improve the global resilience of the chain, despite the fact that it create systemic risk. More precisely, in this paper we try to demonstrate analytically that the insurance motive, through an increasing effect on the marginal cost of inventory investment, gives the financially constrained buyer (the downstream firm) the incentive to finance inventories with trade credit. Furthermore, also when less financially constrained, in term of access to bank credit, it can be optimal to borrow from suppliers, since the "defaultability" of trade credit (i.e. the possibility to defer the repayment of the trade debt, which we assume to be granted by supplier) gives the firm the possibility to absorb expected liquidity shocks through a sort of risk-sharing mechanism. In this way the firm improves its resilience to liquidity shocks and, under liquidity shortage, it is less forced to reduce drastically investments in inventories and, consequently, the future expected output level for the next period. To this end, we start from the model described in Caggese (2007), inside which we consider the trade credit as substitute source of funding respect on bank credit. Using the model, we measure the effects of trade credit on investment decisions and expected output of a financially constrained firm suffering a liquidity shortage condition due to an exogenous liquidity shock.

The paper is organized as follows. In the next section, we review the literature related to this work. In doing this, we describe the most important empirical evidences in the literature related to the trade credit. In section three we model the effect of trade credit in term of inventory investment decisions of a financially constrained firm in a manufacturing supply chain. In doing this we show that, under liquidity shock conditions and with expectations of future financial constraints, trade credit gives the firm an insurance against liquidity shortage, allowing the firm to maintain the optimal desired level of inventory investment. In section four we draw three possible empirical frameworks in order to test the implications of the model over a sample of italian manufacturing firms. The proofs of the propositions used in the paper are collected in the Appendix.

2 Trade Credit in Economic Theory

The widespread use of trade credit, despite its high cost, has attracted the attention of economic theorists who, over the last fifteen years, have provided convincing and exhaustive answers to such a phenomenon.

The existing data show that trade credit constitutes a relevant share of the balance sheet of companies in developed countries and an even larger share in developing countries. Trade credit is the largest source of short-term financing for American corporations. Most firms extensively use trade credit despite its apparent greater cost, and trade credit interest rates commonly exceed 18 percent. Rajan and Zingales (1995) show that trade credit, as a percentage of total credit granted to companies, amounts to 11,5% in Germany, to 17% in France, 15% in the United States, 13,3% in Canada, 13,7% in Great Britain, 14,7% in Italy, and to 15,4% in Japan. These authors also argue that the trade credit granted by companies, as a percentage of their total assets, goes from 13% in Canada to 29% in France and Italy. Cuñat (2007) shows that in Great Britain, trade credit amounts to 17% of total assets, 43% of debts and 52% of short term debts of companies, while in the US these percentages are 18%, 34% and 58%, respectively. This author sustains that these percentages grow during periods when buyers suffer temporary liquidity shortages, and that such an increase in trade credit occurs through defaults on existing debts, where suppliers allow lenders to postpone payments. Suppliers seem to lend to their customers experiencing financial trouble, even when banks are not willing to lend. This additional lending may occur through financing a higher proportion of purchased goods or by extending the agreed maturity of the loans. Furthermore, in many circumstances, this extra lending occurs via late payment of already extended debts. Cannari et al. (2004) study trade credit terms in Italy on the basis of two surveys carried out by the Bank of Italy. The authors show that, on average, 80-90 per cent of sales of the Italian companies surveyed are paid on a deferred payment basis of 90 days on average and a delay of 11 days. 83 per cent of trade credit is extended on net terms (with no discount offered for prompt payment). When two-part terms are offered, the cost of trade credit is normally very high and well above market interest rates. These results point to the importance of marketing determinants for the extension of trade credit (product quality guarantee, customer relationships) more than financial motives. Most of suppliers (80 per cent) differentiate customers by means of price and payment terms in favor of older and bigger firms and long-standing customers; they apply stricter terms to late payers, while being indulgent with those in temporary distress. Two-part terms are offered to not well-known customers and are aimed at extracting information on their creditworthiness. The analysis shows that two-part and net terms respond to very different aims. Contrary to the presumption of most of the literature, generalizing the characteristics of two-part terms contract as pertaining to the whole of credit transactions is unwarranted. Price and payment terms are a flexible and many-faceted device extensively used by firms for building customer relationships, acquiring information on buyers creditworthiness, and exploiting market power.

An important aspect of trade credit, furthermore, is the two-way nature of the transaction. Many companies, particularly those at intermediate points in the value chain, both use trade credit as customers and provide it as suppliers. Trade credit thus represents a substantial component of both corporate liabilities and assets. Prior empirical evidence suggests that larger firms are, in aggregate, net trade credit providers and may therefore provide an important mechanism for channelling finance down to those firms rationed by financial intermediaries. These studies find that small firms, which are more likely to be credit rationed, rely heavily on trade credit when credit market conditions deteriorate. Empirical evidence made economists wonder why such a large share of companies' credit is not provided by banks and financial intermediaries, which are specialized in credit services. The question becomes even more puzzling because of the cost of trade credit, which is much higher than the cost of bank credit. Considering that trade credit is so expensive, why do companies with no liquidity shortage resort to trade credit? Why do firms with binding liquidity constraints grant trade credit to their clients?

Economic theory has responded to the questions at the heart of the trade credit puzzle with a number of arguments. Wilner (2000) argues that trade creditors, desiring to maintain an enduring product market relationship, grant more concessions to a customer in financial distress than would be granted by lenders in a competitive credit market. The debtor firm anticipates larger renegotiation concessions from the supplier part, and so agrees to pay a higher interest rate to a trade creditor than to a credit market lender. In this perspective, the higher trade credit interest rate is fair compensation for the firm's receiving larger concessions if financial distress occurs. The author argues that the borrower in financial distress can take advantage of this trade credit connection if it generates a large percentage of the creditor's profits: if so, the borrower should be willing to accept what appears to be a less favorable contract, anticipating more power in renegotiations should financial distress occur.

In Biais,Gollier (1997) asymmetric information between banks and firms can preclude financing of valuable projects. The authors starts from the assumption that suppliers have private information about their customers, and shows that trade credit can facilitate aggregation of the supplier's information with the bank's information and thus alleviate an information asymmetry which otherwise would preclude financing of positive net present value (NPV) projects. Lending through trade credit serves a non-financial role and suppliers have some comparative advantage in lending, since it is reasonable to assume that suppliers place more value on the collateral of their customer than the bank would. In other terms, trade credit, by incorporating in the lending relation the private information held by suppliers about their customers, can alleviate bank credit rationing due to asymmetric information between banks and firms and this theoretical result appears consistent with empirical studies documenting that firms which are likely to be credit constrained (such as small firms during monetary contractions, or firms without relationships with banks) resort more to costly trade credit. Furthermore, such a theoretical analysis of the link between information-motivated trade credit and information-motivated bank credit rationing provides an interpretation for a number of stylized facts. One of these empirical results, for example, is described in Petersen and Rajan (1995): firms which do not suffer from credit rationing do not use trade credit, while firms for which asymmetric information generates credit rationing react by using trade credit. Furthermore, when the buyer does not have cash and is bank credit constrained, in absence of trade credit it could not buy the good from the seller. Yet by extending trade credit the seller enables the buyer to raise the funds necessary to purchase the good. Another crucial argument in Biais,Gollier (1997) is that trade credit is a credible way for the seller to transmit its private information about the buyer to the bank. If the seller is willing to extend trade credit, and thus to bear the default risk of the buyer, it must be that it has good information about the latter. On observing this, the bank updates positively its beliefs about the buyer, and therefore agrees to lend. In this perspective, trade credit enables the private information of the seller to be used in the lending relationship, and this additional information can alleviate credit rationing due to adverse selection. As a consequence, bank credit and trade credit are complementary because the granting of trade credit by the suppliers of a firm reveals favorable information about the firm, and this constitutes a positive signal for banks and other potential lenders. So, in their model some buyers borrow both from the bank and their supplier, in spite the fact that the rate at which they borrow from the latter is larger than the bank rate. Firms would use simultaneously two sources of financing, one being more expensive than the other, since if they did not use trade credit, then the information from the seller could not be conveyed, and the relatively low bank rate would not be granted. Thus this is an interpretation for the fact that trade credit is used even though it is costly.

Ng, Smith, Smith (1999) also analyze trade credit under the informational perspective. Facing imperfect information about buyer default risk, sellers either demand prepayment or cash or find ways to evaluate creditworthiness. Trade credit terms can be designed to elicit information on creditworthiness because buyers' responses to the terms help identify those who may be having financial difficulties. Their empirical results give support to theories that explain credit terms as contractual solutions to information problems concerning product quality and buyer creditworthiness.

Another argument explaining the trade credit puzzle is connected to a moral hazard problem. Burkart and Ellingsen (2004) present a contract theoretic model of trade credit relations in competitive markets with asymmetric information and argue that trade credit reduces the scope for moral hazard on the part of debtors. For this reason, according to the authors, the weaker the legal protection of creditors, the larger the use of trade credit.

They also argue that trade credit and bank credit are complementary for firms subject to liquidity constraints, while they are substitutes for firms with sufficient 'debt capacity', i.e. firms that have access to external sources of funding. Furthermore, trade credit is more prevalent in less developed credit markets, and accounts payable of large unrated firms are more countercyclical than those of small firms.

Nilsen (2002) presents data in favour of the thesis that firms that face credit rationing use trade credit as a substitute for other sources of funding, and this occurs more markedly during periods of restrictive monetary policies. In other terms, during monetary contractions banks restrict some firms' loans, thus reducing their desired investment independently of interest rates. The author finds that small firms increase trade credit as a substitute credit: more precisely, trade credit is widely used by the small firms suffering the loan decline, indicating a strong loan demand. This analysis supports the bank lending channel: constraint firms do not voluntarily cut bank loans since they increase a less-desirable alternative. But the author also finds that large firms increase trade credit as well as small firms and this appears as a puzzle since they are typically assumed to have wide access to other credit sources.

A new stream of literature focuses on the role of trade credit in default and liquidation. Frank and Maksimovic (2005) and Longhofer and Santos (2003), for example, argue that suppliers have an advantage in liquidating goods in case of their buyers' default, whereas Wilner (2000), as well as Cunat (2007), suggest that suppliers are more willing to offer help when their customers are in trouble (e.g., due to liquidity shocks), given their desire to continue business with those customers in the future. In other terms, trade credit provides coverage of liquidity risk: suppliers, in the face of default of a client, are better off allowing postponement of payments rather than resorting to suing the debtor and possibly contributing to his bankruptcy. The trade credit contract embeds an insurance against liquidity shocks, and this can also explain its costs (Cuñat, 2007). Cunat (2007) analyses the interaction between the financial and the industrial aspects of the supplier-customer relationship. The author examines how, in a context of limited enforceability of contracts, suppliers may have a comparative advantage over banks in lending to their customers. Suppliers are able to enforce debt repayment better than banks, as they hold the threat of stopping the supply of intermediate goods to their customers. Under another perspective, suppliers may act as liquidity providers or lenders of last resort, supporting their customers whenever they experience temporary liquidity shocks and so providing insurance against liquidity shortages that may endanger the survival of their customers. Under this conditions, a financial relationship between a supplier and a customer emerges as a natural consequence of their commercial interaction, despite the existence of a competitive banking sector. As necessary condition for these relationship to exist, the author assume the existence of a surplus that will be split between suppliers and customers if they stay together. In other words, there must be a link between the supplier and the customer that makes it costly for the customer to find alternative suppliers and makes it costly for the supplier to loose its current customers. As a result, the relatively high implicit interest rates of trade credit is justified by the existence of a default premium and an insurance premium. The default premium accounts for the fact that suppliers lend when banks are not willing to lend. and they use their extra enforceability power to lend on the basis of returns that are stochastic. This makes trade credit more risky than bank debt. The insurance premium is related to the fact that suppliers foresee the future needs of liquidity of their customers. As they know that they may have to bail out customers in need of extra liquidity, they will charge them a premium for providing insurance against potential liquidity shocks.

In Smith (1987) the high interest rate incorporated in trade credit contracts is due to a default-risk screening mechanism. Trade Credit terms implicitly define a high interest rate that operates as an efficient screening device where information about buyer default risk is asymmetrically held. By offering trade credit, a seller can identify prospective defaults more quickly than if financial institutions were the sole providers of short-term financing. The information is valuable in cases since it enables the seller to take actions to protect potentially non-salvageable investments in buyers. The author models two-part trade credit as such a screening contract, according to which a buyer who forgoes the discount may pay a very high effective interest rate, as determined by the terms of the two-part offer. Failure to take the discount in such cases signals that monitoring of the buyer's financial position is merited. Depending on the seller's response to default, simple net terms can generate a similar signal, although the penalty for paying after the net date is not explicit.

In Kashyap, Lamont, Stein (1994) firms that are "bank-dependent" are much more prone to shed inventories than their non-bank-dependent counterparts when interest rates are relatively high (i.e. tight monetary policy). Such differences are largely absent when interest rates are low (i.e. during a period of "loose" monetary policy), since much more firms have access to bank credit. The lending channel is likely to be particularly important in explaining inventory fluctuations during downturns.

According to Petersen, Rajan (1997), small firms whose access to capital markets may be limited have incentive to use more trade credit when credit from financial institutions is unavailable. Suppliers lend to constrained firms because they have a comparative advantage in getting information about buyers, they can liquidate assets more efficiently, and they have an implicit equity stake in the firms. Firms with better access to credit offer more trade credit. From the empirical point of view, they analyze the relevance of trade credit for companies of different size and age, using data from Compustat and from the Survey of Small Business Finances of the Federal Reserve. The authors find that i) the amount of trade credit granted by companies is directly correlated with the size and with the age of companies; ii) large and mature companies are often net suppliers of trade credit; iii) companies prefer to resort to bank credit, when available; iv) companies endowed with liquid reserves and with long term relations with banks use less trade credit.

In Carpenter, Fazzari, Petersen (1998) financing constraints explain the dynamic cycles in inventory investment. Cash flow, which is affected by trade credit dynamics, is much more successful than cash stocks or coverage in explaining the facts about inventory investment across firm size, different inventory cycles, and different manufacturing sectors.

Emery (1984) provides a pure financial explanation for the existence of trade credit and for the values of the credit terms offered to customers, consisting of two motives for extending trade credit. First, the pure operating flexibility motive arises because the opportunity to change credit policy provides the seller an efficient way to respond to fluctuations in demand. Second, the pure financial intermediary motive arise because trade credit lending enables the seller and/or the buyers to recapture at least part of the surplus loss due to a wedge between the market prices paid and received for the product plus a loan. This wedge is imposed by the fact that market borrowing rate of interest exceeds the market lending rate of interest and their differential acts as a hindrance to trade or, equivalently, as a financial market tariff caused by transactions costs). The trade credit lender's familiarity with its customers and product provide it with information and collection cost advantages over financial intermediaries

In Yang, Birge (2014) a cursory empirical investigation[1] clearly suggest that the variation in payable amounts is closely related to that in inventory and this gives support the reasonable conjecture that trade credit is extensively used, at least partially, to reduce the

[1] OLS regression which investigates the dynamics of Payable Days versus Inventory Days

widespread inefficiency in supply chains. They argue that, with demand uncertainty, trade credit enhances supply chain efficiency by serving as a risk-sharing mechanism. Two forces determine the optimal trade credit terms: the sales motive (increasing sales through risk-sharing) and the financing motive (minimizing costs of financial distress through financial diversification, that is, employing multiple financing sources). In other terms, trade credit influences supply chain efficiency through two separate effects. First, the sales motive (similar to many other channel coordination mechanisms) thanks to which trade credit allows the retailer to share inventory risks with the supplier, hence inducing a higher order quantity from the. Second, the financing motive, through which trade credit transfers some of the distress costs from the retailer to the supplier, allowing the former to further increase the order quantity. Facing a trade credit contract, the retailer finances inventory using a portfolio of cash, trade credit, and short-term debt, and the structure of this inventory financing portfolio depends on the retailer's financing need and bargaining power and on the retailer's leverage.

In, brief, we can summarize the arguments with which economic theory has responded to the questions at the hearth of the trade credit puzzle by classifying them in four categories categories:

1. Information advantage (monitoring costs): sellers are better informed than banks about their own clients; the receipt of trade credit is a signal of creditworthiness for banks. [Biais and Gollier (1997)];

2. Liquidation value: the collateral assets have a larger liquidation value for the suppliers than for the banks;

3. Moral hazard: i) delayed payments eliminate the risk that suppliers sell goods of a quality inferior to that contracted with the buyer; ii) Diversion theory: trade credit is in kind, suppliers lend goods while banks lend money. Trade credit makes it more difficult for the managers of firms to divert resources from purposes which are consistent with the interests of their creditors [(Burkart and Ellingsen (2004)];

4. Coverage of liquidity risk: suppliers, in the face of default of a client, are better off allowing postponement of payments rather than resorting to suing the debtor and possibly contributing to his bankruptcy. The trade credit contract embeds an insurance against liquidity shocks, and this can also explain its costs (Cuñat, 2007).

The rest of the paper is organized as follows: in the next section, we illustrate the arguments we lay out the basic setting that models the investment decisions of a financially constrained firm with trade credit connections on the supply side and access to two sources of funding, (i.e. bank credit and trade credit). The model, inspired by Caggese (2007), is developed over three periods and characterizes the impact of an exogenous liquidity shock which is sufficiently large to force the firm to postpone the repayment of the trade debt with the aim to maintain a sufficiently high level of next-period inventory investment. In the section 3 we describe the empirical framework aimed to test empirically the validity of the implications of the model.

3 The Model

The aim of this section is to develop a structural model of investment with financing constraints which is able to analyse the impact of a liquidity shortage on the firms' investment behaviour. Our starting point is the model described in Caggese (2007) in which we consider the trade credit as alternative source of funding. An unexpected liquidity shortage can have binding effects on the funding of the working capital of the firm and, consequently, on its production levels. Such effects are transmitted from firms facing a liquidity crisis to their suppliers to an extent that depends on the size of their trade credit obligations. Furthermore, under bank credit rationing, a sufficiently large liquidity shortage is likely to force the firm to reduce its desired level of investments in intermediate goods. This have cascade effects on the expected output level of the subsequent period. According with a number of stylized facts, when the buyer does not have cash and is bank credit constrained, in absence of trade credit it could not buy the good from the seller. By extending trade credit, the supplier enables the buyer to raise the funds necessary to purchase the goods.

In what follows, we analyse those sorts of effects of liquidity shortages which are caused by adverse events that generate unanticipated costs. The model described below lends itself, with simple adaptations, to characterize also the effects of liquidity shocks generated by other causes, such as a credit crunch, fluctuations of the product price and/or of the exchange rate (for exporting firms).

We consider a risk-neutral firm whose objective is to maximize the discounted sum of future expected profits. The discount factor is equal to $1/R$, where $R = 1 + r$ and r is the lending/borrowing risk-free interest rate. The firm finances its production with two sources of funding: the bank credit, b and the trade credit TC. Trade credit arises cause the firm's supplier allows its customer to delay the payment of goods already delivered. We associate the trade credit connection between the firm and its supplier with the purchase of intermediate goods. Financing imperfections are introduced as follows: both the sources of funding are limited by an upper limit and the firm cannot obtain new capital from the shareholders.

The model is developed over three periods: we proceed to characterize the effects of a liquidity shock on the firm activity by considering a succession of production cycles, i.e. a succession of time periods t, $t+1$, $t+2$, and perturbation of the steady state of the supply chain by an exogenous shock which happens in the second period and causes effects in term of investment decisions and output levels over the second and the third period.

The firm operates with three inputs, K_t, L_t and I_t, which denote fixed capital, labour and inventories, respectively. We consider as fixed capital land, buildings, plant, and equipment; as inventories the stock of intermediate goods that is used in the production process: raw materials, work-in-process goods and completely finished goods that are considered to be the portion of a business's assets that are ready or will be ready for sale.

Banks behave competitively and are willing to lend and borrow at an interest rate equal to r. Firms' supplier provides the firm with the intermediate goods necessary for production and is willing to concede a deferral of payment at an interest rate $r_1 > r$ over one period, and at an interest rate $r_2 > r_1 > r$ over two period, in case the firm is suffering a liquidity shortage and is unable to repay the trade credit debt within the initial contracted terms.

Time is divided in periods; the length of a period corresponds to a production cycle. The timing of the model is as follows:

At the beginning of period t the firm has a stock of permanent and fixed-term workers equal to L_t. The firm observes θ_{t-1}, realizes revenues Y_t , repays the debt $b_{t-1} \cdot R$ and the trade credit obtained in the previous period, $TC_{t-1} \cdot R_1$. Furthermore, the firm obtains new bank credit b_t from the bank and new trade credit TC_t from its supplier in order to finance production factors which will be used to produce the output Y_t. Output Y_t is obtained with factors K_t, I_t, L_t (decided at t) and observed at time $t+1$; the entire output, produced with factors purchased at t, is sold to its client at $t+1$. During the second period (at time $t+1$), the firm observes a *Liquidity Shock* which perturbs the steady state and inflicts a loss on it. We assume that the loss is sufficiently large to force the firm to reduce the investment in inventories if it has no possibility to posticipate the repayment of the trade credit obtained in the previous period. In other terms, despite the fact that the loss is partially absorbed by the current profits π_{t+1}, its extent is such that, after the payment of wages due to labour contracts and the repayment of bank credit obtained in the previous period, the firm is not able to repay the trade credit obtained at time t, TC_t, without reducing the optimal and desired level of investment in inventories I_{t+1}.

On the basis of the theoretical and empirical results cited above, we assume that a supplier never asks for the liquidation of a defaulting buyer and always accepts to defer the payment to the next period. So, the Firm can posticipate to the subsequent period the repayment of a fraction $(1-D)$ of the trade credit TC_t at an higher interest rate $r_2 > r_1 > r$. The firm, suffering a liquidity shortage and subject to bank credit constraints, can choose to repay the portion $D \cdot TC_t$, with $0 < D < 1$, at an interest rate r_1,within the initial contractual terms (i.e. at time $t+1$), deferring the repayment of the portion $(1-D) \cdot TC_t$ to the subsequent period $t+2$, at an interest rate $r_2 > r_1$

The production function is strictly concave in all the three factors. We assume a Cobb-Douglas functional form:

$$Y_t = \theta_t K_t^{\alpha} I_t^{\beta} L_t^{\gamma} \text{ with } \alpha + \beta + \gamma = 1 \tag{1}$$

All prices are constant and normalized to one. This simplifying assumption will be relaxed in the empirical section of the paper. The factor θ_t is a productivity shock that follows a stationary AR(1) stochastic process. For simplicity we assume that inventories stock is nondurable and fully depreciates after one period, while fixed capital is durable and depreciates at the rate δ_k,

$$0 < \delta_k < 1 \tag{2}$$

Gross fixed capital investment, ΔK_t, is irreversible, that is

$$\Delta K_t \geq 0 \tag{3}$$

and is given by

$$\Delta K_t = K_t - (1 - \delta_k) K_{t-1} \tag{4}$$

Financial imperfections are introduced by assuming that new share issues and risky debt are not available. At time t the firm can borrow one-period debt from or lend one-period debt to the banks at the market riskless rate r, where the face value of debt is denoted by b_t, and $b_t \cdot R$ is the amount the firm must repay at time $t+1$. A positive (negative)

b_t indicates that the firm is a net borrower (lender). Banks only lend secured debt, and the only collateral they accept is physical capital. Therefore, at time t the bank-borrowing capacity of the firm is limited by the following constraints:

$$b_t \leq vK_t \tag{5}$$

$$\pi_t \geq 0 \tag{6}$$

and

$$0 < v \leq 1 - \delta_k \tag{7}$$

where π_t are the profits with which the firm remunerate the shareholders, and v is the share of fixed capital that can be used as collateral.

Analogously, at time t the firm can obtain one-period trade debt from its supplier which is limited by the following constraint:

$$TC_t \leq \overline{TC} \tag{8}$$

TC_t denotes the face value of trade debt; $\overline{TC}$ denotes the upper limit of the amount of the credit the supplier is willing to grant. If $\overline{TC} = I_t$, the entire amount of inventories can be financed with trade credit and the constraint is not binding.

It is useful to define the financial wealth of the firm W_t^F, after the debts b_{t-1} and TC_{t-1} are repaid, as

$$W_t^F = Y_{t-1} - b_{t-1} \cdot R - TC_{t-1} \cdot R_1 \tag{9}$$

where $R = (1 + r)$ and $R_1 = (1 + r_1)$ and $TC_{t-1} \cdot R_1$ is the amount that firm repays to its supplier within the terms established in the trade credit contract.

After producing, the firm allocates W_t^F plus the new borrowed funds between profits, fixed capital investment, wages and inventory investment according to the following flow of funds equation (i.e. the budget constraint) for the period t:

$$\pi_t + I_t + \omega L_t + \Delta K_t = W_t^F + b_t + TC_t \tag{10}$$

or, analogously:

$$\pi_t + I_t + \omega L_t + \Delta K_t = Y_{t-1} - b_{t-1} \cdot R + b_t - TC_{t-1} \cdot R_1 + TC_t$$

In the second period, $t+1$, we introduce an expected *Liquidity Shock,* which is affected by the magnitude of θ_t, and which generate a loss for the firm, with a positive probability

p^2. As well as in the first period, the firm uses financial wealth plus new borrowing b_{t+1} to pay dividends and wages, and obtains new trade credit TC_{t+1} to finance investment in inventories I_{t+1}. The budget constraint for the second period is the following:

$$\pi_{t+1} + I_{t+1} + \omega L_{t+1} + \Delta K_{t+1} = W_{t+1}^F + b_{t+1} + TC_{t+1} - Liq.Shock \tag{11}$$

or, analogously.

$$\pi_{t+1} + I_{t+1} + \omega L_{t+1} + \Delta K_{t+1} = Y_t - b_t \cdot R + b_{t+1} - D \cdot TC_t \cdot R_1 + TC_{t+1} - Liq.Shock$$

where D is the fraction of TC_t which the firm is able to repay at $t+1$, at an interest rate r_1 and $(1-D)$ is the defaultable fraction of TC_t, i.e. the portion of trade debt obtained in the previous period which the firm can choose to repay at $t+2$, at an interest rate $r_2 > r_1$ in order to absorb the liquidity shortage condition. In accordance with empirical and theoretical results cited above, we assume that the supplier never asks for the liquidation of a defaulting buyer and always accepts to defer the payment to the next period.

It is useful to define the "defaultable" trade credit as

$$(1-D)TC_t = TC_t^D \tag{12}$$

Condition 1 *The liquidity shock is sufficiently large to force the firm to reduce the desired investment in inventories in absence of defaultable trade credit*

$$I_{t+1}^* + \omega L_{t+1} + \Delta K_{t+1}^* < Y_t - b_t \cdot R + b_{t+1} - D \cdot TC_t \cdot R_1 + TC_{t+1} - Liq.Shock$$

where $D = 1$. If $D = 1$, it means that the supplier is not willing to concede a deferral of payment, so $TC_t^D = 0$. $\Delta K_{t+1} = 0$ and $K_t = (1-\delta_k)K_{t-1}$, since the irreversibility constraint. Profits π_{t+1} cover partially but not totally the liquidity shortage due to the shock. In this situation the firm would be forced to reduce the desired optimal level of investment in inventories, and $I_{t+1} < I_{t+1}^*$. This produce a detrimental and decreasing effects on the level of expected output Y_{t+1} (to be obtained at the beginning of third period) and therefore on the level of expected revenues.

[2] In this version of the paper, without loss of generality, we consider the case in which the probability of the occurrence of the *Liquidity Shock* equals to 1. In the next version, we will specify in the model the case $0 < p < 1$, and we will specify the relation between the extent of the *Liquidity Shock* and the productivity term θ_t.

if $0 < D < 1$, it means that the supplier is willing to concede a deferral of payment, so $TC_t^D > 0$. In this case the firm decides to repay the fraction $(1-D)TC_t = TC_t^D$ in the third period: in this way it has the possibility to maintain the level of investment in inventories as much as possible close to the optimal desired level I_{t+1}^*. For a sufficiently low value of the Liq.Shock, it could exist a value of D for which the above flow of funds equation holds with equality. The advantage of the defaultable trade credit for the firm consists on the minimum possible investment reduction in response to the liquidity shock.

As a consequence of the Condition 1, the budget constraint for the third period is the following:

$$\pi_{t+2} + I_{t+2} + \omega L_{t+2} + \Delta K_{t+2} = W_{t+2}^F + b_{t+2} + TC_{t+2} \tag{13}$$

or, analogously.

$$\pi_{t+2} + I_{t+2} + \omega L_{t+2} + \Delta K_{t+2} = Y_{t+1} - b_{t+1} \cdot R + b_{t+2} - TC_t^D \cdot R_2^2 - TC_{t+1} \cdot R_1 + TC_{t+2}$$

where $TC_t^D = (1 - D)TC_t$

In our model the amount of defaultable trade credit influences the spending power of the firm in the second period and, as a consequence, its investment decisions. Under bank credit constraint and liquidity shortage condition, the possibility to defer the repayment of the trade debt increases the available cash flow to finance the inventory investments. As a consequence, the extent of the defaultable trade credit produces effects on the level of the expected output to be observed in the third period. Formally, the output Y_{t+1} is a function of TC_t^D:

$$Y_{t+1} = V[I_{t+1}(\underset{-}{D})] = V[I_{t+1}(\underset{+}{TC_t^D})] \tag{14}$$

Under liquidity shortage, the effects of D (and hence of TC_t^D) on I_{t+1} (and hence on Y_{t+1}) are as follows:

$$\frac{\partial Y_{t+1}}{\partial I_{t+1}} > 0; \frac{\partial I_{t+1}}{\partial D} < 0 \Longrightarrow \frac{\partial Y_{t+1}}{\partial D} < 0 \text{ and} \tag{15}$$

$$\frac{\partial TC_t^D}{\partial D} < 0; \frac{\partial I_{t+1}}{\partial TC_t^D} > 0 \Longrightarrow \frac{\partial Y_{t+1}}{\partial TC_t^D} > 0 \tag{16}$$

The term $\frac{\partial Y_{t+1}}{\partial TC_t^D}$ is the marginal productivity of the defaultable trade credit concerning the output Y_{t+1}: in other terms, it is the incremental quantity of output Y_{t+1} which the firm is able to obtain by purchasing an incremental unit of I_{t+1} thanks to the possibility to defer the repayment of a fraction of TC_t. In this framework, the trade credit obtained from the supplier (upstream firm) in the first period, TC_t, produce effects not only on the output obtained at the beginning of the second period, Y_t (by affecting the amount of I_t), but also on the expected output to be observed at the beginning of the third period, Y_{t+1}, through the effects of the defaultable portion $TC_t^D = (1-D)TC_t$, (which affects the amount of I_{t+1})

Let us denote the firm's value at time t, after θ_{t-1} is realized, by V_t. Formally, the firm's maximization problem is as follows:

$$V_t = \underset{b_t, I_t, TC_t, K_t}{MAX} \left\{ \pi_t + \frac{1}{R} E_t \left[\pi_{t+1} \right] + \frac{1}{R^2} E_t \left[\pi_{t+2} \right] \right\} \tag{17}$$

The firms maximizes (17) subject to Eqs. (3), (5), (6), (8), which are respectively the irreversibility constraint, the bank-borrowing constraint, the non-negativity constraint of profits and the trade credit constraint, and subject to the three flow-of-funds equations (10), (11), and (13).

In order to describe the optimality conditions of the model, we use Eqs. (10), (11), and (13) to substitute π_t, π_{t+1}, π_{t+2} in the value function (17). Let μ_t, λ_t, ϕ_t and σ_t be the Lagrangian multipliers associated, respectively, with the irreversibility constraint (3), the bank-borrowing constraint (5), the non-negativity constraint of profits (6) and the trade credit constraint (8). The solution of the problem is defined by the following conditions:

$$1 + \phi_t = \lambda_t + Et \left[1 + \phi_{t+1} \right] \tag{18}$$

$$1 + \phi_t = \sigma + \frac{R_1}{R} Et \left[1 + \phi_{t+1} \right] - \frac{1}{R^2} Et \left[\frac{\partial Y_{t+1}}{\partial TC_t^D} \right] \left(1 + Et \left[\phi_{t+2} \right] \right) - \frac{1}{R^2} COV \left(\phi_{t+2}; \frac{\partial Y_{t+1}}{\partial TC_t^D} \right) \tag{19}$$

$$Et \left[\frac{\partial Y_t}{\partial I_t} \right]_{TC} = R \cdot \Psi_t^D + R \cdot \Psi_t^I + \frac{1}{R} \cdot InsuranceEffectTerm \tag{20}$$

where

$$\Psi_t^D = \left\{ \left(\frac{R_2}{R} \right)^2 \cdot \frac{1 + Et \left[\phi_{t+2} \right]}{1 + Et \left[\phi_{t+1} \right]} + D \left[\frac{R_1}{R} - \left(\frac{R_2}{R} \right)^2 \cdot \frac{1 + Et \left[\phi_{t+2} \right]}{1 + Et \left[\phi_{t+1} \right]} \right] \right\} \tag{21}$$

$$\Psi_t^I = \left\{ \frac{\sigma_t - \frac{1}{R} COV \left(\phi_{t+1}; \frac{\partial Y_t}{\partial I_t} \right) - \frac{1}{R^2} COV \left(\phi_{t+2}; \frac{\partial Y_{t+1}}{\partial TC_t^D} \right)}{1 + Et \left[\phi_{t+1} \right]} \right\} \tag{22}$$

$$InsuranceEffectTerm = -Et\left[\frac{\partial Y_{t+1}}{\partial TC_t^D}\right]\frac{1+Et\left[\phi_{t+2}\right]}{1+Et\left[\phi_{t+1}\right]} \tag{23}$$

and

$$Et\left[\frac{\partial Y_t}{\partial I_t}\right]_b = R\cdot\left(1+\Phi_t^I\right) \tag{24}$$

where:

$$\Phi_t^I = \left\{\frac{\lambda_t - \frac{1}{R}COV\left(\phi_{t+1};\frac{\partial Y_t}{\partial I_t}\right)}{1+Et\left[\phi_{t+1}\right]}\right\} \tag{25}$$

and

$$Et\left[\frac{\partial Y_t}{\partial K_t}\right] = \left\{R\left[1+\Phi_t^K\right]-(1-\delta_k)\right\} + \frac{1-\delta_k}{1+Et\left[\phi_{t+1}\right]}Et\left[\mu_{t+1}\right] \tag{26}$$

where

$$\Phi_t^K = \left[\frac{(1-v)\,\lambda_t - \mu_t - \frac{1}{R}COV\left(\phi_{t+1};\frac{\partial Y_t}{\partial I_t}\right)}{1+Et\left[\phi_{t+1}\right]}\right] \tag{27}$$

Eqs. (18), and (19) are the first-order conditions of b_t and TC_t respectively. Eq. (20) is the first-order condition of I_t when financed with trade credit TC_t. Eq. (24) is the first-order condition of I_t when financed with bank credit b_t. Eq. (26) is the first-order condition of K_t, financed with bank loan. In

As well as in Caggese (2007), by iterating forward Eq. (18) over the three periods, we obtain

$$\phi_t = \sum_{j=0}^{2} E_t\lambda_{t+j} \tag{28}$$

ϕ_t is the shadow value of money, i.e. the shadow cost of the non-negativity constraint of profits. λ_t is the shadow cost of a binding bank financing constraint. Eq. (28) shows that ϕ_t is equal to the sum of the current and future costs of a binding financing constraint. As long as $\phi_t > 0$, then the return from investing earnings inside the firm is higher than r, and the firm does not distribute profits, so $\pi_t = 0$. σ_t is the shadow cost of a binding trade credit constraint.

Now we combine the flow-of-funds equation (10) with the bank borrowing (collateral) constraint (5) and the trade credit constraint (10) in order to define the maximum investment capacity of the firm at time t:

$$I_t + \omega L_t + K_t \leq W_t^F + (1-\delta_k)K_{t-1} + vK_t + \overline{TC} - \pi_t \tag{29}$$

or, analogously

$$I_t + \omega L_t + (1 - \upsilon) K_t \leq W_t^F + (1 - \delta_k) K_{t-1} + \overline{TC} - \pi_t \tag{30}$$

Eq. (30) implies that the down payment necessary to buy K_t and I_t must be lower than the residual net worth, defined as $W_t^F + (1 - \delta) K_{t-1}$, after distributing profits plus the maximum level of trade credit which the supplier is willing to concede.

In the Eq. (26), the optimality condition of K_t, the left-hand side is the marginal productivity of fixed capital and the right-hand side the marginal cost of fixed capital. The term Φ_t^K is equal to zero if the firm is not financially constrained (in term of bank credit) today or in the future. The terms μ_t and $Et\left[\mu_{t+1}\right]$ measure the shadow cost of a currently binding irreversibility constraint and of future expected irreversibility constraints, respectively.

Eq. (20) is the optimality condition of I_t when financed with trade credit TC_t. As well as in the optimality condition of K_t, the left-hand side is the marginal productivity of inventories and the right-hand side the marginal cost of inventories. The term Ψ_t^I is directly related to σ_t , the Lagrange multiplier of the trade credit constraint. Eq. (24) is the optimality condition of I_t when financed with bank credit b_t. Analogously, the left-hand side is the marginal productivity of inventories and the right-hand side the marginal cost of inventories. The term Φ_t^I is directly related to λ_t , the Lagrange multiplier of the bank borrowing-constraint.

In our model, the firm, in dependence of the intensity of financing constraints in term of bank credit and trade credit, can consider the optimal inventory financing portfolio composed of bank credit and trade credit, considering the two sources of funding as substitutes or complementary. Formally, the optimality condition of inventories, when financed with both financing sources, is the following:

$$\widehat{Et\left[\frac{\partial Y_t}{\partial I_t}\right]} = M \cdot Et\left[\frac{\partial Y_t}{\partial I_t}\right]_{TC} + (1 - M) \cdot Et\left[\frac{\partial Y_t}{\partial I_t}\right]_b \tag{31}$$

Where M is the fraction of inventory investment the firm decides to finance with trade credit.

If constraint (29) is not binding then $\lambda_t, \sigma_t = 0$. In this case the firm can finance all the desired amount of inventories with the bank credit (at a lower interest rate) and Eqs. (24) and (26) determine the optimal unconstrained capital levels K_t^u and I_t^u. If K_t^u is greater than $(1 - \delta_k) K_{t-1}$ then the irreversibility constraint (3) is not binding, the Lagrange multiplier μ_t is equal to zero, and $\{K_t^*, I_t^*\}$, the optimal investment choices are determined by $\{K_t^u, I_t^u\}$. If K_t^u is smaller than $(1-\delta_k)K_{t-1}$ then the irreversibility constraint is binding, K_t is constrained to be equal to $(1-\delta_k)K_{t-1}$, and Eqs. (24) and (26) can be solved to determine I_t^{ic} and μ_t^{ic} In this case the optimal investment choices $\{K_t^*, I_t^*\}$ are determined by $\{(1-\delta_k)K_{t-1}, I_t^{ic}\}$. Alternatively the collateral constraint and the trade credit constraint are both binding if financial wealth is not sufficient as a down payment for $\{K_t^*, I_t^*\}$, even if distributed profits are zero:

$$I_t^* + \omega L_t + (1-v)K_t^* > W_t^F + (1-\delta_k)K_{t-1} + \overline{TC} - \pi_t \tag{32}$$

In this case the constrained levels of capital K_t^c and inventories I_t^c are such that:

$$I_t^c + \omega L_t + (1-v)K_t^c = W_t^F + (1-\delta_k)K_{t-1} + \overline{TC} - \pi_t \tag{33}$$

3.1 Model Implications

In the Eq. (20), the term Ψ_t^D captures the effect of trade credit interest rates r_1 and r_2 on the marginal cost of inventories. Depending of the magnitude of D, i.e. the fraction of TC_t which will be repaid within the trade credit contract terms, r_1 and r_2 affect the value of the marginal cost of inventories. More precisely:

- if $D = 1$ (i.e. $TC_t^D = 0$ since the total amount of TC_t is repaid within the contracted terms):

$$\Psi_t^D = \frac{R_1}{R} \tag{34}$$

- if $D = 0$ (i.e. $TC_t^D = TC_t$ since the total amount of TC_t is repaid one period later):

$$\Psi_t^D = \left(\frac{R_2}{R}\right)^2 \cdot \frac{1 + Et\left[\phi_{t+2}\right]}{1 + Et\left[\phi_{t+1}\right]} \tag{35}$$

If the firm expects to be equally financially constrained at $t+1$ and at $t+2$ (i.e. the expected shadow value of money does not change from period two to period three, and $Et\left[\phi_{t+1}\right] = Et\left[\phi_{t+2}\right]$) the term Ψ_t^D is decreasing in D. In other terms the greater the portion of trade credit TC_t which will be repaid one period later, the higher is the marginal cost of inventories when financed with trade debt, due to the combined effect of the trade credit interest rates. Formally, $\dfrac{\partial \Psi_t^D}{\partial D} < 0$, and the term D capture the increasing effect of higher trade credit interest rates on the marginal cost of inventories and, consequently, their decreasing effect on inventory investment level.

For the tractability of the next sections, it's useful to define the spending power of the firm at time t as the amount of financial liquid resources with which it can finance its inventory investment decisions. Formally:

- Spending power at time t, not considering trade credit as source of funding:

$$SP_t = W_t^F + b_t \tag{36}$$

- Spending power at time t, considering trade credit as source of funding:

$$SP_t^{TC} = W_t^F + b_t + TC_t \tag{37}$$

The term Ψ_t^I summarizes the effect of trade credit financing constraint on inventory investment and is a monotonously decreasing and convex function of W_t^F and TC_t as stated in the following proposition, inspired by Proposition 1 in Caggese (2007):

Proposition 1 *We define $\overline{SP_t^{TC}}$ as the level of spending power (i.e. internal financial wealth plus bank credit and trade credit obtained at t) such that the firm does not expect to be financially constrained now or in the future. In other terms $\overline{SP_t^{TC}}$ is the level of spending power that allows the firm to purchase the desired levels of investments in fixed capital and inventories. It follows that, for a given value of the state variables θ_{t-1} and K_t and for $SP_t^{TC} < \overline{SP_t^{TC}}$, Ψ_t^I is positive and is decreasing and convex in the amount of internal finance financial wealth and in the amount of trade credit granted, that is, $\dfrac{\partial \Psi_t^I}{\partial W_t^F} < 0$, $\dfrac{\partial \Psi_t^I}{\partial TC_t} < 0$ and $\dfrac{\partial^2 \Psi_t^I}{\partial \left(W_t^F\right)^2} > 0$, $\dfrac{\partial^2 \Psi_t^I}{\partial \left(TC_t\right)^2} > 0$. This implies that, for $SP_t^{TC} < \overline{SP_t^{TC}}$, Ψ_t^I is positive and is decreasing and convex in the amount of the spending power SP_t^{TC}. Conversely, if the amount $(W_t^F + vK_t + \overline{TC})$ is such that $SP_t^{TC} \geq \overline{SP_t^{TC}}$, then $\Psi_t^I = 0$*

Proof. See Appendix 1 ■

Conversely the term Φ_t^I summarizes the effect of bank credit financing constraint on inventory investment and, analogously to the term Ψ_t^I , is a monotonously decreasing and convex function of W_t^F and TC_t

Proposition 2 *For a given value of the state variables θ_{t-1} and K_t and for $SP_t^{TC} < \overline{SP_t^{TC}}$, Φ_t^I is positive and is decreasing and convex in the amount of internal financial wealth and in the amount of trade credit granted, that is,* $\frac{\partial \Phi_t^I}{\partial W_t^F} < 0$, $\frac{\partial \Phi_t^I}{\partial TC_t} < 0$ *and* $\frac{\partial^2 \Phi_t^I}{\partial (W_t^F)^2} > 0$, $\frac{\partial^2 \Phi_t^I}{\partial (TC_t)^2} > 0$. *This implies that, for $SP_t^{TC} < \overline{SP_t^{TC}}$, Φ_t^I is positive and is decreasing and convex in the amount of the spending power SP_t^{TC}. Conversely, if the amount $(W_t^F + \overline{TC})$ is such that $SP_t^{TC} \geq \overline{SP_t^{TC}}$, then $\Phi_t^I = 0$*

Proof. The proof is analogous to the proof of the proposition 1. See Appendix 1 ■

Proposition 2 applied to Eq. (24) establishes a link between trade credit availability and the inventory investment decisions of firms. It says that when a firm is financially constrained due to credit rationing from the bank-credit market side, the availability of trade credit increases the investment in inventories and reduces its marginal return, due to an increasing effect on the firm's spending power. The intensity of trade credit financing constraint affect directly the value of λ_t and, through it, the marginal return of inventories. It is important to note that Proposition 3 cannot be applied to fixed capital investment because of the presence of the irreversibility constraint. As well as in Caggese (2007), if the irreversibility constraint is binding, then $K_t = (1 - \delta_k)K_{t-1} \Longrightarrow \Delta K_t$ and $\mu_t > 0$. In this case a change in the intensity of trade credit financing constraints, which causes a change in Φ_t^K in Eq. (26) affects the value of μ_t but does not affect fixed capital investment.

Proposition 2 applied to Eq. (20), analogously, shows the effect of a relaxed trade credit constraint on the marginal cost of inventories. An increase of $\overline{TC}$ (i.e. the max possible amount of the credit the supplier is willing to grant), increases the spending power and has a decreasing effect on Ψ_t^I, by affecting the value of σ_t. The consequence is an increase in inventory investment and a reduction of its marginal return.

Eq. (20) also shows the effect of the defaultable portion of the trade credit obtained at time t, i.e. TC_t^D, which is captured by the $InsuranceEffectTerm$. According to Eqs. (36) and (37), we define the spending power of the firm at time $t+1$:

$$SP_{t+1}^{TC} = W_{t+1}^F + b_{t+1} + TC_{t+1} - Liq.Shock \tag{38}$$

where $W_{t+1}^F = Y_t - b_t \cdot R - D \cdot TC_t \cdot R_1$.

Proposition 3 *For a given value of the state variables θ_t and K_{t+1}, the term $Et\left[\frac{\partial Y_{t+1}}{\partial TC_t^D}\right]$ in Eq. (23) is positive and is decreasing and convex in the amount of the spending power* *available at time $t+1$. Furthermore, it is decreasing and convex in the amount of the internal financial wealth W^F_{t+1}, the bank credit b_{t+1}, and the defaultable portion of trade credit TC^D_t. Formally* $\dfrac{\partial Et\left[\dfrac{\partial Y_{t+1}}{\partial TC^D_t}\right]}{\partial SP^{TC}_{t+1}} < 0$, $\dfrac{\partial Et\left[\dfrac{\partial Y_{t+1}}{\partial TC^D_t}\right]}{\partial W^F_{t+1}} < 0$, $\dfrac{\partial Et\left[\dfrac{\partial Y_{t+1}}{\partial TC^D_t}\right]}{\partial b_{t+1}} < 0$

Proof. See Appendix 1 ■

Proposition 3 applied to Eq. (20) establishes a link between the defaultable portion of trade credit obtained at t (the repayment of which can be deferred to $t+2$) and the amount of inventory investment decided at t and financed with trade credit. The higher the intensity of financing constraint the firm expects to be subject to (due to credit rationing or due to the extent of the Liquidity Shock) in the subsequent period, the higher the expected marginal productivity of the defaultable portion of trade credit obtained at t. More precisely, if the firm expects to be subject to a liquidity shortage and that the amount of expected revenues plus the expected amount of available bank credit (which both affect the spending power available at $t+1$) will be not sufficient to absorb the liquidity shock, the expected marginal productivity of inventories financed with the defaultable portion of trade credit is relatively high. Consequently, the expected level of inventory investment, when financed only with internal finance plus constrained bank credit, without considering the defaultable portion of trade credit, is relatively low, far from the desired optimal level. This means that the financially constrained firm, in presence of a positive probability of suffering future liquidity shortage tomorrow, is has incentive to finance inventories with trade credit today, because the possibility to defer the repayment of a fraction of the trade debt beyond the contractual terms makes possible for the firm to maintain a level of expected inventory investments (and, as a consequence, a level of future expected output) as close as possible to the optimal desired level. In other terms, the trade credit contract yields an insurance effect, i.e. an insurance coverage against liquidity risk, which explains why the liquidity constrained firm has incentive to finance inventories with trade credit, in spite of its high implicit cost. The higher the intensity of expected future financing constraints, the higher the value of the expected marginal productivity of the defaultable portion of the trade credit and, consequently, the higher the incentive for the firm to finance inventory investment decisions with trade credit today.

If the firm, considering the bank credit market side, is not currently financially constrained and expects to be not financially constrained in the future, the terms Φ^I_t and Ψ^I_t

are equals to zero. In this case, Eqs. (20) and (24) becomes as follows:

$$Et\left[\frac{\partial Y_t}{\partial I_t}\right]_b = R \tag{39}$$

and

$$Et\left[\frac{\partial Y_t}{\partial I_t}\right]_{TC} = R \cdot \Psi_t^D + \frac{1}{R} \cdot InsuranceEffectTerm \tag{40}$$

where, in absence of future expected financial constraints,

$\Psi_t^D = \left\{\left(\frac{R_2}{R}\right)^2 + D\left[\frac{R_1}{R} - \left(\frac{R_2}{R}\right)^2\right]\right\}$, and

$InsuranceEffectTerm = -Et\left[\frac{\partial Y_{t+1}}{\partial TC_t^D}\right]$.

Eq. (39) shows that, if the firm is not financially constrained today or in the future, the marginal cost of inventories, when financed with bank credit, is equal to R. Without liquidity constraints, the extent of the $InsuranceEffectTerm$ is relatively low (due to Proposition 4) and, considering that the firm is able to repay the total amount of trade debt within the initial contractual terms (i.e. $D = 1$), $\Psi_t^D = \frac{R_1}{R}$. In this case, Eq. (40) shows that the trade credit one-period interest rate has an increasing effect on marginal cost of inventories which is balanced by the decreasing effect of the insurance motive. But, in absence of future expected liquidity constraints, the latter is weaker compared to the case in which the firm expects to be financially constrained. As a consequence, for sufficiently high values of R_1, it would be optimal to finance the total amount of inventories with bank credit. Conversely, if the firm expects to be financially constrained today and in the next period, $\Phi_t^I > 0$ and $\Psi_t^I > 0$, and both increase as the amount of the spending power decrease (Propositions 2,3). Furthermore, the $InsuranceEffectTerm$ is relatively high and increases as the future expected spending power decreases (Proposition 4). When the bank-borrowing constraint is binding, if the firm expects to suffer a future liquidity shortage, it would be optimal for the firm to finance inventories with a portfolio of bank credit and trade credit, since the higher decreasing effect of the $InsuranceEffectTerm$ (i.e. the marginal productivity of the defaultable portion of trade credit on the expected output) on the marginal cost of inventories balances in a stronger fashion the increasing effect of trade credit interest rates guided by the fraction D. In other terms, the higher the intensity of future expected financial constraints (or, equivalently, the higher the intensity of the expected bank-borrowing constraint which affect the expected spending power), the higher the marginal productivity of TC_t^D on Y_{t+1} (i.e. the higher the absolute value of the $InsuranceEffectTerm$), the lower the marginal cost of inventories I_t and, since the production function is concave in all the factors, the higher is the level of inventory investment financed with trade credit at time t, for a given value of K_t. If the firm expects the occurrence of a sufficiently high liquidity shock for which it expects to be subject to a binding financial constraint from the bank-lending market side, for sufficiently low values of trade credit one-period and two-periods interest rates, r_1 and r_2 respectively, it would be optimal to finance the largest possible portion of inventories with trade credit today, because the possibility to defer the repayment of the trade debt to the next period (at an interest rate equal to r_2), makes the firm able to maintain a sufficiently high level of future expected inventory investment and so a sufficiently high level of future expected output, as close as possible to the optimal desired level.

3.2 Model Extension

In this extension of the model, we consider the term D, i.e. the fraction of TC_t which the firm repays at $t+1$, as a choice variable of the maximization problem. More precisely, the firm is allowed by the supplier to decide the extent of D, and, consequently, the amount of TC_t which will be repaid at $t+2$ at an interest rate equal to $r_2 > r_1$. The firm now maximizes the current profit plus the discounted sum of expected profits of the two subsequent periods, considering D as choice maximization variable. The first order condition of D is as follows:

$$TC_t = \frac{Et\left[\frac{\partial Y_{t+1}}{\partial D}\right] + \frac{COV\left(\frac{\partial Y_{t+1}}{\partial D}; \phi_{t+2}\right)}{\left(1 + Et\left[\phi_{t+2}\right]\right)}}{\left(R_1 R \frac{\left(1 + Et\left[\phi_{t+1}\right]\right)}{\left(1 + Et\left[\phi_{t+2}\right]\right)} - R_2^2\right)} \tag{41}$$

Where the term $\frac{\partial Y_{t+1}}{\partial D}$ is the marginal productivity of D concerning the expected output Y_{t+1} which will be observed at $t+2$. Intuitively, $\frac{\partial Y_{t+1}}{\partial D}$ is negative, because repaying an additional unit of trade credit TC_t at $t+1$ is equivalent to subtract an additional unit of cash flow available to purchase inventories I_{t+1}, and this produces a decreasing effect on the expected output Y_{t+1}. If we assume that the firm expects to be equally financially constrained at $t+1$ and at $t+2$ (i.e. the expected shadow value of money does not change from period two to period three, and $Et\left[\phi_{t+1}\right] = Et\left[\phi_{t+2}\right]$) and change the sign of both the denominator and numerator, we obtain

$$TC_t = \frac{-Et\left[\dfrac{\partial Y_{t+1}}{\partial D}\right] - \dfrac{COV\left(\dfrac{\partial Y_{t+1}}{\partial D};\phi_{t+2}\right)}{(1+Et\left[\phi_{t+2}\right])}}{(R_2^2 - R_1 R)} \tag{42}$$

The marginal productivity of D (with reference to output Y_{t+1}) is negative and has opposite sign respect on the marginal productivity of $(1-D)$. More precisely, an additional unit of the fraction $(1-D)$ of the trade credit TC_t, which will be repaid one period later (i.e. at $t+2$ instead of at $t+1$), is equivalent to add an additional unit of cash flow available at $t+1$ to purchase inventories I_{t+1}, and this produces an increasing effect on the expected output Y_{t+1}. We can rewrite Eq. (42) as follows:

$$TC_t = \frac{Et\left[\dfrac{\partial Y_{t+1}}{\partial(1-D)}\right] - \dfrac{COV\left(\dfrac{\partial Y_{t+1}}{\partial D};\phi_{t+2}\right)}{(1+Et\left[\phi_{t+2}\right])}}{(R_2^2 - R_1 R)} \tag{43}$$

Eq. (43) is the first order condition of the trade credit obtained at t. It shows that the optimal level of TC_t is decreasing in trade credit interest rates r_1 and r_2[3], and increasing in the term $Et\left[\dfrac{\partial Y_{t+1}}{\partial(1-D)}\right]$, i.e. the expected marginal productivity of the defaultable fraction of TC_t with reference to the expected output Y_{t+1} which will be observed at $t+2$. In other terms, the higher the expected marginal productivity of the defaultable fraction of TC_t, the higher the incentive for the financially constrained firm to ask its supplier for more credit, in a context in which the firm expects a liquidity shock to happen in the next period with a positive probability.

4 Empirical Framework (draft)

In this section, in order to verify the validity of the model implications, we propose an empirical framework which is able to determine that, under liquidity and financial constraints, the dynamics of trade credit is more significant than bank credit dynamics in explaining the dynamics of inventory investments. To this end, we derive a reduced-form inventory

[3]Since $r_2 > r_1$, if r_1 increases, r_2 must increase too, for at least the same extent.

investment equation and test it on a sample of small and medium Italian manufacturing firms. The sample is obtained by the dataset AIDA (Italian Company Information and Business Intelligence), provided by Bureau Van Dijk, which contains comprehensive information on companies in Italy, with up to five years of history. Aida covers 1 million companies in Italy over a period of ten years.

For the empirical specification of the test framework we consider the following specification of the production function in the Eq. (1),

$$Y_{i,t} = \theta_{i,t} K^{\alpha}_{i,t-1} I^{\beta}_{i,t} L^{\gamma}_{i,t} \tag{44}$$

defining the variables (in real terms) as follows: $Y_{i,t}$ is total revenues (during period t, firm i); $K_{i,t-1}$ is the value of plant, equipment, and intangible fixed capital (end of period $t-1$, firm i); $I_{i,t}$ is the stock of inventories, more precisely, the stock of intermediate goods that is used in the production process: raw materials, work-in-process goods and completely finished goods that are considered to be the portion of a business's assets that are ready or will be ready for sale (during period t, firm i); $L_{i,t}$ is the labor cost (during period t, firm i). We assume that fixed capital installed in period t will become productive from period $t+1$ on. More detailed information about all the variables is reported in Appendix 2.

By using Eq. (44) in the optimality condition specified by Eq. (20), we obtain:

$$\beta \theta_{i,t} K^{\alpha}_{i,t-1} I^{\beta-1}_{i,t} L^{\gamma}_{i,t} = R \cdot \Psi^D_t + R \cdot \Psi^I_t + \frac{1}{R} \cdot InsuranceEffectTerm \tag{45}$$

If we take logs of both sides of Eq. (45) and solve for $\ln I_{i,t}$, we obtain:

$$\begin{aligned} \ln I_{i,t} \quad = \quad & \frac{1}{1-\beta} \ln \frac{\beta}{R} + \frac{1}{1-\beta} \ln Et\theta_{i,t} + \frac{\alpha}{1-\beta} \ln K_{i,t-1} + \frac{\gamma}{1-\beta} \ln L_{i,t} + \\ & - \frac{1}{1-\beta} \ln \left[\Psi^D_t + \Psi^I_t + \frac{1}{R^2} \cdot InsuranceEffectTerm \right] \end{aligned} \tag{46}$$

As discussed in the previous section, the term Ψ^D_t is increasing in the amount of trade credit interest rates. This allows us to approximates it with a positive function of $\frac{\overline{R^{TC}_t}}{R}$, where $\overline{R^{TC}_t} = (1 + \overline{r^{TC}_t})$ and $\overline{r^{TC}_t}$ is the trade credit implicit interest rate on annual basis. We approximate Ψ^D_t as follows:

$$\Psi^D_t = \left(\frac{\overline{R^{TC}_t}}{R} \right)^{\eta_1} \tag{47}$$

where η_1 is an indicator of the sensitivity of the inventory investment to a change in the trade credit interest rates.

Proposition 2 allows us to approximate the term Ψ_t^I with a negative and convex function of the terms $\left\{\frac{\overline{SP_t^{TC}}}{W_t^F}, \frac{\overline{SP_t^{TC}}}{b_t}, \frac{\overline{SP_t^{TC}}}{TC_t}\right\}$, where $\overline{SP_t}$ is the level of total spending power at time t, considering internal financial wealth plus bank credit and trade credit, that allows the financing of all profitable investment projects, i.e. the level of financial wealth such that the firm does not expect to be financially constrained now or in the future (and for which the bank borrowing constraint and the trade credit constraint, defined by the upper limits vK_t and $\overline{TC}$ respectively, are not binding). We approximate Ψ_t^I as follows:

$$\Psi_t^I = \left(\frac{\overline{SP_t^{TC}}}{W_t^F}\right)^{\eta_2} \cdot \left(\frac{\overline{SP_t^{TC}}}{b_t}\right)^{\eta_3} \cdot \left(\frac{\overline{SP_t^{TC}}}{TC_t}\right)^{\eta_4} \tag{48}$$

where η_2, η_3, η_4 are measures of the responsiveness of the inventory investment to a change in the available internal finance, bank credit and trade credit, respectively. In other terms η_2, η_3, η_4 are indicators of the intensity of the internal finance constraint, the bank borrowing constraint and the trade credit constraints respectively. The more the firm is financially constrained (in the model, this corresponds to lower values of v and $\overline{TC}$) which tightens the financing constraints, the more the investment of the firm is sensitive to internal finance (meaning that Ψ_t^I increases more rapidly as W_t^F decreases), and the larger η_2 is. Furthermore, η_3 and η_4 isolate the intensities of the bank borrowing constraint and the trade credit constraint: the more the investment of the firm is sensitive to bank credit, (meaning that Ψ_t^I increases more rapidly as b_t decreases), the larger η_3 is; the more the investment of the firm is sensitive to trade credit, (meaning that Ψ_t^I increases more rapidly as TC_t decreases), the larger η_4 is.

Eq. (23) shows that the $InsuranceEffectTerm$ is a negative function of the term $Et\left[\frac{\partial Y_{t+1}}{\partial TC_t^D}\right]$ and Proposition 4 explains that the latter is decreasing and convex in the amount of the spending power available at time $t+1$, SP_{t+1}^{TC}, and so it is decreasing and convex in the amount of the internal financial wealth W_{t+1}^F and the bank credit b_{t+1}. This means that the $InsuranceEffectTerm$ is negative and is increasing in W_{t+1}^F and b_{t+1}. It follows that if the firm expects to be financially constrained in the next period, cause a combined effect of the binding bank borrowing constraint and the extent of the liquidity shock, the expected marginal productivity of the defaultable portion of trade credit obtained in the previous period is relatively high. According to Eqs. (23) and (46), this means that the firm has a relatively high incentive to increase the amount of inventory investment financed with trade credit TC_t. The lower the internal financial wealth $W_{t+1}^F$and

the higher the intensity of the bank borrowing constraint (corresponding to a lower upper limit for b_{t+1}) and/or the higher the extent of the expected Liquidity Shock (and so, the lower the value of the spending power SP^{TC}_{t+1})[4], the higher is the increasing effect of the $InsuranceEffectTerm$ on the amount of inventory investment purchased at time t and financed with TC_t. The marginal productivity of the defaultable trade credit TC^D_t on the output Y_{t+1} is not observable in reality. But it can be approximate with a positive function of the total trade credit to internal financial wealth ratio at time $t+1$. The intuition is as follows: if the firm expects to be financially constrained from the bank borrowing side and to suffer a liquidity shortage due to a liquidity shock at time $t+1$ (i.e. expected future values of W^F_{t+1} and SP^{TC}_{t+1} are relatively low), it increases the portion of inventories purchased with trade credit TC_t, since the therm $Et\left[\dfrac{\partial Y_{t+1}}{\partial TC^D_t}\right]$ is relatively high and so the $InsuranceEffectTerm$ affects negatively the marginal cost of inventories and positively the level of inventory investment (since the concavity of the production function). If the firm will actually be under liquidity constraints in the next period $(t+1)$, it intuitively will decide to defer the repayment of the fraction $(1-D)$ of the trade debt TC_t in order to have sufficient cash flow to purchase a level of inventories I_{t+1} as close as possible to the optimal desired level I^*_{t+1}. The consequence is that the firm, at time $t+1$, is shifting its inventory financing portfolio towards an higher fraction of trade credit compared to the fractions of internal financial wealth and bank debt, i.e. the firm is increasing the following ratios:
$\dfrac{TC^D_t}{SP^{TC}_{t+1}}, \dfrac{TC^D_t}{W^F_{t+1}}, \dfrac{TC^D_t}{b_{t+1}}$.

This allows us to approximate the absolute value of the $InsuranceEffectTerm$ with a positive function of one of the ratios above. We define:

$$TC^{ratio}_{t+1} = \frac{TC^D_t}{SP^{TC}_{t+1}} \tag{49}$$

where SP^{TC}_{t+1} is the total spending power of the firm at time $t+1$, as defined in Eq. (38). We approximate the $InsuranceEffectTerm$ in modulus as follows:

$$|InsuranceEffectTerm| = \left(\frac{TC^D_t}{SP^{TC}_{t+1}}\right)^{\eta_5} = \left(TC^{ratio}_{t+1}\right)^{\eta_5} \tag{50}$$

where η_5 is a measures of the responsiveness of the inventory investment to a change in TC^{ratio}_{t+1}. Considering Eqs. (47), (48), (49), (50), and considering that the $InsuranceEffectTerm$

[4] All the three phenomena correspond to a lower value of the spending power SP^{TC}_{t+1}

enters in Eq. (46) with negative sign, we set:

$$\left[\Psi_t^D + \Psi_t^I + \frac{1}{R^2}\cdot IET\right] = \frac{\left(\frac{\overline{R_t^{TC}}}{R}\right)^{\eta_1} \cdot \left(\frac{\overline{SP_t^{TC}}}{W_t^F}\right)^{\eta_2} \cdot \left(\frac{\overline{SP_t^{TC}}}{b_t}\right)^{\eta_3} \cdot \left(\frac{\overline{SP_t^{TC}}}{TC_t}\right)^{\eta_4}}{\left(TC_{t+1}^{ratio}\right)^{\eta_5}} * \frac{1}{R^2} \quad (51)$$

Inserting Eq. (51) in Eq. (46) we obtain the following reduced-form inventories equation:

$$\begin{aligned} \ln I_{i,t} &= \Pi_0 + \Pi_1 \ln \theta_{i,t-1} + \Pi_2 \ln K_{i,t-1} + \Pi_3 \ln L_{i,t} + \Pi_4 \ln \overline{R_t^{TC}} + \\ &\quad + \Pi_5 \ln W_{i,t}^F + \Pi_6 \ln b_{i,t} + \Pi_7 \ln TC_{i,t} + \Pi_8 \ln TC_{i,t+1}^{ratio} + \varepsilon_{i,t} \qquad (52) \end{aligned}$$

where the subscript i indicates the i^{th} firm, $\varepsilon_{i,t}$ is the error term and $\theta_{i,t-1}$ is the productivity shock[5]. The coefficients are as follows:

- $\Pi_0 = \left[\frac{1}{1-\beta}\ln\frac{\beta}{R} + \frac{1}{1-\beta}\ln R^2 + \frac{\eta_1}{1-\beta}\ln R - (\frac{\eta_2+\eta_3+\eta_4+\eta_5}{1-\beta})\ln \overline{SP_t^{TC}}\right]$
- $\Pi_1 = \frac{1}{1-\beta}$
- $\Pi_2 = \frac{\alpha}{1-\beta}$
- $\Pi_3 = \frac{\gamma}{1-\beta}$
- $\Pi_4 = -\frac{\eta_1}{1-\beta}$
- $\Pi_5 = \frac{\eta_2}{1-\beta}$
- $\Pi_6 = \frac{\eta_3}{1-\beta}$
- $\Pi_7 = \frac{\eta_4}{1-\beta}$
- $\Pi_8 = \frac{\eta_5}{1-\beta}$

[5] In this version of the paper we omit the specification of the stochastic idiosyncratic productivity shock θ_t and its estimation. In the next version of the paper we will model it as a stochastic process and specify how it is related with the liquidity shock.

In this version of our paper we skip the estimation of the productivity shock θ, and we consider only the liquidity shock during the second period, occurring with probability 1: this means that we consider $\theta_{i,t-1} = 1$ and, as a consequence, $\ln \theta_{i,t-1} = 0$. We lagged the ratio of one period and we use the trade credit amount $TC_{i,t}$ as numerator of the ratio, instead of TC_t^D, considering that TC_t^D is empirically not observable and that, conceptually, all the trade credit amount is defaultable. Futhermore, in order to avoid possible multicollinearity problems, we decide to maintain the variable $TC_{i,t}$ only inside the ratio $TC_{i,t}^{ratio}$[6]The Eq. (52) becomes as follows:

$$\begin{aligned} \ln I_{i,t} &= \Pi_0 + \Pi_1 \ln K_{i,t-1} + \Pi_2 \ln L_{i,t} + \Pi_3 \ln \overline{R_t^{TC}} + \\ &\quad + \Pi_4 \ln W_{i,t}^F + \Pi_5 \ln b_{i,t} + \Pi_6 \ln TC_{i,t}^{ratio} + \varepsilon_{i,t} \end{aligned} \tag{53}$$

The variables $I_{i,t}$, $b_{i,t}$ and $TC_{i,t}$ (as numerator of the ratio) are in term of stocks. In the following subsection we consider an alternative empirical framework in which we put inside the regression framework the changes in the variables, i.e. the flows, instead of the absolute values (i.e. the stocks). The rationale under considering the stocks is to consider the investment decisions of the firm from a mid-term perspective. The current stocks of I, b and TC are consequences of investment decisions taken during a certain number of previous periods.

The coefficients Π_4, Π_5 measure the functional relationship between the investment in inventories (the dependent variable of our regression framework) and the internal financial wealth and the bank credit, respectively. The coefficient Π_6 explains how the dynamics of inventory investment is related to the value of the $TC_{i,t}^{ratio}$. This term, which is specified by lagging Eq. (49) by one period, is the ratio between the portion of trade credit the firm has decided to repay one period later respect on the initial contractual terms (i.e. TC_t^D) and the total spending power amount (i.e. SP_t^{TC}). The term TC_t^D is not observable in reality. As a proxy we use the value of Accounts Payable plus any other debts towards suppliers, no distinguishing on the basis of the repayment terms.

We compute $W_{i,t}^F$, net financial wealth of firm i at the beginning of period t, by using the budget constraint (10) at time $t-1$ to substitute $b_{i,t}$ in (8), considering the amount of trade credit obtained in the previous period as totally defaulted[7]:

$$W_t^F = Y_{t-1} - R(I_{t-1} + \omega L_{t-1} + \Delta K_{t-1}) + R(W_{t-1}^F - \pi_{t-1}) \tag{54}$$

[6]In the empirical framework, the ratio becomes as follows: $TC_{i,t}^{ratio} = \dfrac{TC_{i,t}}{SP_{i,t}^{TC}}$

[7]I.e. the term TC_{t-1} in the budget constraint is zero.

In the model, the term $Y_{t-1} - R(I_{t-1} + \omega L_{t-1} + \Delta K_{t-1})$ represents the beginning-of-period t profits generated from the investment in period $t-1$. We therefore estimate $Y_{t-1} - R(I_{t-1} + \omega L_{t-1} + \Delta K_{t-1})$ as the operating profits during period $t-1$ (value of production minus the cost of production inputs). Moreover, we estimate $(W^F_{t-1} - \pi_{t-1})$ as the net short-term financial assets at the beginning of period $t-1$. Moreover, we estimate SP^{TC}_t as the operative profits during period $t-1$ (value of production minus the cost of production inputs)[8], plus net short-term financial assets of period $t-1$ multiplied by one plus the nominal interest rate, plus the amount of short-term debt obtained at time t from the lending-market side (bank credit), plus the amount of Accounts Payable at time t.

On the basis of the model implications, our theoretical results suggest the following hypotheses:

- Π_1, $\Pi_2 > 0$
- $\Pi_3 < 0$
- Π_4, $\Pi_5 > 0$
- $\Pi_6 > 0$ with $\Pi_6 > \Pi_5$ and $\Pi_6 > \Pi_4$

The regression framework is under implementation at the moment of the PhD thesis defense day. After a set of preliminary regressions, the results seem to confirm the predictions of the analytical model.

4.1 Estimation of the trade credit interest rate

The estimation of $\overline{R^{TC}_t}$ give rise to a brief analysis of the trade credit conditions and forms. Generally, a trade transaction between firms can be settled in three ways:

1. Cash on delivery
2. Cash before delivery
3. Cash after delivery

[8] I.e. the beginning-of-period t profits generated from the investment in period $t-1$.

Only the third case is identified as trade credit.

Basicly trade credit transactions can be regulated in two ways:

1. Net Terms. This is the simplest form: the payment has to be made within certain time. In other terms, no interest will be charged if payment is made within the specified period, usually 30, 60, 90, or 120 days

2. Two-part Terms: This is the most complex form: it provides the possibility for the buyer to receive a discount on the price agreed if he/she pays within a shorter period. In other terms, a two-part offer of trade credit adds a discount period during which the purchaser may take a discount if payment is made within an even shorter period. For example, a "2/10 Net 30" offer means that the buyer has the option of taking a 2 percent discount if payment is made within ten days. Otherwise, full payment is expected within 30 days.

In the last case, there are three elements defining the trade credit conditions: the discount percentage offered by the supplier, the time period within which the buyer must repay the discounted price and the time period within which the buyer must repay the full price. These three elements define ex-ante what is the implicit cost of the trade credit for the buyer[9].

---

[9] In their paper Cannari, Chili, Omiccioli (2014) provide a formula to estimate the trade credit interet rate as a measure of the ex-ante trade credit implicit cost on annual basis, adopting the following specification:

$$\overline{R_t^{TC}} = 1 + \left[\left(\frac{100}{100-s} \right)^{\frac{360}{N+rit-n}} - 1 \right] * 100 \tag{55}$$

where s is the discount (in percentage) offered for an early repayment within contractual defined terms, N is the number of days within which the firm must pay the full price, n is the number of days within which the firm can pay the discounted price.

In another alternative specification, a repayment delay in payment, beyond the deadline agreed upon, and the possible application of penalties are further elements that define the ex-post implicit cost of trade credit.

In this case:

$$\overline{R_t^{TC}} = 1 + \left[\left(\frac{100+pen}{100-s} \right)^{\frac{360}{N+rit-n}} - 1 \right] * 100,$$

where rit is the number of delay days and pen is the penalty apllied due to the delay.

Two-part trade credit form has an implicit interest charge built in. That is, if the purchaser chooses the "Net 30" option over the "2/10" option and fails to take the trade discount offered, then the purchaser is in effect paying an interest charge on the 30 days of credit that has been extended. When annualized, the interest rate on most trade credit far exceeds that offered by banks and other financial institutions. Consequently, two-part trade credit offers provide suppliers/sellers with information about the creditworthiness of their customers. Creditworthy customers would always take the trade discount, because they could find third-party financing at better rates than are offered by the two-part trade credit offer. The actual credit terms of trade credit offers appear to be standardized within industries, although they may vary from industry to industry. They tend to remain constant and not be affected by supply and demand. The extending of trade credit can serve a business firm's informational and financial needs. In addition to providing sellers with information on the creditworthiness of their customers, trade credit offers can serve to bond relationships between buyers and sellers. Sellers who offer trade credit have a financial interest in maintaining a continuing relationship with their buyers. The extending of trade credit also gives the purchaser time to verify the quality of the goods purchased and evaluate the seller's performance.

According to the Mediocredito Centrale Survey (2001), the diffusion in Italy of the two-part term contract form is very limited. According to the Survey of Small Business Finances, conducted by Federal Reserve in 1998, the percentage of suppliers that offer a discount for a cash repayment is, on average, 22%; for the first quartile of firms, this percentage is zero, the median is 5%. Danielson, Scott (2000) found that, considering a sample of small american firms, the percentage of trade debts with offers of a discount for an earlier repayment is, on average, 24%, with a median of 10%. Summers, Wilson (2003) found that the most part of small-medium firms in UK generally offer net-terms contracts; the percentage that offer two-part terms contracts is 17%., only the 2% of the firms requires a cash payment. Therefore, the most part of trade credit is offered in form of net-terms contracts. Futhermore, the most part of supply firms declares that trade credit is granted without an interest rate, i.e. without additional costs associated to a deferred payment.

On the basis of this empirical evidence, we consider two options in order to include the trade credit interest rate among the determinants of the investment decisions in inventories.

The first option is to include the term $\overline{R_t^{TC}}$ inside the intercept Π_0 of the Eqs. (52), (53), considering it as a constant.

The second option is to consider the interest rate for bank loans to non-financial italian firms as a proxy of the term $\overline{R_t^{TC}}$: more precisely we estimate $\overline{R_t^{TC}}$ as the interest rate for bank loans to non-financial italian firms, provided by Bank of Italy, times two.

4.2 Alternative Testing Strategies

4.2.1 Considering changes (flows) in the variables

In this alternative specification, instead of looking at absolute values, we focus on the changes in inventories, trade credit (accounts payable) and bank credit (short-term debt towards financial institutions) instead of stocks, and how they are related to the change in inventory investment. We modify Eq.(53) as follows:

$$\begin{aligned} \ln \Delta I_{i,t} &= \Pi_0 + \Pi_1 \ln K_{i,t-1} + \Pi_2 \ln L_{i,t} + \Pi_3 \ln \overline{R_t^{TC}} + \\ &\quad + \Pi_4 \ln W_{i,t}^F + \Pi_5 \ln \Delta b_{i,t} + \Pi_6 \ln \Delta TC_{i,t}^{ratio} + \varepsilon_{i,t} \end{aligned} \tag{56}$$

where $\Delta I_{i,t} = I_{i,t} - I_{i,t-1}$, $\Delta b_{i,t} = b_{i,t} - b_{i,t-1}$ and $\Delta TC_{i,t}^{ratio} = \dfrac{\Delta TC_{i,t}}{\Delta SP_{i,t}^{TC}}$ (where, $\Delta TC_{i,t} = TC_{i,t} - TC_{i,t-1}$ and $\Delta SP_{i,t}^{TC} = SP_{i,t}^{TC} - SP_{i,t-1}^{TC}$).

The rationale under considering the flows is to analyse and to investigate the investment decisions of the firm under a short-term perspective. The flows of I, b and TC represents the investment decisions taken in the last period.

4.2.2 Considering absolute values (stocks) over two periods

In this specification, we consider the following modified empirical framework:

$$\begin{aligned} \ln I_{i,t} &= \Pi_0 + \Pi_1 \ln K_{i,t-1} + \Pi_2 \ln L_{i,t} + \Pi_3 \ln \overline{R_t^{TC}} \\ &\quad + \Pi_4 \ln W_{i,t-1}^F + \Pi_5 \ln b_{i,t-1} + \Pi_6 \ln TC_{i,t-1} \\ &\quad + \Pi_7 \ln W_{i,t}^F + \Pi_8 \ln b_{i,t} + \Pi_9 \ln TC_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{57}$$

The regression framework specified in Eq. (57) is aimed to establish the functional relationships between the inventory investment and the three sources of funding considered in this paper: internal financial wealth (after debts repayment), bank credit and trade credit, which together constitutes the spending power available for the firm in order to finance the investment decisions. The test is aimed to establish how the two-period dynamics of the three sources of funding are able to explain the inventory investment decision of the firm. To this end, we consider the absolute value of internal financial wealth and the stock of trade credit and bank credit over two periods (the current period t and the previous period $t-1$) as explanatory variables of the stock of inventory investment in the current period. This means that the amounts of the three sources of funding at time t are related to the amount of inventories at time t but also at time $t-1$. In other terms, internal financial wealth, trade credit and bank credit (and their distribution in the inventory financing portfolio) are able to explain ex-post the firm investment behavior over two periods. Futhermore, in this framework, we consider the stock of trade credit TC, in substitution of the TC^{ratio}, i.e. the proportion of the total spending power of the firm in form of trade debt.

5 Conclusions

In this paper we have illustrated a structural model of a profit maximizing firm subject to bank borrowing constraints and with three sources of funding: self-financing, bank credit and trade credit. The solution of the optimal investment problem shows that, under current and future expected financing constraints, the amount of current and future available trade credit affects investment decisions. The model shows that trade credit yields effects on the investment decisions of a financially constrained firms in manufacturing supply chains, with particular reference to a context of bank borrowing constraints. Trade Credit enhances the resilience of firms to liquidity shocks and, due to the flexibility of repayment terms, embeds an insurance coverage against liquidity risk. The implicit cost of this insurance effect is incorporated in the trade credit interest rate. Due to this insurance effect, trade credit is an optimal source of funding for a financially-constrained firm under future expected liquidity shortage, because the firm can maintain a level of expected inventory investment and, as a consequence, future expected levels of output and revenues, as close as possible to the optimal desired level.

We also proposed an empirical framework in order to test conjectures and implications of the analytical model. A set of econometric regressions over a sample of italian manufacturing firms is under implementation. The sample is obtained by the dataset AIDA (Italian Company Information and Business Intelligence), provided by Bureau Van Dijk: it is consisting of 1 million small and medium companies in Italy over a period of ten years. The results of the empirical analysis will be drawn in the next version of this article.

References

[1] Atanasova, C.V., Wilson, N., 2003, Bank borrowing constraints and the demand for trade credit: evidence from panel data, *Managerial and Decision Economics*

[2] Biais, B. e Gollier, C. (1997), Trade credit and credit rationing, *Review of Financial Studies*, Vol. 10, No.4, 903-937.

[3] Boissay, F. (2006), Credit chains and the propagation of financial distress, European Central Bank, working paper series n° 573.

[4] Boissay, F. e Gropp,R. 2007, Trade Credit Defaults and Liquidity Provision by Firms, European Central Bank, working paper series n° 753.

[5] Breig, H, 1994, Bank lending and corporate finance in major industrial countries: are France and Germany really similar?

[6] Burkart, M. e Ellingsen, T. 2004, In-Kind Finance: A Theory of Trade Credit, *American Economic Review*, Volume 94, Number 3, pp. 569-590(22).

[7] Caggese, A. 2007, Financing Constraints, Irreversibility, and Investment Dynamics, *Journal of Monetary Economics*, Volume 54, pp 2102–2130.

[8] Caggese, A. 2007, Testing Financing Constraints on Firm Investment using Variable Capital, *Journal of Financial Economics*, Volume 86, pp 683–723.

[9] Caggese, A. e Cunat, V. 2008, Financing Constraints and Fixed-Term Employment Contracts, *The Economic Journal*, Volume 118, pp 2013–2046.

[10] Caggese, A. e Cunat, V. 2013, Financing constraints, firm dynamics, export decisions, and aggregate productivity, *Review of Economic Dynamics*, Volume 16, pp 177–193.

[11] Cannari L., Chiri S. e Omiccioli M. 2004, Condizioni di credito commerciale e differenziazione della clientela, Banca d'Italia, Temi di Discussione, n. 495.

[12] Carpenter, R.E., Fazzari, S.M., Petersen, B., 1993, Inventory Investment, Internal-Finance Fluctuations, and the Business Cycle, Brookings Papers on Economic Activity 2 (1994), 75-138.

[13] Carpenter, R.E., Fazzari, S.M., Petersen, B., 1998, Financing constraints and inventory investment: a comparative study with high-frequency panel data, *The Review of Economics and Statistics* 80, 513-519.

[14] Carpenter, R.E., Petersen, B., 2003, Is the growth of small firms constrained by internal finance?, *Review of Economics and Statistics* 84, 298-309.

[15] Cunat, V., 2007, Trade Credit: Suppliers as Debt Collectors and Insurance Providers, *Review of Financial Studies*, 20 (2): 491-527.

[16] Danielson, M. G., Scott, J. A., 2000, Additional Evidence on the Use of Trade Credit by Small Firms: The Role of Trade Credit Discounts, Fox School of Business and Management, Temple University, mimeo; URL: http://ssrn.com/abstract=236260.

[17] Elliehausen, G. E.; Wolken J. D., 1993, The Demand for Trade Credit: An Investigation of Motives of Trade Credit Use by Small Businesses Board of Governors of the Federal Reserve System.

[18] Emery, Gary W., 1984, A Pure Financial Explanation for Trade Credit, *Journal of Financial and Quantitative Analysis* Vol 19 (3) September Pages 271-285.

[19] Ferris, J. Stephen, 1981, A Transactions Theory of Trade Credit Use, *Quarterly Journal of Economics*; 96(2), May, pages 243-70.

[20] Frank, M., Maksimovic, V., 2005, Trade Credit, Collateral, and Adverse Selection, URL: http://ssrn.com/abstract=87868

[21] Gertler, M.,Gilchrist, S., 1993a, The role of credit market imperfections in the monetary transmission mechanism: arguments and evidence, *The Scandinavian Journal of Economics*, Vol. 95, No. 1 (Mar., 1993), pp. 43-64.

[22] Gertler, M.,Gilchrist, S., 1993b, The cyclical behavior of short-term business lending: Implications for financial propagation mechanisms. *European Economic Review*, Volume 37, pp 623–631.

[23] Giannetti, M., Burkart, M. e Ellingsen, T. (2011), What You Sell Is What You Lend? Explaining Trade Credit Contracts, *Review of Financial Studies*, 24 (4): 1261-1298.

[24] Guedes, Jose and Mateus, Cesario (2009) Trade credit linkages along the supply chain: evidence for the Italian textile sector. In: European Financial Management Association 2009 Annual Meetings, 24-27 Jun 2009, Milan, Italy

[25] Kashyap, A., Lamont, O., Stein, J., 1994, Credit conditions and the cyclical behavior of inventories, *The Quarterly Journal of Economics* 109, 565-592.

[26] Kiyotaki, N. e Moore, J.H., 1997, Credit Chains, mimeo, London School of Economics.

[27] Longhofer, S.D., Santos, J.A.C., 2003, The paradox of priority, *Financial Management*, Vol. 32, No. 1 (Spring, 2003), pp. 69-81

[28] Mian, S. L., Smith, C. W., Jr., 1992, Accounts Receivable Management Policy: Theory and Evidence, *Journal of Finance*, 47(1), March, pages 169-200.

[29] Mian, S. L., Smith, C. W., Jr., 1994, Extending trade credit and financing receivables, *Journal of Applied Corporate Finance*, Volume 7, pp 75–84.

[30] Ng, C. K., Smith,J. K. e Smith, R.L, 1999, Evidence on the Determinants of Credit Terms Used in Interfirm Trade, *The Journal of Finance*, 54(3), p. 1109-29

[31] Nilsen, J., 1994, The Impact of Credit Markets on Monetary Policy Princeton University, PhD.

[32] Nilsen, J. 2002, Trade credit and the bank lending channel, *Journal of Money, Credit and Banking*, vol 34, no 1.

[33] Petersen, M.A. e Rajan, R.G., 1997, Trade Credit: Theories and Evidence, *Review of Financial Studies*, Vol. 10, no. 3, p. 661-691.

[34] Petersen, M.A. e Rajan, R.G., 1994, The Benefits of Lending Relationships: Evidence from Small Business Data, *The Journal of Finance*, Volume 49, pp 3–37

[35] Petersen, M.A. e Rajan, R.G., 1995, The Effect of Credit Market Competition on Lending Relationships, *The Quarterly Journal of Economics*, Vol. 110, pp. 407-443.

[36] Piga, G., 2002, Debt and firms' relationships: the Italian evidence, *Review of Industrial Organization* 20, 267-282.

[37] Rajan, R.G. e Zingales, L.,1995, What Do We Know about Capital Structure? Some Evidence from International Data, *The Journal of Finance*, Vol. 50, No. 5.

[38] Schwartz, 1974, An Economic Model of Trade Credit, *Journal of Financial and Quantitative Analysis*, 9(4) September, pages 643-57.

[39] Smith, J. K., 1987, Trade Credit and Informational Asymmetry, *Journal of Finance*, 42(4), September, pages 863-72.

[40] Summers, B., Wilson, N., 2003, Trade Credit and Customer Relationships, *Managerial and Decision Economics*, Volume 24, pp 439–455.

[41] Wilner, B.S., 2000, The Exploitation of Relationships in Financial Distress: The Case of Trade Credit, The Journal of Finance

[42] Yang, S.A., Birge J.R., 2014, How Inventory Is (Should Be) Financed: Trade Credit in Supply Chains with Demand Uncertainty and Costs of Financial Distress, Working Paper Available at SSRN 1734682, 2011 - papers.ssrn.com

Appendix 1: proofs of propositions

Proof of Proposition 1. Let $\overline{SP_t^{TC}}$ be the level of spending power that allows the financing of all profitable investment projects. If $SP_t^{TC} \geq \overline{SP_t^{TC}}$, then $I_t = I_t^*$ and $K_t = K_t^*$. It follows that $\overline{SP_t^{TC}}$ satisfies the condition:

$$I_t^* + \omega L_t + K_t^* = \overline{SP_t^{TC}} + (1-\delta_k)K_{t-1} \tag{58}$$

where: $\overline{SP_t^{TC}} = W_t^F + vK_t + \overline{TC}$. Suppose now that SP_t^{TC} decreases below $\overline{SP_t^{TC}}$. Eq.() cannot be satisfied with equality. If the irreversibility constraint is binding with equality, then $K_t = (1-\delta_k)K_{t-1}$. In this case a reduction of SP_t^{TC} causes a reduction in I_t below I_t^*. The proof of Proposition 2 follows by the fact that the Cobb-Douglas production function implies that $Et\left[\frac{\partial Y_t}{\partial I_t}\right]$ is decreasing and convex in I_t conditional on $K_t = (1-\delta_k)K_{t-1}$. If the irreversibility constraint is not binding, then both I_t and K_t must decrease as SP_t^{TC} decreases below $\overline{SP_t^{TC}}$ because the two factors of production are complementary. This still implies that $Et\left[\frac{\partial Y_t}{\partial I_t}\right]$ is decreasing and convex in SP_t^{TC} because the production function is concave in all the factors. Since, according to Eq. (37), $\frac{\partial SP_t^{TC}}{\partial W_t^F} > 0$ and $\frac{\partial SP_t^{TC}}{\partial TC_t} > 0$, it follows that $Et\left[\frac{\partial Y_t}{\partial I_t}\right]$ is decreasing and convex in W_t^F and TC_t. ■

Proof of Proposition 2. The proof of Proposition 2 is the same as the proof of Proposition 1. ■

Proof of Proposition 3. The proof follows from the fact that the Cobb-Douglas production function is concave in all the factors and from the definition of spending power at time $t+1$, SP_{t+1}^{TC}. Eq. (38) shows that SP_{t+1}^{TC} is positive function of W_{t+1}^F, which, by definition, is negative function of $D \cdot TC_t$. But, if $D \cdot TC_t$ increases, by definition, $TC_t^D = (1-D)TC_t$ decreases. It follows that SP_{t+1}^{TC} is positive function of TC_t^D. Analogously to the proof of Proposition 2, Let $\overline{SP_{t+1}^{TC}}$ the level of spending power at $t+1$ that allows the financing of all profitable investment projects at time $t+1$. If $SP_{t+1}^{TC} \geq \overline{SP_t^{TC}}$, then $I_{t+1} = I_{t+1}^*$ and $K_{t+1} = K_{t+1}^*$. It follows that $\overline{SP_{t+1}^{TC}}$ satisfies the condition:

$$I_{t+1}^* + \omega L_{t+1} + K_{t+1}^* = \overline{SP_{t+1}^{TC}} + (1-\delta_k)K_t \tag{59}$$

where: $SP_{t+1}^{TC} = W_{t+1}^F + b_{t+1} + TC_{t+1} - Liq.Shock$ and $W_{t+1}^F = Y_t - b_t \cdot R - D \cdot TC_t \cdot R_1$. Suppose now that SP_{t+1}^{TC} decreases below $\overline{SP_{t+1}^{TC}}$. It can happen through an increase in $D \cdot TC_t$ (i.e.

a decrease in TC_t^D) which causes a decrease in W_{t+1}^F. Eq.(62) cannot be satisfied with equality. If the irreversibility constraint is binding with equality, then $K_{t+1} = (1-\delta_k)K_t$. In this case a reduction of SP_{t+1}^{TC} causes a reduction in I_{t+1} below I_{t+1}^*. The proof of Proposition 4 follows by the fact that the Cobb-Douglas production function implies that $Et\left[\frac{\partial Y_{t+1}}{\partial I_{t+1}}\right]$ is decreasing and convex in I_{t+1} conditional on $K_{t+1} = (1-\delta_k)K_t$. If the irreversibility constraint is not binding, then both I_{t+1} and K_{t+1} must decrease as SP_{t+1}^{TC} decreases below $\overline{SP_{t+1}^{TC}}$ because the two factors of production are complementary. This still implies that $Et\left[\frac{\partial Y_{t+1}}{\partial I_{t+1}}\right]$ is decreasing and convex in SP_{t+1}^{TC} because the production function is concave in all the factors. But, the cash flow available for the firm at time $t+1$ in order to finance the investment I_{t+1} is increasing in the amount of TC_t^D: in other terms, deferring the repayment of the portion $(1-D)$ of the trade debt TC_t, the firm reacts to the liquidity constraint by increasing the amount of the inventory investment I_{t+1} to a level as close as possible to the optimal level I_{t+1}^*. An additional unit of TC_t^D means an additional unit of available cash flow at $t+1$, and this means an additional unit of I_{t+1} if $I_{t+1} < I_{t+1}^*$. Since, according to Eq. (38), $\frac{\partial SP_{t+1}^{TC}}{\partial W_{t+1}^F} > 0$ and $\frac{\partial SP_t^{TC}}{\partial TC_t^D} > 0$, it follows that $Et\left[\frac{\partial Y_{t+1}}{\partial I_{t+1}}\right]$ and $Et\left[\frac{\partial Y_{t+1}}{\partial TC_t^D}\right]$ are both decreasing and convex in W_{t+1}^F and TC_t^D. ■

Appendix 2: definition of the variables used in the empirical framework

We describe here the variables we are using in the empirical analysis section of the paper:

- $Y_{i,t}$: total revenues realized during year t
- $K_{i,t-1}$: sum of the values of i) plants and equipment, and ii) intangible fixed capital (Software, Advertising, Research and Development). We include in $K_{i,t-1}$ all capital purchased before the end of time t
- $I_{i,t}$: stock of intermediate goods that is used in the production process: raw materials, work-in-process goods and completely finished goods that are considered to be the portion of a business's assets that are ready or will be ready for sale. It is computed as follows: beginning-of-period t input inventories (materials and work in progress), plus new purchases of materials in period t, minus end-of-period t input inventories.

- $L_{i,t}$: the total cost of the labor in year t

- $TC_{i,t}$: Accounts payables (Total Invenories, Raw and Consumable Materials, Work in Progress and Semi-Finished Goods)

- $W^F_{i,t}$: operating profits during period $t-1$ (value of production minus the cost of production inputs) plus net short-term financial assets.

- SP^{TC}_{t+1} : operating profits during period t (value of production minus the cost of production inputs) plus net short-term financial assets at the beginning of period t multiplied by one plus the nominal interest rate (i.e. $W^F_{i,t+1}$), plus short-term debts towards financial institutions, plus Accounts Payable.

- SP_{t+1} : SP^{TC}_{t+1} minus Accounts Payable

Chapter 4

The effects of connectivity and centralization of financial networks on their exposure to systemic risk: a numerical investigation

Mario Eboli, Bulent R. Ozel, Andrea Teglio, Andrea Toto

february 2016

Abstract

This paper studies the effects that two characteristics of the topology of a financial network, namely its degrees of connectivity and of centralization, have on the response of the network to external shocks that can generate phenomena of default contagion. We put forward some conjectures about such effects, conjectures based on some analytic results that shed some light on the exposure to systemic risk of three highly stylised classes of networks: i) complete networks, which are the most connected ones; ii) circular networks (also known as 'wheels'), which are the least connected and least centralised networks; and iii) the star-shaped networks, which are the least connected and most centralised networks. We conjecture that the more a network is connected, the more it displays a robust-yet-fragile nature, in the sense that it is completely resilient to relatively small shocks but it is exposed to the risk of a total melt down (the default of all agents in the network) if it is hit by large enough shocks. We also conjecture that the centralization of a financial network has the same effects: the more centralised a network is, the more it is robust-yet-fragile. Conversely, we conjecture that a sparse and decentralised network, likewise the circular networks, has the opposite feature: the more sparse and decentralised a network is, the more it displays a vulnerable-yet-resilient nature, in the sense that it is exposed to episodes of local contagion due top relatively small shocks, while it faces a rather small risk of complete contagion. The numerical simulations that we run confirm these conjectures.

Keywords: systemic risk, financial contagion, financial networks

1 Introduction, motivation and related literature

The exposure of interbank and financial networks to the risk of contagion has become a major concern for authorities as well as for academic economists. A large and growing stream of literature investigates the response of banking and financial networks to shocks, with the aim of understanding the structural features that determine the contagiousness of a financial networks, i.e. its exposure to systemic risk.

Eisenberg and Noe (2001) is the seminal contribution which has provided the analytical basis and the computational tool to many authors (see the below cited papers) who perform numerical simulations to study direct contagion. As well as their paper, the present one studies the properties of the same object: a directed and weighted graph that represents a financial system. Differently from the present one, their paper resorts to a different analytical approach: Eisenberg and Noe resort to matrix algebra and lattice theory. These authors investigate the domino effect generated by the default of agents that participate in a single payment system. In so doing, they study the existence and the uniqueness of a vector of payments that clears a network of interdependent financial claims, where the capability of an agent to repay in full his debts depends on the solvency of his own debtors which, in turn, depends on the solvency of their debtors, and so forth. They express such a vector as a function of the operating cash flows of the members of the financial network. This function is defined on a lattice, representing such a financial system, and complies with the requirements of limited liability, debt priority and pro-rata reimbursements.

Eisenberg and Noe, as well as the present paper, do not investigate agents' behaviour in a financial network and focus on the mechanics of contagion as governed by the rules of limited liability, debt priority and pro-rata reimbursements.

This marks a major difference with respect to theoretical analyses of direct financial contagion that take explicitly into account the behaviour of banks and depositors. In Allen and Gale (1998), the initial failure of one or more banks, capable of generating a widespread financial crises, can be due to exogenous causes. Financial crises arise as a consequence of downturns in the economic cycle. Recessions can cause losses in the value of the assets held by banks, these losses are capable of rendering them insolvent. If depositors foresee the recession, they will protect themselves from possible bank defaults by withdrawing their deposits and, in so doing, they create the conditions for the occurrence of a widespread crisis. In Allen and Gale (2000) the failure of a bank is due to an idiosyncratic shortage of liquidity that forces the bank to liquidate long-term assets, incurring the costs of such fire sales. They show that a complete network - a network where all banks are equal to one another, all having mutual bilateral obligations and of the same amount - is more robust than an incomplete network, i.e., a network with fewer links among the banks.

Concernings the literature related on simulation techniques to study financial contagion, Upper (2007, page 2 and 3) says: ”Unfortunately, analytical results on the relationship between market structure and contagion have been obtained only for a limited number of highly stylized structures of interbank markets, which are of limited use when it comes to assessing the scope for contagion in real world banking systems.[...] Given the scarcity of theoretical results, researchers have increasingly turned to computer simulations to study contagion.” Upper refers to several authors who, in order to assess the robustness of different network structures, have studied the mechanics of default contagion using numerical simulations, foregoing the microeconomic behaviour of banks and depositors. Such papers - which includes the works by Sheldon and Maurer (1998), Furfine (2003), Wells (2002), Elsinger, Lehar and Summer (2006), Upper and Worms (2004), Degryse and Nguyen (2004), Blavarg and Nimander (2002), Cifuentes (2003), Mistrulli (2005,2006), Canedo and Martínez Jaramillo (2009) - have analyzed national banking systems, in most cases estimating the structure of national interbank networks, using simulations to evaluate their exposure to default contagion.

Numerical simulations are also used by Shin et al. (2005) and Nier et al. (2007), who analyze generic network structures, rather than specific national ones. Shin et al. present a model where default contagion is exacerbated by the effects of fire sales. Nier et al., using a computing device, generate random banking networks, in the fashion of the random graphs a là Erdos-Rényi, and use them to run numerical simulations aiming at evaluating the exposure to systemic risk of different network structures.

Recent analytic results (Acemoglu et al. 2013, 2015; Eboli 2013, 2016) have shown that complete networks, i.e. networks where everybody is connected to everybody else, confirm the conjecture by Haldane (2009) that highly dense interbank networks have a ‘robust-yet-fragile’ nature:

“In a nutshell, interconnected networks exhibit a knife-edge, or tipping point, property. Within a certain range, connections serve as a shock-absorber. The system acts as a mutual insurance device with disturbances dispersed and dissipated. Connectivity engenders robustness. Risk-sharing – diversification – prevails. But beyond a certain range, the system can flip the wrong side of the knife-edge. Interconnections serve as shock-amplifiers, not dampeners, as losses cascade. The system acts not as a mutual insurance device but as a mutual incendiary device. Risk-spreading – fragility - prevails." Eboli (2013) shows that star-shaped networks display the same feature: the bank at the center of such a kind of network acts as a hub, distributing losses evenly among the other members of the network, in case of a crisis. Acemoglu et al. (2013) challenge the conclusion of Allen and Gale (2000), that the complete network structure is the most robust. As the authors put it: "One of our main results is that as the magnitude or the number of negative shocks cross certain thresholds, the types of financial networks that are most prone to contagious failures change dramatically. In particular, more financial interconnections are no longer a guarantee for stability. Rather, in the presence of large shocks, interbank liabilities facilitate financial contagion and create a more fragile system. Our results show that, in the presence of large shocks, weakly connected financial networks - for example, one consisting of a collection of pairwise connected banks with only a minimal amount of shared assets and liabilities with the rest of the system - are significantly less fragile than the more complete networks. [Acemoglu et al. (2013), pages 2 and 3].

Conversely, Acemoglu et al.(2013) and Eboli (2016) prove that circular networks, where each agent is connected to just one or two neighbours, forming a cycle, behave in the opposite way: they are vulnerable with respect to episodes of local contagion, caused by relatively small shocks, while they are less exposed than complete and star networks to the risk of a complete system melt down. These analytic results describe the behaviour of highly stylized examples of financial networks:[1] the complete network is the most densely connected network, the star is the most centralised and sparse network while the circle is the most sparse and decentralised network. We believe that the effects that connectivity and centralization have on such stylized financial networks are also present in generic networks, according to their degrees of connectivity and centralization. We conjecture that the more a financial network is a) densely connected or b) sparse and highly centralised, the more the network has a robust-yet-fragile nature, likewise the complete and the star-shaped networks; while the more a network is sparse and decentralised – likewise the circular

[1]Eboli (2013, 2015) analyse the behaviour of complete, star-shaped and unilateral circular networks, while Acemoglu et al.(2013, 2015) focus on complete and bilateral circular networks.

networks – the more it displays a vulnerable-yet-resilient behaviour. The purpose of the present work is to test such conjectures by means of numerical simulations. To this end, we look at the response of financial networks to external shocks in terms of *first* and *last thresholds* of default contagion. The first threshold of contagion is the value of the smallest external shock capable of causing a secondary default. i.e. the default of an agent due to the losses received by its neighbours. The last contagion threshold is the value of the smallest shock capable of causing the default of all agents in the network. The more the first and the final thresholds of contagion in a network are close to one another, the more the network is 'robust-yet-fragile'. Conversely, the more apart are these thresholds in a network, the more the network displays a 'fragile-yet-resilient' nature. More specifically, we conjecture that, ceteris paribus: i) the more dense a financial network, the closer its first and its final contagion threshold; ii) the more centralised a financial network, the closer its first and its final contagion thresholds; and iii) the more sparse and decentralised a financial network, the larger the gap between its first and its final contagion thresholds. We test these conjectures generating random networks with various degrees of density and centralization and perturbating them with the shocks of increasing value, shocks that diminish the value of the external assets of the banks in the networks in a random fashion.

The paper is organised as follows. In the next section, we introduce the network model used in this paper and describe the above mentioned analytic results by Eboli (2013, 2016), results concerning the response to shocks of complete, star-shaped and circular financial networks. On the basis of these analytic results, in section three we put forward some conjectures about the relation between the degree of connectivity and the degree of centralization of financial networks, on one hand, and their exposure to episodes of default contagion, on the other hand. In the same section we present the results of the numerical simulations that we run to test such conjectures. Conclusions are drawn in section four.

2 Complete, star and circular networks

In this section, as well as in the numerical simulations that we run, we model an *interbank network* as a connected, directed and weighted graph $N := (\Omega, \Lambda)$, where the node $\omega_i (i = 1, 2, ..., n)$ in Ω represents a bank and the links in $\Lambda \subseteq \Omega^2$ represent the interbank deposits that connect the members of Ω among themselves. The liabilities of a bank ω_i in Ω comprise customers (households) deposits, h_i, and interbank deposits, d_i, and its own equity e_i. On the asset side, a bank ω_i holds long-term assets, a_i, which are liabilities of agents that do not belong to Ω, and short-term assets c_i, which are deposits made by bank ω_i in other banks of the network. The budget identity of a bank is: $a_i + c_i = h_i + d_i + e_i$. A link $l_{ij} \in \Lambda$ represents the interbank obligations, and its direction goes from the debtor node ω_i to the creditor node ω_j. The weight of the link l_{ij} is equal to the amount of money c_{ji} that bank ω_j has deposited in bank ω_i. For simplicity, we assume that all interbank deposits are reciprocal, i.e. $l_{ij} = l_{ji}$ for all $l_{ij} \in \Lambda$.

In a *complete* interbank network, each bank places a deposit in every other bank: $\Lambda^c = \{l_{ij} | i \neq j; i, j = 1, ..., n\}$. Let $N^c = \{\Omega, \Lambda^c\}$ be a complete interbank network where all the links in Λ^s have the same capacity c_{ij}, and $\sum_j c_{ij} = (n-1)c_{ij} = c_i$. In other words, each bank in N^c evenly allocates its own interbank deposits c_i among all other banks in the network. Eboli (2013) shows that in a complete network N^c the first threshold and the final threshold of contagion coincide and are equal to

$$\tau^c = ne_i + (n-1)e_i \frac{h_i}{d_i} = E\left(1 + \frac{1}{\phi}\right) - \frac{E}{n}\frac{1}{\phi} \tag{1}$$

where $E = \sum_{i=1}^{n} e_i$ is the total equity of the banks in Ω, and $\phi = d_i/h_i$ is the ratio between the interbank debt and the external debt of a bank.

This result shows that the complete network, on one hand, is entirely resilient to relatively small shocks, i.e. faces no defaults for shocks smaller than τ^c. On the other hand, for large enough shocks – larger than or equal to τ^c – this network induces a complete system melt down. The same applies to the star-shaped network, as shown below.

A *star-shaped* interbank network consists of a central node, ω_c, that places a deposit in each of the remaining $n-1$ peripheral banks which, in turn, place their deposits in ω_c and exchange no deposits among themselves. Let $N^s = \{\Omega, \Lambda^s\}$ be a star-shaped interbank network that complies with the above assumptions, i.e. $\Lambda^s = \{l_{ic}, l_{ci} | i \in \Omega \backslash \omega_c\}$, and where each link in Λ^s has a capacity equal to c_s. That is, c_s is the amount deposited by each peripheral bank in ω_c and is the amount that the central node deposits in each of the peripheral ones. In the star-shaped network the two thresholds of contagion may coincide or not, depending on whether the central node is in the set of primary defaults or not. If the central node ω_c is in the set of primary defaults, i.e. if $\Delta = \{\omega_c, \omega_p | \text{for some } p \in \Omega\}$, then in a star-shaped network N^s the first threshold of contagion τ_1^s and the final threshold of contagion τ_f^s coincide, $\tau_1^s = \tau_f^s = \tau^s$, and we have

$$\tau^s = E + ne_p \frac{1}{\phi} - e_p \alpha = E\left(1 + \frac{1}{\phi}\right) - \frac{E}{n}\alpha \tag{2}$$

where $\alpha = \frac{\phi - \varepsilon(1+2\phi)}{\phi^2 \varepsilon(1+\phi)}$.

If the central node ω_c is not in the set of primary defaults, the first and the final thresholds of contagion of the star-shaped network do not coincide. In this case they are, respectively, equal to

$$\tau_1^s = me_p + e_c \left(1 + \frac{1}{\phi}\right) \tag{3}$$

and

$$\tau_2^s = E\left(1 + \frac{1}{\phi}\right) + \frac{E}{n}\frac{\phi + 1}{\phi^2}, \tag{4}$$

where m is the minimum number of peripheral primary defaults which is sufficient to induce the default of the central node.[2]

Note that, for arbitrarily large values of n, which is the cardinality of the set of nodes Ω in the network, the above contagion thresholds τ^c, τ^s and τ_2^s become arbitrarily close to

$$\tau = E\left(1 + \frac{1}{\phi}\right). \tag{5}$$

The response to external shocks of a circular network is rather different from the one described above. Formally, a cycle-shaped financial network $N^o = \left\{\Omega, A, T, H, L^\Omega, L^A, L^T, L^H, \Gamma\right\}$ is such that $L^\Omega = \{l_{ij} | i = 1, 2, ..., n;\ j = i + 1 \text{ for } i = 1, ..., n-1, \text{ and } j = 1 \text{ for } i = n\}$. Acemoglu et al.(2015) and Eboli (2016) find that a circular network is 'vulnerable yet resilient' in as much as it is exposed to episodes of local contagion when hit by relatively small shocks (while the star and the complete are not), and is less exposed to the risk of a complete melt down because, for sufficiently large number n of banks in the network, there is always a positive probability that at a least a bank survives even in case of large shocks. Eboli (2016) shows that in the unilateral circular network, also known as the wheel, the bankruptcy of a bank –in the worse possible scenario, where the bank looses all of its assets – induces the defaults of at the most k successive banks along the wheel, where k is the smallest integer such that

$$\frac{\left(\frac{\phi}{1+\phi}\right)^k}{(1+\phi)\left[1 - \left(\frac{\phi}{1+\phi}\right)^k\right]} > \frac{\varepsilon}{1 - \varepsilon}$$

It follows that the final threshold of contagion in the wheel, i.e. the default of all banks, can be achieved by relatively small external shocks (smaller than τ), as long as these

[2]Eboli (2013) shows that $m \geq \frac{\varepsilon + \varepsilon\phi(n-1))}{1-\varepsilon}\frac{1+\phi}{\phi} = n\left[\frac{\varepsilon(1+\phi)}{1-\varepsilon}\right] + \frac{\varepsilon(1-\phi^2)}{\phi(1-\varepsilon)}$.

shocks are evenly distributed among k/n banks that are located at a regular distance k from one another. However, the probability that such a specific allocation of an external shock occurs is sharply decreasing in the number of banks in the wheel. For example, for the ranges of parameters $\varepsilon \in (0.06, 0.12)$ and $\phi \in (0.1, 0.6)$, k is at the most equal to 2. Within this plausible range of values of ε and ϕ, for $n > 4$, and for a number of primary defaults $m < n - 2$ (that is, for external shocks that spare just two banks out of n) there is always a positive probability that at least one bank survives (while for large enough m, i.e. for shocks larger than τ, there are no survivors at all in the star and in the complete networks).

As mentioned above, on the basis of these analytic results we make the following conjectures:

1. in the class on networks with the same degree of connectivity, the more a network is centralised, the closer its first and final thresholds of contagion. In our numerical simulations, we expect that moving from high centralised networks to progressively less centralised networks, keeping connectivity constant, the gap between the first and the final thresholds increases.

2. in the class on networks with the same degree of centralization, the more dense is a network, the closer are its first and final thresholds of contagion. In our numerical simulations, we expect that the gap between the first and the final thresholds increases as we move from the highly dense networks to progressively more sparse networks, keeping centralization constant.

In order to bring these properties in the foreground in a neat fashion, we first present simulations on the effects of centralization keeping connectivity at its minimum. That is, we track the behaviour of the first and final contagion thresholds as we progressively transform circular interbank networks into star-shaped networks. We then present simulations on the effects of connectivity keeping centralization at its minimum. That is, we explore the response to shocks of the class of regular networks, starting from the least dense one, the circle, and progressively moving towards the most dense one, the complete network. Finally, we investigate numerically the behaviour of networks with varying degrees of centralization and connectivity.

3 The simulations

In this section we present the results of several numerical simulations that describe the behavior of different network configurations when subjected to exogenous shocks, i.e. the resilience of different types of financial networks in a stress scenario. We devise a simulation engine that (a) is able to generate interbank networks according to the network shapes that are under investigation in this paper and that (b) allows us to study the resilience of the realised banking system to shocks and how resilience to shocks depends on the key parameters of the system. The aim is to investigate the effects of connectivity and centralization on the resilience of the network in term of their exposure to systemic risk. In doing this, we simulate a contagion process triggered by exogenous shocks over different types of network and analyze the behavior of the different networks in terms of local and global resilience, robustness and fragility. The starting point configuration is a circular graph, unilateral or bilateral, with zero centralization and minimum level of connectivity; the final configurations are the complete network and the star shaped network.

We think of a banking system as a network of nodes, where each node represents a bank and each link represents a directional lending relationship between two nodes. Importantly, the banking system needs to obey bank-level as well as aggregate balance sheet restrictions.

For every simulation framework, the network is composed of 64 banks (nodes); each bank is characterized by its own balance sheet. On the asset side, let a_i be the value of the sum of *external assets* owned by the bank i. Besides the external assets, the bank i can hold *internal assets* which are liabilities of other banks in Ω, and let c_i be the sum of the such assets held by agent i. On the liability side of the balance sheet, let d_i be the sum of the debts that the bank i owes to other agents in Ω, while h_i is the *external debt* of i, i.e., the amount of debt claims against i held by households, and $\sum_j d_{ij}$ is the *internal debt* of agent i, i.e., the claims against i held by other banks of Ω. For simplicity, we assume that all debts have the same seniority. Finally, the value of the equity of the i-th agent, e_i, is set by the budget identity $e_i \equiv a_i + c_i - d_i - h_i$. We assume that the value of the external assets is set by the market and take the other balance sheet headings c_i, d_i, h_i, as well as the debts d_{ij}, at their book values. For the sake of simplicity, we also assume that all the shares issued by the members of Ω are held by households, i.e., there is no cross-holding of shares among the financial intermediaries.

The key parameters to set the initial balance sheet entries are: i) the ratio between intra-network debt and external debt, $\phi = \frac{d}{h}$, and the capitalization ratio, $\varepsilon = \frac{e}{(a+c)}$. For all the experiments, the capitalization ratio, ε, is invariant and fixed equal to 0.1. Differently, we consider three different values for the parameter ϕ: 0.3, 0.4, 0.5.

In the connectivity experiment, the simulation engine generates the initial network configuration, the unilateral circle (also know as the wheel) with connectivity equal to one and zero centralization, where all the banks have the same balance sheet configuration. Then, in 63 steps, the engine increases the connectivity of the initial network, creating 63 different network configurations, each one with a different higher degree of connectivity. The final network configuration is a complete network where the bank i is connected with all the other $n-1$ banks. In other terms, the engine starts from a sparse decentralized network and move through more dense network configurations, keeping invariant the zero-centralization level, until reaching the complete network.

In the centralization experiment, the starting point configuration is the bilateral circle, where each node is connected with two neighbours and the centralization level is zero. In order to move towards more centralized configurations, the engine randomly designates a candidate center node and, step by step, rewires each of the other nodes to the designated center. During the entire experiment, the simulation engine creates 64 different more centralized network shapes; the final configuration is a perfect star with a single node in the center and 63 pendant nodes.

In both the simulation frameworks, for each network configuration, the simulation engine perturbs the system with sequential idiosyncratic shocks that hit each of nodes in a random fashion. In order to obtain a sufficiently large sample of different shock scenarios, the engine produces, for each network configuration, 100 different round of idiosyncratic shocks. The target of the exogenous shock is the external asset a: the amount of the loss is the total share of a and it is partially or totally absorbed by the amount of the equity e. If the total loss is smaller than e, the shock is totally absorbed and the node is solvent; conversely, we are in presence of a primary default. The defaulted bank transmits losses to its creditors and customers: upon the occurrence of a shock, the propagation of the losses across the system is governed by the rules of limited liability, debt priority and pro-rata reimbursement of creditors.

The graphic representations of the different experiments show, for each network configuration, two boxplots that depict graphically the distribution of two contagion thresholds over 100 shock simulations, by displaying the median value and their quartiles[3].

In Figure 1 and Figure 2 we analyze the effects of an increase in connectivity on the exposure of the network to systemic risk. The starting point is a wheel network, i.e. the unilateral circle network, where every node is connected with only one neighbour, the connectivity is 1 and the centralization is zero and constant for all the subsequent configurations. The final configuration, i.e. where the connectivity is maximum, is a complete network. The difference between the two figures lies purely in the graphic style. On the horizontal axis we have the different degrees of connectivity, from 1 to $n-1$; on the vertical axis we have the amount of exogenous shocks that hit the network. T1 is the first threshold, i.e. the minimum amount of the exogenous shock that is sufficient to yield one secondary default[4]. Tfin is the final threshold, i.e. the minimum amount of exogenous shock that is sufficient to yield the complete meltdown of the system.

The path of the two thresholds show the effects of connectivity of the interbank network on its exposure to systemic risk. The two paths are clearly convergent in a quasi-monotonic fashion. For low levels of connectivity, there is a distance between the two thresholds. The small values of T1 show that a relatively small exogenous shock is sufficient to trigger a secondary default: this means that sparse networks are vulnerable with respect to episodes of local contagion when hit by relatively small shocks. Conversely, the high values of Tfin demonstrate that sparse networks are resilient with respect to the risk of a complete meltdown of the system because, for a sufficiently large number of agents, there is always a positive probability that at least a bank survives even in case of large shocks. For high level of connectivity, the two thresholds converge towards the same values and, in the case of a the complete network, they are coincident. This shows that dense networks behave in the opposite way: they are resilient to relatively small shocks, i.e. they exhibit no secondary defaults for shocks smaller than a quite large threshold, but they are vulnerable for large enough exogenous shock, i.e. they are exposed to the complete failure of the system when hit by shocks larger than a quite small threshold. Furthermore, the complete network exhibits a perfect coincidence of the first and the final thresholds, and this constitutes a confirmation of one of the analytical results of the model. The convergence value of the two thresholds in the final configuration network (the complete network) confirms perfectly the analytical result in Eq.(5): substituting the total equity present in the network and

[3]The bottom and top of the box are always the first and third quartiles

[4]A secondary default is a default due to losses transmitted from other defaulted nodes, i.e. not directly caused by an exogenous idiosyncratic shock.

the initial set value of ϕ, the final threshold is exactly calculated by the analytical formula. In summary, as conjectured in the previous section, in the class of networks with the same degree of centralization, the more dense is a network, the closer are its first and final thresholds of contagion. The numerical simulations confirm this conjecture: the first threshold decreases and the final threshold increases as we move from the highly dense networks to progressively more sparse networks, keeping the degree of centralization constant and fixed at the minimum value.

Figures 3, 4, 5 are related to simulations on the effects of centralization keeping constant connectivity at its minimum. The simulation framework is the same for the three experiments. The initial configuration is a circular network with bilateral links, i.e. a decentralised network. Then, step by step, the engine creates 64 more centralized configurations, until reaching a star-shaped network, with one center node, i.e. the most centralized network configuration. In this framework, the higher the grade of centralization, the higher the aggregate amount of the interbank lending, while the balance sheet entries of the sole peripheral (pendant) nodes are kept constant during the entire experiment. Conversely, the center node increases the amount of its intra-network debt, keeping constant the other balance sheet entries, so becoming progressively more exposed to peripheral shocks. In other terms, while the d/h ratio of the peripheral nodes is constant, the d/h ratio of the central node increases until reaching its maximum value in the final configuration. Toghether with the the stock of equity e, the d/h ratio between internal and external debt is the only heading of the balance sheets of the agents that determines the contagion thresholds of the above analysed networks. Moreover, all the above characterized thresholds are increasing in the equity endowments, e, and decreasing in the d/h ratio. The protective role played by the equity stock is not surprising: the larger the equity of the members of a network, the larger the amount of losses that can be absorbed by those agents, the higher the contagion thresholds of the network (and, of course, the smaller the set of defaults induced by any given shock). The relevance of the d/h ratio lies in the fact that this ratio governs the allocation of the flow of losses, released by defaulting nodes, between external creditors (households) and internal ones (other nodes in Ω). The larger this ratio between internal and external debt, the smaller the portion of losses that, at each default, is sent into the sink H, and the larger the flow of losses that continues to circulate among the nodes in Ω, and vice versa. Therefore, the larger the d/h ratio: i) the larger the portion of an external shock that overflows from the primary defaults towards the rest of the network; ii) the smaller the smallest shocks capable of causing secondary defaults (the contagion thresholds), and iii) the larger the number of defaults induced by a shock. This is evident examining the Figures 3, 4, 5. The variant parameter of the three experiments is the d/h ratio between internal and external debt, relatively to all the peripheral nodes, that varies from 0.3 to 0.5. For higher values of the ratio d/h, ceteris paribus, the paths of both the thresholds are characterized by lower values. It's important to note that, together with the final threshold, the other threshold under examination here is not T1, but T2, i.e. the minimum amount of the exogenous shock that is sufficient to yield two secondary defaults. This is due to the fact that here the initial network configuration is a circular one with bilateral links: under this condition, for $\phi \in [0.3; 0.5]$, a primary default on a peripheral node yields at least two secondary defaults. This first experiment on the effect of centralization demonstrate that, basically, an increase in the degree of centralization, keeping constant the degree of connectivity, has the same effect of an increase in connectivity, keeping constant the degree of centralization, i.e. the convergence of the two fundamental thresholds (the second and the final threshold, in this case). For low levels of centralization, there is a big distance between the two thresholds. The small values of T2 show that a relatively small exogenous shock is sufficient to trigger two secondary defaults: this means that decentralised networks are vulnerable with respect to episodes of local contagion when hit by relatively small shocks. Conversely, the high values of Tfin demonstrate that decentralised networks are resilient with respect to the risk of a complete meltdown of the system. For high level of centralization, the two thresholds converge towards the same values and, in the case of a the star-shaped network, they are coincident. This shows that highly centralised networks behave as well as highly dense networks: they are entirely resilient to relatively small shocks, i.e. they exhibit no secondary defaults for shocks smaller than a quite large threshold, but they are vulnerable for large enough exogenous shock, i.e. they are exposed to the complete failure of the system when hit by shocks larger than a quite small threshold. In other terms, they have a have a robust-yet-fragile nature. Furthermore, the star-shaped network exhibits a perfect coincidence of the first and the final thresholds, and this, as well as in the case of the complete network, constitutes a confirmation of one of the analytical results of the model. As conjectured in the previous section, in the class of networks with the same degree of connectivity, the more a network is centralised, the closer its second and final thresholds of contagion. The numerical simulations confirm this conjecture: the second threshold decreases and the final thresholds increases as we move from high centralised networks to progressively less centralised networks, keeping connectivity constant.

Figures 6, 7, 8 show the result of the second simulation framework on the effect of centralization. Differently from the previous experiment on centralization, here the balance sheet entries of both the pendant nodes and the designated central node are not constant during the entire centralization process. The invariant parameters in this framework are the total interbank lending and the initial set value of ϕ, which is constant and fixed to, respectively, 0.3, 0.4 and 0.5 for every single node. This means that the central node, keeping constant the value of ϕ, as well as the other peripheral nodes, preserves the same robustness and the same exposure to external shocks of the peripheral nodes. As well as in the first experiment on the effects of centralization, the second experiment confirm that, keeping constant the connectivity, the more a network is centralised, the closer its second and final thresholds of contagion. Furthermore, decentralised networks exhibit a vulnerable-yet-resilient behavior, while highly centralised networks show a robust-yet-fragile nature.

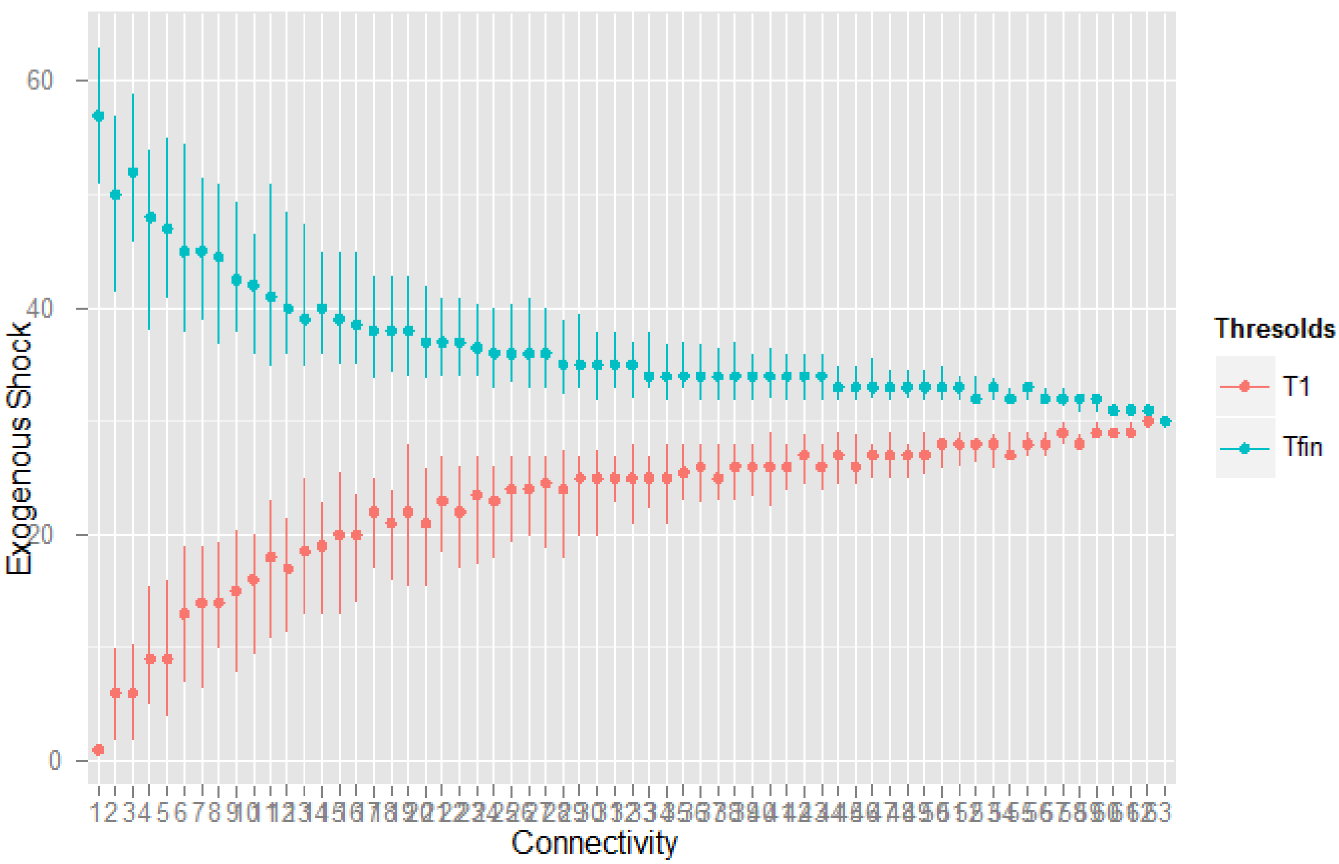

Figure 1: Threshold T1 and Final Threshold as function of connectivity; $\phi = 0.4$, $\varepsilon = 0.1$

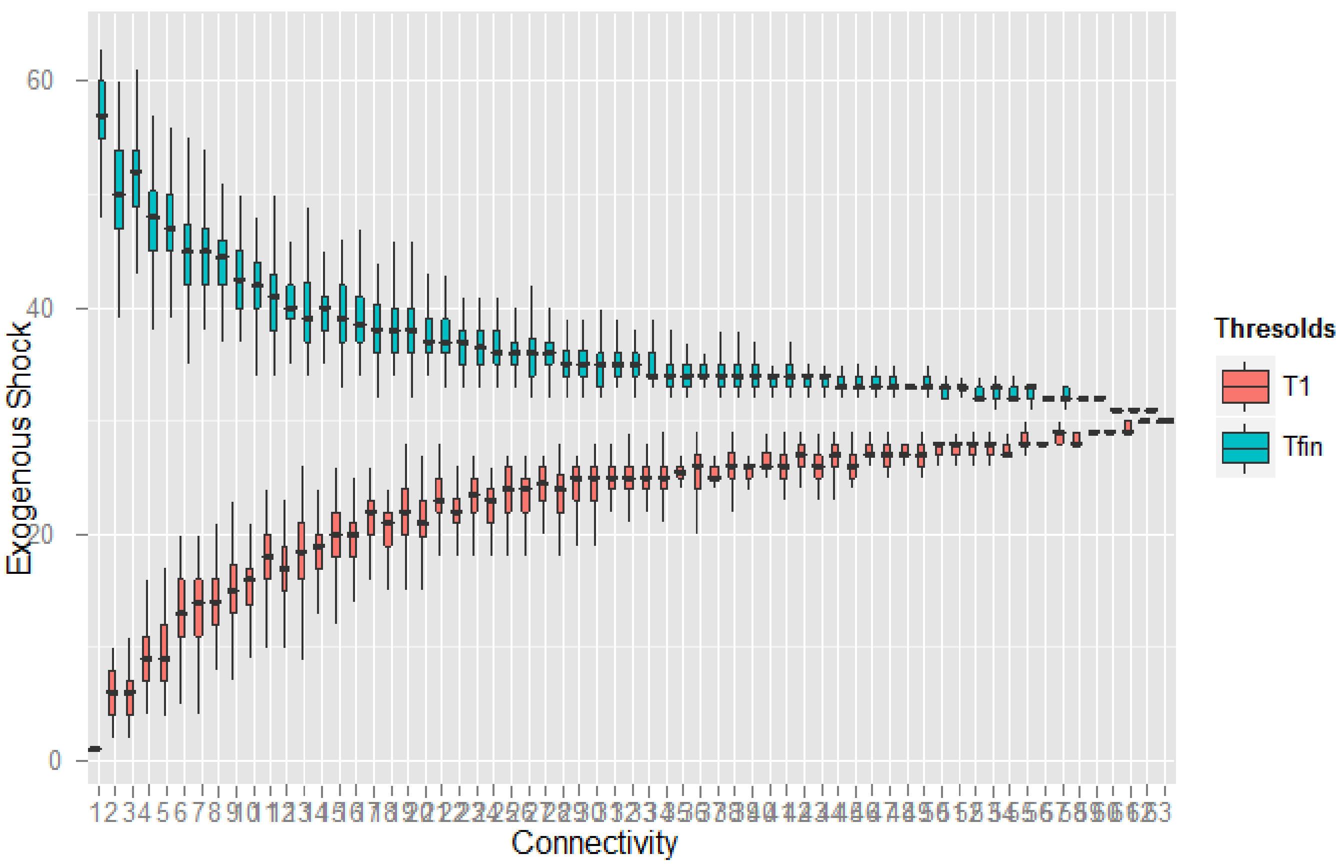

Figure 2: Threshold T1 and Final Threshold as function of connectivity; $\phi = 0.4$, $\varepsilon = 0.1$

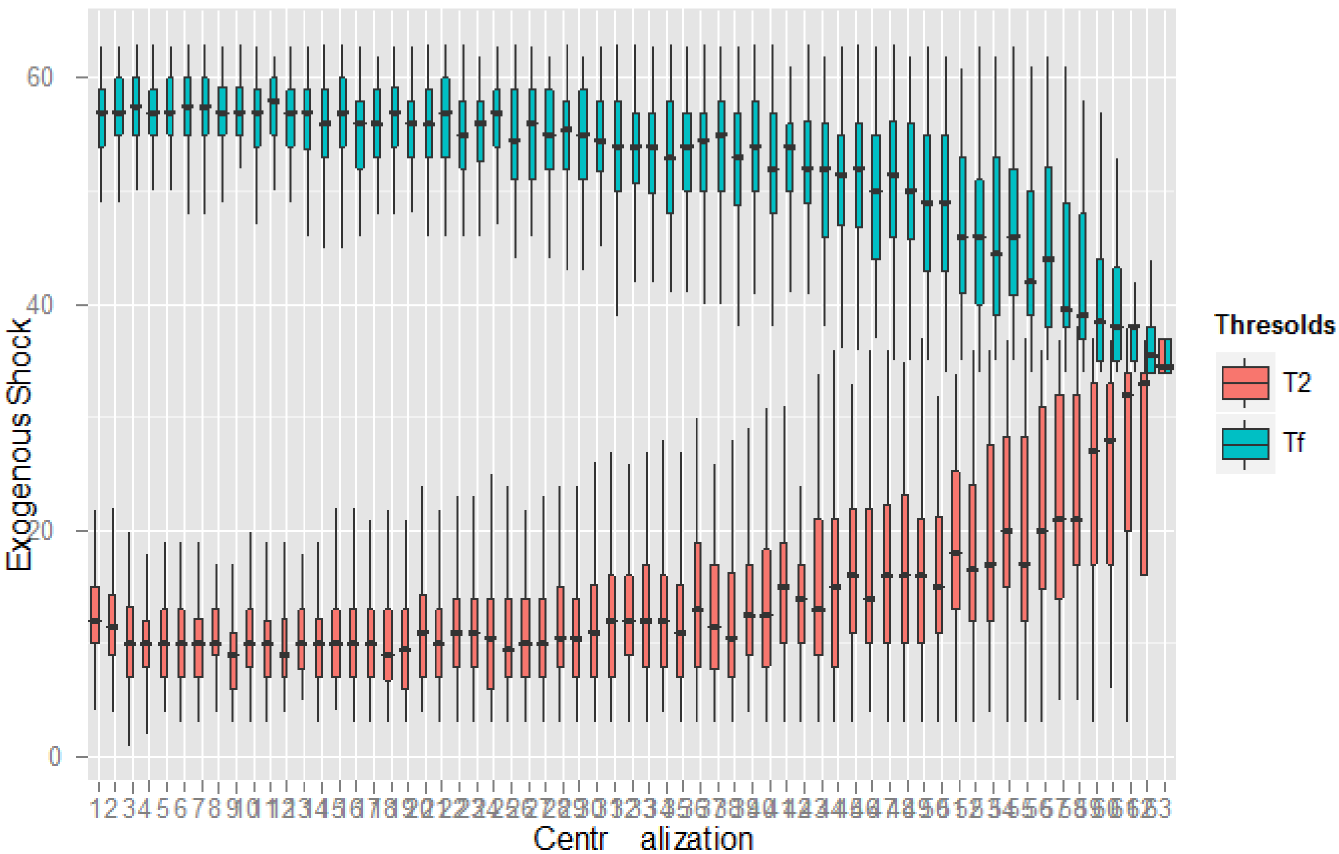

Figure 3: Threshold T2 and Final Threshold as function of centralization; $\phi = 0.3$, $\varepsilon = 0.1$, constant for all nodes but the center

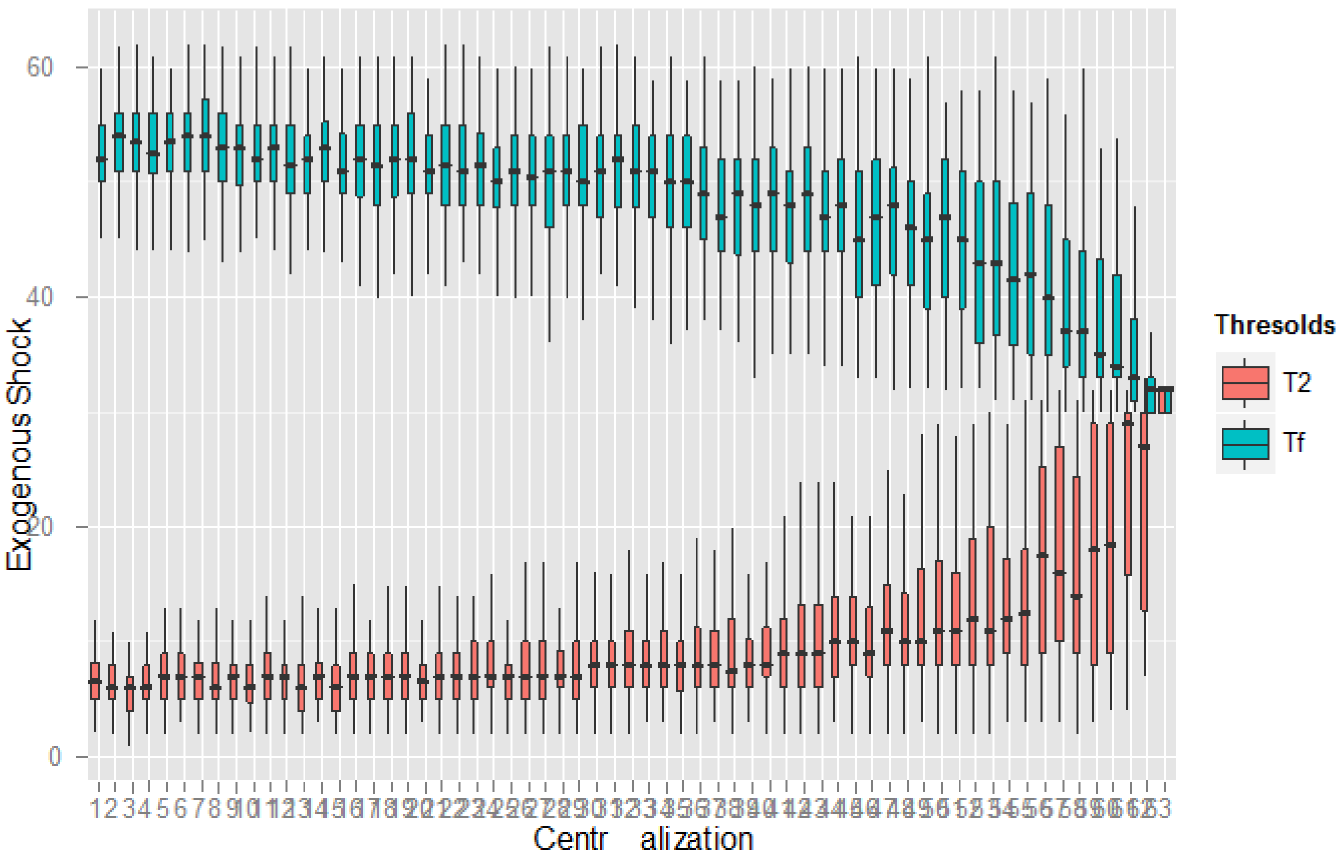

Figure 4: Threshold T2 and Final Threshold as function of centralization; $\phi = 0.4$, $\varepsilon = 0.1$, constant for all nodes but the center

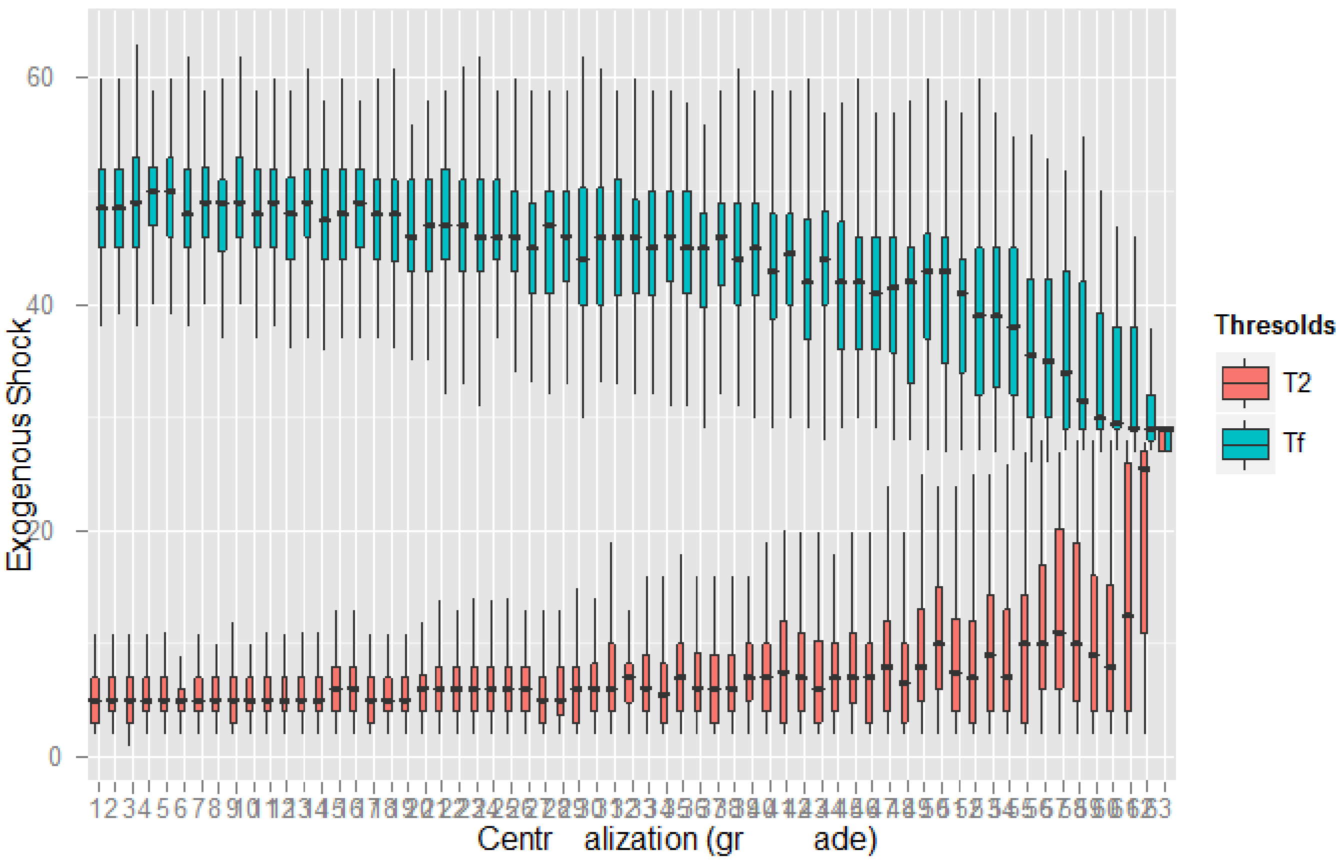

Figure 5: Threshold T2 and Final Threshold as function of centralization; $\phi = 0.5$, $\varepsilon = 0.1$, constant for all nodes but the center

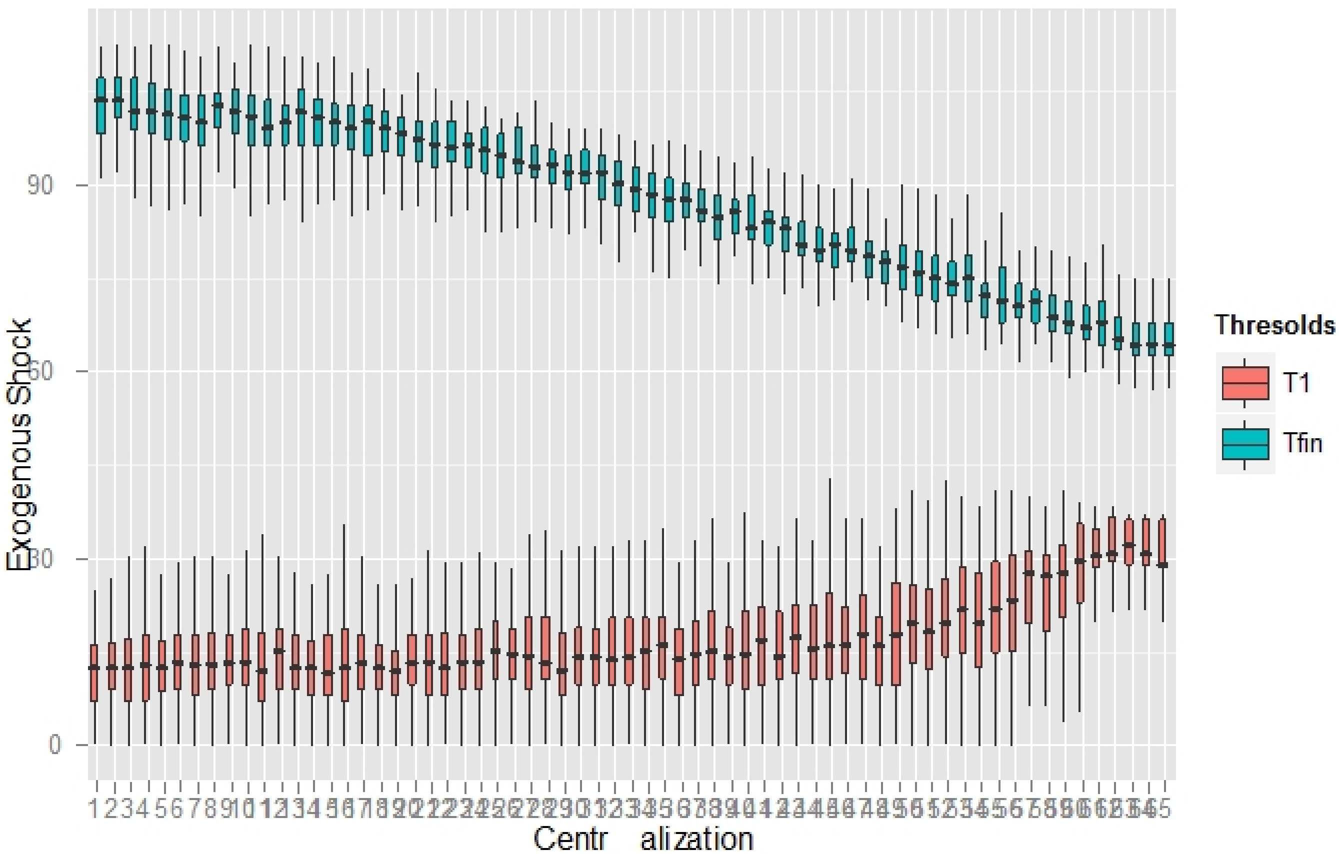

Figure 6: Threshold T2 and Final Threshold as function of centralization; $\phi = 0.3$, $\varepsilon = 0.1$, constant for all nodes

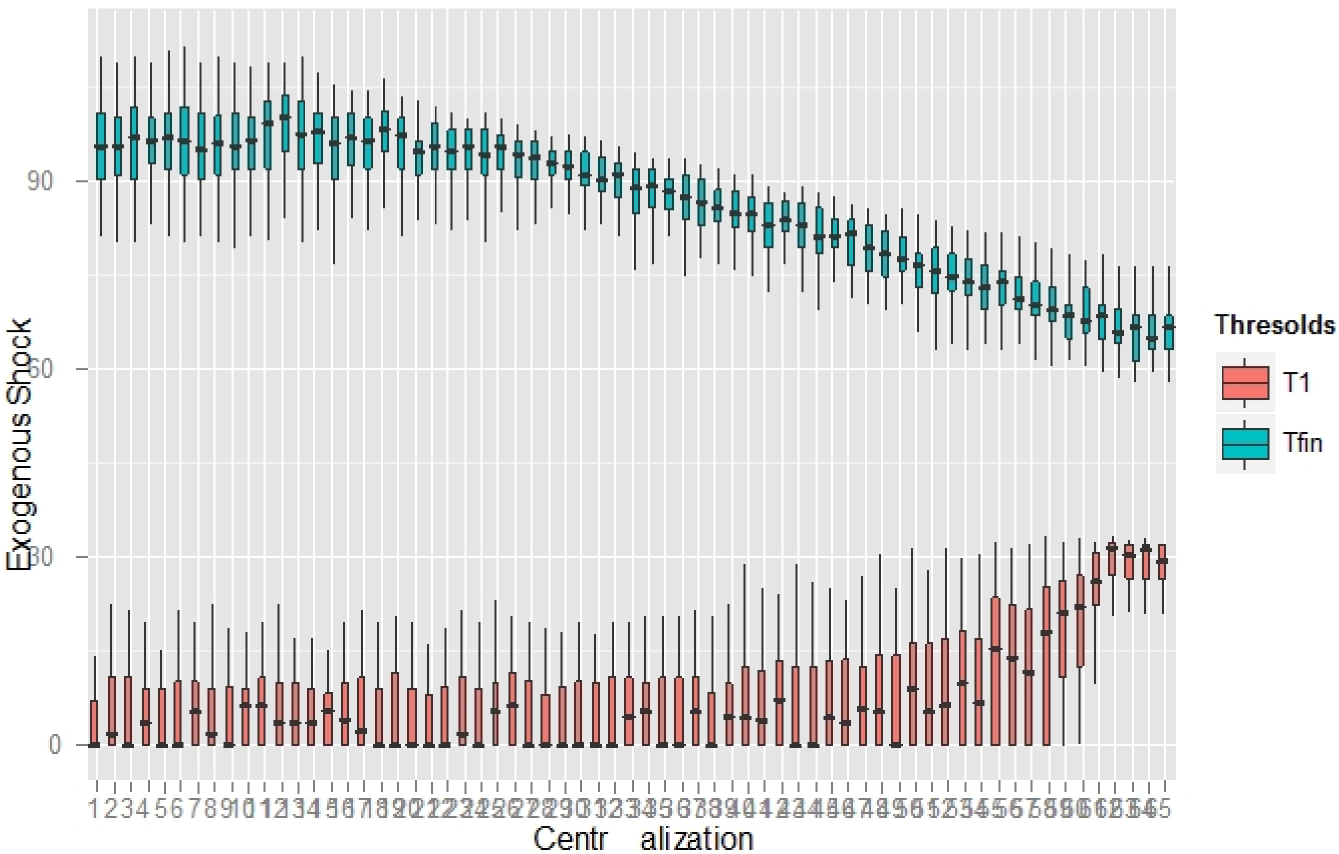

Figure 7: Threshold T2 and Final Threshold as function of centralization; $\phi = 0.4$, $\varepsilon = 0.1$, constant for all nodes

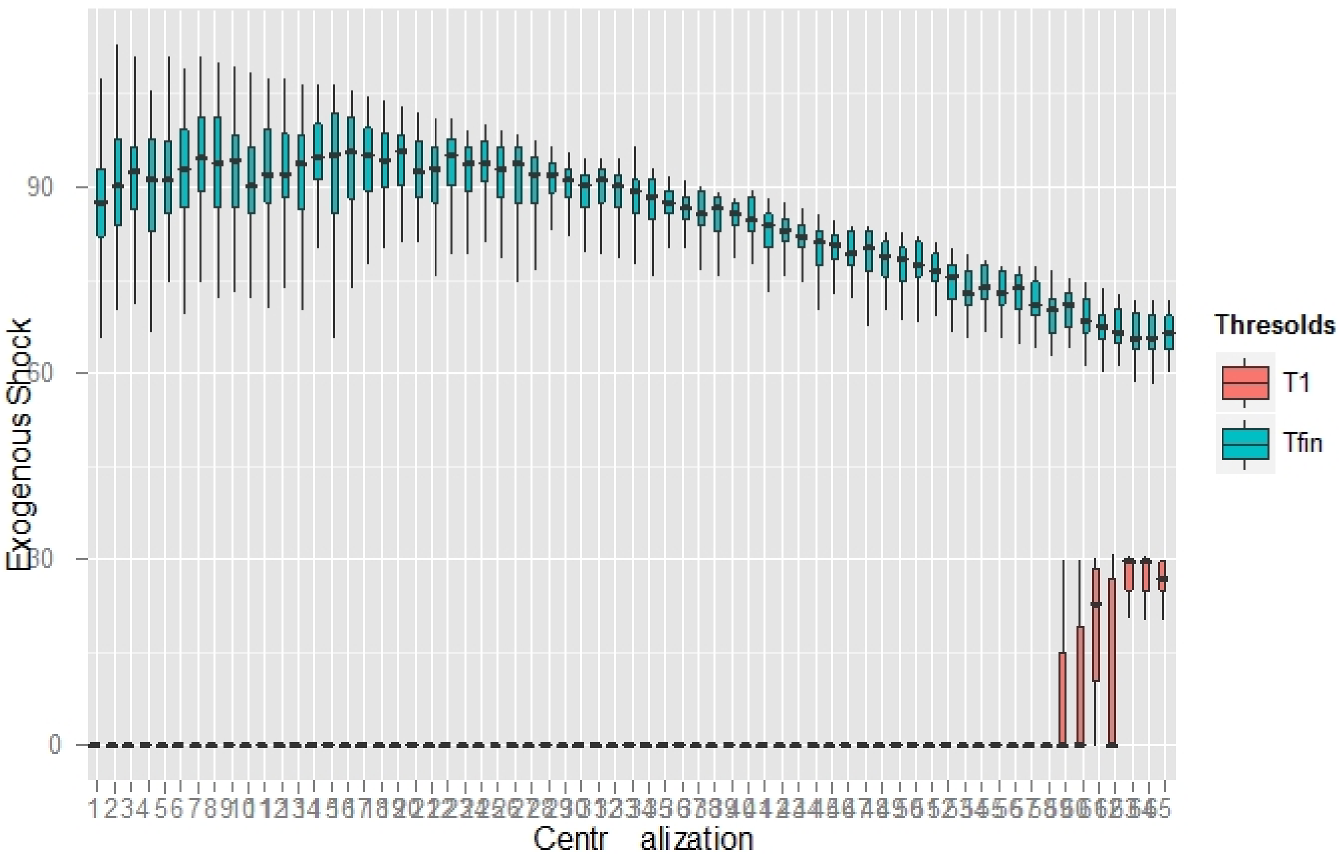

Figure 8: Threshold T2 and Final Threshold as function of centralization; $\phi = 0.4$, $\varepsilon = 0.1$, constant for all nodes

4 Conclusions

In this paper we make some conjectures about the impact of centralization and connectivity on the response of financial network to exogenous shocks to the assets held by the members of the networks. These conjectures are based on analytic results that pointed out that, in stylized examples of networks, i) complete connectivity as well as maximum centralization render a network robust-yet-fragile, in the sense that these types of networks have a single contagion threshold: for shocks smaller than such a threshold, no secondary defaults occur, while for shocks larger than such a threshold all members of the network default; and, conversely, ii) the network with minimum connectivity and minimum centralization is characterised by a large gap between the first and the final thresholds of contagion, thus it displays a vulnerable-yet-resilient behaviour: it is exposed to episodes of local contagion due to relatively small shocks while it is resilient with respect to large shocks.

We conjecture that these effects of centralization and connectivity apply to generic network in a fashion that is proportional to their degrees of density and centralization. To test these conjectures, we run numerical simulations on randomly generated networks with varying degrees of connectivity and centralization. The results we obtained confirm our conjectures. We tested the conjectures of the effects of connectivity by running simulations on the class of regular networks, where all node have the same degree of centrality, hence centralization is kept at zero. We obtain that, as density increases, the networks become progressively more robust-yet-fragile: the first and the final contagion thresholds converge to the unique threshold of the pure star-shaped configuration in a quasi-monotonic fashion. Similarly, we tested the effects of centralization on a class of networks with constant and almost minimum connectivity, moving from circular networks towards star-shaped networks. We find that the gap between the first and the final thresholds of contagion decreases as we move from sparse and decentralised networks towards sparse and highly centralised networks, showing that the vulnerable-yet-resilient features of the circular networks is progressively replaced by the robust-yet-fragile nature of the highly centralised star-shaped network. Also in these test, the convergence of the first and the final thresholds towards the unique threshold of the star network in quasi-monotonic.

Interestingly, our results show that the pattern of the first threshold in the first set of experiments is clearly convex, while the pattern of the same threshold in the second set of experiments is clearly concave. The exact opposite applies to the pattern of the final threshold of contagion: in the simulations aimed to test the effects of connectivity, the final threshold shows a concave pattern while, in the simulations that test the effects of centralization, the final threshold has a neatly convex pattern. This result indicates that the effects of increasing connectivity, in rendering a network robust-yet-fragile, became noticeable starting from relatively low levels of connectivity. Conversely, increasing centralization, that also makes a network increasingly robust-yet-fragile, yields evident effects only for high values of centralization. In other words, the losses due to exogenous shocks are distributed among the members of a network in an even fashion - generating the robust-yet-fragile phenomenon - starting from relatively low levels of density. Conversely, increasing centralization delivers the same effect but only for high levels of centralization.

References

[1] Acemoglu, D., A. Ozdaglar and A. Tahbaz-Salehi (2013), 'Systemic Risk and Stability in Financial Networks', NBER Working Paper, No. 18727.

[2] Acemoglu, D., Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2015), "Networks, shocks, and systemic risk", No. w20931. National Bureau of Economic Research.

[3] Ahuja, K. R., Magnanti, T. L. and Orlin, J. B. (1993) *Network Flows: theory, algorithms and applications*, Prentice Hall, New Jersey.

[4] Allen, F. and Gale, D. (1998) "Optimal Financial Crises", *Journal of Finance*, 53, 1245-48.

[5] Allen, F. and Gale, D. (2000) "Financial Contagion", *Journal of Political Economy*, 108(1): 1-34.

[6] Blavarg, M. and Nimander, P., (2002) "Interbank exposures and systemic risk". *Sveriges Riksbank's Economic Review*, no 2, 2002, pp 19-45

[7] Canedo, J. M. D., and Martínez Jaramillo, S., (2009) "A network model of systemic risk: stress testing the banking system", *Intelligent Systems in Accounting, Finance & Management*, Volume 16 Issue 1-2, p. 87-110.

[8] Degryse, H., and Nguyen, G. (2004) "Interbank Exposure: An Empirical Examination of Systemic Risk in the Belgian Banking System". CentER Discussion Paper No. 2004-04. URL http://ssrn.com/abstract=556944

[9] Diestel, R. (2000) *Graph Theory*, Springer-Verlag, New York

[10] Dow, James (2000) "What is systemic risk? Moral hazard, initial shocks and propagation", *Monetary and Economic Studies*, December.

[11] Eboli, M. (2013) "A flow network analysis of direct balance-sheet contagion in financial networks. Technical report 1862, Kiel Institute for the World Economy.

[12] Eboli, M. (2014) "On the Indeterminacy of the Clearing Payment Vectors in Numerical Simulations on Financial Networks", *Advances in Computational Social Science Agent-Based Social Systems*, Volume 11, pp 61-72.

[13] Eboli, M. (2016) 'Robust-yet-fragile' and 'vulnerable-yet-robust'financial networks: A flow network analysis of systemic risk. Working Paper.

[14] Eisenberg, L., Noe, T. H., (2001) "Systemic risk in financial systems", *Management Science*, 47 (2), 236-249.

[15] Elsinger, L., Lehar, A. and Summer, M. (2002) "Risk assessment of banking systems" Austrian Central Bank, mimeo.

[16] Evrensel, A. (2008), "Banking crisis and financial structure: A survival-time analysis", *International Review of Economics and Finance*, 17, 589-602.

[17] Freixas, X., L. Parigi and J. C. Rochet. (2000) "Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank", *Journal of Money, Credit and Banking*, n.32.

[18] Furfine, C.H , (1999) "Intebank exposure: quantifying the risk of contagion", BIS working papers, n° 70.

[19] Haldane, Andrew G. (2009), "Rethinking the financial network." Speech delivered at the Financial Student Association in Amsterdam.

[20] Hojman, D. and Szeidl, A., (2008) "Core and periphery in endogenous networks", *Journal of Economic Theory*, Volume 139, Issue 1, March 2008, Pages 295–309

[21] Mistrulli, P. E. (2005), "Interbank lending patterns and financial contagion". The Bank of Italy Working paper.

[22] Mistrulli, P. E. (2006), "Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns", *Journal of Banking & Finance*, vol. 35, issue 5, May 2011, pages 1114–1127.

[23] Nier, E., Yang, J., Yorulmazer, T. and Alerton, A. (2007) "Network models and financial stability", forthcoming in the *Journal of Economic Dynamics and Control*.

[24] Pokutta, S., Schmaltz, C. and Stiller, S. (2011) "Measuring Systemic Risk and Contagion in Financial Networks". Available at SSRN: http://ssrn.com/abstract=1773089.

[25] Rochet, J-C. and Tirole, J. (1996) "Interbank Lending and Systemic Risk", *Journal of Money, Credit and Banking*, 28, 733-62.

[26] Shin, H.S., Cifuentes, R. and Ferrucci, G. (2005) "Liquidity Risk and Contagion", *Journal of the European Economic Association*, Vol. 3, No. 2-3, Pages 556-566.

[27] Upper, C., and Worms, A. (2004) "Estimating bilateral exposure in the german interbank market: is there a danger of contagion?", *European Economic Review*, Volume 48, pp 827–849

[28] Upper, C. (2007) "Using Counterfactual Simulations to Assess the Danger of Contagion in Interbank Markets". August,BIS Working Paper No. 234

[29] Wells, S. (2002) "U.K. Interbank exposure: systemic risk implications", *Financial Stability Review* 13, 175-82, Bank of England.