Theory of Systemic Risk

Chen Chen

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Theory of Systemic Risk

Chen Chen

Systemic risk is an issue of great concern in modern financial markets as well as, more broadly, in the management of complex business and engineering systems. It refers to the risk of collapse of an entire complex system, as a result of the actions taken by the individual component entities or agents that comprise the system. We investigate the topic of systemic risk from the perspectives of measurement, structural sources, and risk factors. In particular, we propose an axiomatic framework for the measurement and management of systemic risk based on the simultaneous analysis of outcomes across agents in the system and over scenarios of nature. Our framework defines a broad class of systemic risk measures that accommodate a rich set of regulatory preferences. This general class of systemic risk measures captures many specific measures of systemic risk that have recently been proposed as special cases, and highlights their implicit assumptions. Moreover, the systemic risk measures that satisfy our conditions yield decentralized decompositions, i.e., the systemic risk can be decomposed into risk due to individual agents. Furthermore, one can associate a shadow price for systemic risk to each agent that correctly accounts for the externalities of the agent's individual decision-making on the entire system. Also, we provide a structural model for a financial network consisting of a set of firms holding common assets. In the model, endogenous asset prices are captured by the marketing clearing condition when the economy is in equilibrium. The key ingredients in the financial market that are captured in this model include the general portfolio choice flexibility of firms given posted asset prices and economic states, and the mark-to-market

wealth of firms. The price sensitivity can be analyzed, where we characterize the key features of financial holding networks that minimize systemic risk, as a function of overall leverage. Finally, we propose a framework to estimate risk measures based on risk factors. By introducing a form of factor-separable risk measures, the acceptance set of the original risk measure connects to the acceptance sets of the factor-separable risk measures. We demonstrate that the tight bounds for factor-separable coherent risk measures can be explicitly constructed.

Contents

Acknowledgements

First of all, I feel very fortunate to have had the opportunity to work with my advisors, Professor Garud Iyengar and Professor Ciamac Moallemi. I would like to thank them for their invaluable guidance and encouragement along the way to the completion of this dissertation. It has been a wonderful experience working with them. They are undoubtedly two of the most intelligent people I have ever worked with. Thanks to them, I have the opportunity to explore the interesting field of optimization and risk management. During our meetings, it has been a luxury for me to be able to discuss academic problems with two brilliant minds. Both of them helped me on my research with great patience. I would not have been accomplished my thesis without them. Both Professor Iyengar and Professor Moallemi have been great advisors for my PhD study and mentors for my life. Their consistent support and enthusiasm on research are unforgettable to me. What I have learned from my advisors helps me grow into a researcher, an educator, and a responsible person.

I would like give my most sincere gratitude to my other dissertation committee members, Professor Paul Glasserman, Professor Tim Siu-Tang Leung, Professor Alireza Tahbaz-Salehi, for their careful reading of my thesis manuscript. They have raised great questions and provided me with constructive comments, which helped improve the thesis.

Also, I am very grateful to my schoolmates and officemates. My study and research has benefited from the collaboration and discussions with them. It was a privilege to have worked with such a group of talented people. During the past five years at Columbia University, I have met a lot of people who have become very good friends and an integral part of my life. They have made my life more colorful and the office always feel like home. I wish to express my sincere gratitude to

Jing Guo, Ningyuan Chen, Brian Lu, Cun Mu, Xinyun Chen, Bo Huang, Juan Li, Chun Wang, Jing Dong, Yupeng Chen, Anran Li, Chun Ye, Xingbo Xu, Yina Lu, Xin Li, Zheng Wang, Zhen Qiu, Chen Xie, Jingfei Jia, Lijian Lu, Yunan Liu, Shiqian Ma, Ruxian Wang, Haowen Zhong, Zhiwei Qin, Linan Yang, Yunru Han, Hua Zheng, Hailey Song Hee Kim, Rodrigo Carrasco, Necdet Serhat Aybat, Yixi Shi, Gordon Pang, Carlos Adrian Abad Lopez, Itai Feigenbaum, Marco Santoli, Peter Maceli, Shyam Sundar Chandramouli, Aya Wallwater, Matthieu Plumettaz.

I am extremely indebted to my parents, whose are always unconditional supportive and have helped me to fulfil my educational goals and complete my journey to the PhD degree.

Chen Chen

June 30, 2014

To my parents

Chapter 1

Introduction

1.1 Background and Motivation

Systemic risk refers to the risk of collapse of an entire complex system, as a result of actions taken by the individual component entities or agents that comprise the system. The management of systemic risk is of fundamental importance in many applications, since engineering and economic systems have grown ever larger, more complex, and more central in providing for a more and more complicated set of societal needs. The reliability of these systems has, therefore, become very critical for societal well being. It is increasingly the case that the risk of individual components of such systems is reasonably well characterized, and there are quantitative tools available to measure this risk. However, the critical issue of systemic risk, i.e., understanding the reliability of the overall system, is poorly understood. Much of the academic literature on the theoretical foundations for risk management, as well as the main regulatory standards, has been focused on the study of individual agents in isolation. Systemic risk, on the other hand, poses a new set of challenges.

Systemic risk emerges from the interaction of the components through mechanisms such as contagion, i.e., the possibility of a cascading series of failures in an interconnected system. Managing systemic risk and identifying too-big-to-fail firms capable of igniting contagion when they fail is one of the central issues being addressed in the new financial regulations that are currently being debated. This is similar to the way smart grid related technologies attempt to develop techniques to prevent blackouts in a critically stressed power network. Despite the importance of systemic risk, systemic risk measures and the tools available to measure and control systemic risk are very ad hoc and extremely application specific. This thesis will develop a unified underlying theory for systemic risk by employing axiomatic approaches, structural properties of networks and factor analysis to understand the impact of interacting agents in a system facing uncertain future scenarios. The proposed theory allows for the identification of a set of risk management principles that are common to all instances of systemic risk, and is flexible enough to be adapted to suit particular application domains. This thesis yields analytical and computational methods that can be applied to minimize catastrophic failures in a broad range of settings. A number of recent events across many disparate settings (for example, the Deepwater Horizon oil spill in 2010, the financial crisis in 2008, the blackout in the Northeastern United States in 2003) have all highlighted the critical importance of risk management in general, and systemic risk management in particular. Progress on developing tools for systemic risk management will help prevent, or at least better manage, future systemic failures in financial, public health, environmental, national security, and governmental domains. This thesis propose a theoretical framework for understanding, measuring, and estimating systemic risk.

1.2 Measurement of Systemic Risk

The measurement and management of systemic risk is of fundamental importance in many business and engineering domains. The manager of a diversified firm has to assess and control the collective risk of all individual divisions or business units. The manager of a supply chain network must evaluate the overall risk associated with a complex network of suppliers and sub-contractors. The manager of an electric power distribution network needs to determine the aggregate risk of the generating stations, transmission facilities, and other entities in the network. As highlighted by the financial crisis of 2007–2008, one example of particular interest is the measurement and regulation of systemic risk of an economy or a financial market. While the methods proposed in this thesis are general, and we seek to develop a broad understanding of systemic risk management, we will focus on this last case and use the language of financial markets to present the results.

Fundamentally, the study of systemic risk in a financial market involves the simultaneous analysis of outcomes across all entities (firms) in the economy. On the other hand, in the academic literature and major regulatory standards, the only topic has been well studied is the risk for individual firms. We seek to bridge this gap by developing an axiomatic framework for a broad class of systemic risk measures.

Specifically, we are interested in an approach to systemic risk that is based on the analysis of the joint distribution of profits and losses across all firms in the economy and states of nature. We consider systemic risk from the perspective of a regulator, who wishes to express a preference over sets of possible distributions of outcomes for the entire economy. One approach to defining a systemic risk measure is to apply a traditional, single-firm risk measure such as value-at-risk or conditional value-at-risk to the distribution of the *total* profits and losses for all firms in the economy [e.g., Adrian and Brunnermeier, 2009, Acharya et al., 2010a, Tarashev et al., 2010]. This approach treats the entire economy as a portfolio consisting of the constituent firms, and the regulator as a portfolio manager. However, the portfolio approach suffers from a number of modeling shortcomings. It implicitly allows the netting of profits and losses across the portfolio components. This is reasonable from the perspective of the manager of an investment portfolio. However, such netting may be undesirable from the perspective of a systemic regulator who, typically, is not able to directly cross-subsidize different firms with distinct ownership interests. Moreover, by considering only the total outcome, the portfolio approach lacks the modeling flexibility to accommodate preferences over the cross-sectional distribution of outcomes in an economy. For example, the regulator may have views on whether it is preferable for a single firm to have a large loss or many firms to have small losses, or whether profits at one firm can subsidize losses at another.

In the literature, the measurement of risk has been explored by many studies over the past decades. In particular, Artzner et al. [1999] introduced several axioms that define coherent risk measures. Based on their work, many variations of risk measures have been investigated. However, most of the work only focuses on single-firm risk measures. To extend the framework from a single-firm setting to a systemic setting requires non-trivial analysis. Inspired by the work by Artzner et al. [1999], this thesis develops a general framework for defining system risk measures for a set of interacting agents with uncertain joint outcomes. We identify a set of axioms that characterize a very general class of systemic measures. This class of risk measures includes all the ad hoc risk measures that have been proposed in the literature. The approach leads to a primal-dual representation for the systemic risk that allows us to distribute the overall risk to individual agents in a fair manner, as well as to develop a decentralized algorithm for the social welfare maximization that is sensitive to systemic risk. We also propose to extend this representation to include a larger class of risk functions.

1.3 Contagion in Financial Networks

To understand the mechanism of systemic interactions that drives systemic risk, contagion is one of the critical topics to investigate. Contagion is an endogenous phenomenon that characterizes negative outcomes for one agent having an impact throughout the system. It features the possibility of a cascading series of failures in an interconnected system. Understanding systemic risk requires careful analysis of the mechanisms of interaction between the individual agents. Managing systemic risk and identifying systemically important ingredients that amplify contagion is one of the central issues being addressed in the new financial regulations. In particular, during the recent financial crisis, contagion effects played an important role in causing catastrophic results. The contagion in the financial system stems from complex sources.

Two major types of financial interactions contribute to the process of contagion. One is the relationship of interbank lending and borrowing, and the other is the asset-firm holding connection. The former type produces the possible propagation of counterparty defaults that stem from institutions, while the latter highlights the cascading effects through market activities, for example, portfolio re-balancing from asset holders and asset price changes. In the recent literature, there have been articles that emphasize the interbank debt obligations. One of the most influential papers was written by Eisenberg and Noe [2001] who proposed a structural model for an interbank lending network that allows for an algorithmic debt clearing mechanism. On the other hand, there has been less discussion on the topic of asset price based contagion in the current literature. The lack of analytical understanding of asset-firm interaction has been the inspiration for the work in this paper, since the contagion through change in asset prices and portfolio re-balancing is a major contributor, if not the most crucial one, to the systemic risk in financial systems compared to other contributing elements of contagion. In fact, Glasserman and Young [2014] demonstrates that the obligation-based "interconnectedness" of the financial network may have little influence on severe financial contagion. Meanwhile, other investigations point out that some commonly-held toxic assets may have caused the problem. For example, Wallison [2012] from The Wall Street Journal states that "None of these firms was weakened by its exposure to Lehman or anyone else. They were weakened by the fact that virtually all of them held, or were suspected of holding, large amounts of what the media came to call toxic assets." This statement points out an insightful truth: massive price fluctuations of assets across the market impose the systemic risk events of the most severe consequences. Thus, the investigation of asset-based contagion is a critically relevant topic in understanding the systemic issues in financial crisis. This thesis seeks to develop new structural models to understand the role of contagion in complex systems. In what follows, we will also explore the design of the network structure underlying complex systems to minimize overall risk of contagion.

1.4 Risk Factors

Another key issue in risk management is to provide accurate estimates of risk measurement. For example, in financial applications, a typical problem is how to estimate the risk of the outcomes for portfolios. In practice, the most commonly applied method of tackling this problem is to model the random outcomes of the portfolio through a set of risk factors. This approach is appealing since the selected risk factors are often better understood than the distributions of random outcomes of the portfolio per se. For instance, many macroeconomic factors are chosen as risk factors, including stock indices and crude oil prices. These factors have more complete historic data available for analysis. Meanwhile, some risk factors are updated more frequently for rapid updates of the consequent risk estimates.

Generally speaking, risk factors explain only part of the risk of the system. How much of the risk is explained depends on how the system interplays with the risk factors that are chosen. It is possible that the residuals that have not been covered by the risk factors remain a significant source of the overall risk. The commonly used approach in current practice is to assume that the residuals follow a simplified distribution, e.g., a normal distribution. For instance, for a typical stress testing, the regulator specifies certain hypothetical scenarios where the states of the macroeconomic risk factors are given. For those scenarios, firms need to report the estimates of the outcomes of their portfolios. With these estimates of outcomes, the regulator can evaluate the risk level of the firm.

In this process of estimating risk through stress testing, we wish to point out two observations. One observation is that the risk evaluation from the regulator implicitly imposes an acceptance set that determines whether a profile of the outcome estimates is acceptable or not. This implicit acceptance set is equivalent to defining a risk measure ρ that takes as input the outcome estimates (See Artzner et al. [1999]). The other observation is that when firms report the outcome estimates of their portfolios, they essentially take the expected outcomes of whatever models they are using for the estimates. There are some drawbacks in the described practice. It is not difficult to prove that putting the expectation of outcomes given factor realizations as inputs into the risk measure ρ always provides an underestimate of a convex risk measure without conditioning on values of risk factors (See Cherny and Madan [2006]). As far as risk management is concerned, constantly producing underestimates as the reported value of the risk not only offers misleading information, but can also lead to bad risk decisions due to the lack of awareness of the risk resulting from the residuals. By applying the expected loss function, this type of risk model tries to explain the risk contributed by all the factors and implicitly assumes that the residuals are diversifiable given factor values. It does not capture the uncertainty that lies in the residuals. In situations where the chosen risk factors explain outcomes poorly, the risk embedded in the residuals may deliver a significant contribution to the overall risk. In this thesis, we propose to develop a methodology for the analysis of systemically important factors and the risk measures based on these factors. This methodology would lead to robust estimation of systemic risk.

1.5 Organization of This Thesis

The balance of this thesis is organized as follows:

Chapter 2. In this chapter, we define an axiomatic framework defining systemic risk, and establish an associated structural decomposition. We establish a dual representation for systemic risk, that allows attribution of risk to individual agents.

Chapter 3. In this chapter, we provide a structural model for a financial network consisting of a set of firms holding common assets. We provide a price sensitivity analysis for the financial network. We characterize the key features of financial holding networks that minimize systemic risk, as a function of overall leverage.

Chapter 4. In this chapter, we provide a framework to analyze risk measures based on factors. Also, we provide tight bounds for factor-separable coherent risk measures.

Chapter 5. In this chapter, we offer some concluding remarks, and discuss directions for future work.

Chapter 2

Systemic Risk Measures: An Axiomatic Approach

2.1 Introduction

In this chapter, we investigate fundamental properties of the measurement of systemic risk. Motivated by the fact that there have been a variety of ad hoc systemic risk measures proposed in the literature, we seek to propose a unified theory that subsumes most of them. Thus, we define a broad class of systemic risk measures that can accommodate a rich set of regulatory preferences and practical applications. The main contributions of this chapter are as follows:

- *We define an axiomatic framework defining systemic risk, and establish an associated structural decomposition.*

  Our work parallels the axiomatic approach to single-firm risk measures by Artzner et al.

[1999]. Schied [2006] provides a very good survey of the extensive literature on coherent and convex risk measures for a single firm. Unlike the single-firm case, however, we consider a system or economy that consists of multiple components or firms. Systemic risk is then defined functional on the joint distribution of outcomes across firms in an economy and scenarios (states of nature) that satisfies a set of axioms. While we impose many axioms developed originally for single-firm risk measures to systemic risk measures to address similar concerns; we introduce two new axioms. In particular, as in the case of coherent risk measures for a single firm, we assume the monotonicity and positive homogeneity of systemic risk. Besides the usual notion of convexity, we introduce a new risk convexity concept for situations where outcomes are not directly combined. Additionally, we assume a preference consistency condition that relates to the interactions between different firms across scenarios. The latter condition is novel and fundamentally specific to systemic risk; it has no analog among the typical conditions for single-firm risk measures, and it becomes trivial if the economy consists of a single firm.

We demonstrate that any systemic risk measure satisfying our definition can be characterized by two independent components: (1) an aggregation function that expresses a preference over the cross-sectional profile of outcomes across firms in a single scenario, and (2) a base risk measure, similar to existing single-firm risk measures, that expresses a preference over the profile of aggregated outcomes across scenarios of nature. This structural decomposition provides a clear structural characterization of systemic risk, and suggests a well-defined procedure to construct such risk measures by choosing constituent aggregation functions and base risk measures. This decomposition highlights the power of the preference consistency condition.

Our framework includes many recently proposed systemic risk measures as special cases. For example, a number of authors analyze systemic risk by applying single-firm risk measures to a portfolio consisting of all firms in the economy [e.g., Gauthier et al., 2010, Tarashev et al., 2010]. The 'systemic expected shortfall' risk measure of Acharya et al. [2010a] employs a portfolio approach, with an expected shortfall base risk measure. An alternative reduced-form approach to systemic risk involves considering the price of deposit insurance or other credit insurance [e.g., Lehar, 2005, Huang et al., 2009]. Giesecke and Kim [2011] consider a risk measure defined through the fraction of failed firms in the economy. The general framework in this chapter subsumes a number of these approaches. In Section 2.2, we illustrate how portfolio-based approaches to systemic risk measures such as the systemic expected shortfall or deposit insurance can be modeled as special cases in our framework. However, our framework provides greater flexibility in modeling systemic risk, allowing, for example, complex non-linear interactions between firms.

- *We establish a dual representation for systemic risk, that allows attribution of risk to individual agents.*

  We show that any systemic risk measure can be expressed as the worst-case expected loss over a family of distributions over scenarios of nature and the cross-sectional profiles of firms, a generalization of the dual representation for single-firm coherent risk measures [Artzner et al., 1999]. In many cases, this representation provides operational benefits by permitting decentralized computation of systemic risk by the firms in the economy. Moreover, we show that the dual variables are, in fact, shadow prices for systemic risk: they represent the marginal increase in systemic risk as a function of a marginal increase in the loss of a particular firm in a particular scenario.

In our setting, the dual representation provides a mechanism for risk attribution. The total systemic risk can always be apportioned across the constituent firms in a way that satisfies a 'no-undercut' condition: the systemic risk allocated to any subset of the firms is no more than the systemic risk those firms would face as a stand-alone economy. Our allocation rule is a generalization of the Aumann-Shapley prices for fair allocation of costs or the Euler allocation rule for allocating the capital requirements of a portfolio across constituent subportfolios [Denault, 2001, Buch and Dorfleitner, 2008]. Similarly, the 'marginal expected shortfall' risk attribution of Acharya et al. [2010a] is a special case of our attribution rule.

We show that the risk attribution can properly account for the externalities imposed on the system when making decisions involving risk. Specifically, through a decentralized taxation scheme, the objective of the regulator can be aligned with the incentives of individual firms. Here, each individual firm maximizes the difference between its individual utility function and a tax payment that is derived from the firm's contribution to the systemic risk.

- *Our methodology extends to a general class of risk measures.* The structural decomposition of a systemic risk function into an aggregation function and a base risk measure follows from the preference consistency condition, and therefore, can be extended to broader classes of risk measures. In Section 2.5, we consider homogeneous systemic risk measures. These are systemic risk measures that are positively homogeneous and monotonic, but not necessarily convex. One example of such a risk measure is that of Adrian and Brunnermeier [2009], who define a risk measure based on the value-at-risk of the economy-wide portfolio. We show that homogeneous systemic risk measures that satisfy preference consistency can be decomposed into a single-firm homogeneous base risk measure and a homogeneous aggregation function. We describe a risk attribution scheme for a special class of piecewise linear homogeneous systemic risk measures that is a generalization of Aumann-Shapley prices. Similarly, a convex, monotonic, but not necessarily positively homogeneous, systemic risk measure that satisfies preference consistency can also be decomposed into a convex monotonic single-firm risk measure and a convex monotonic aggregation function.

Other authors have sought to model the structural mechanisms of interaction between firms in a financial crisis. Such models explicitly describe the contagion of credit events across firms in an economy through different structural mechanisms. For example, Acharya et al. [2010b] and Staum [2011a] consider asset price contagion, while Eisenberg and Noe [2001], Liu and Staum [2010], and Cont et al. [2010] consider counterparty contagion. Staum [2011b] provides an excellent survey of the literature on contagion and systemic risk. In this chapter, we take as given a collection of exogenous outcomes across firms and scenarios of nature. However, we can accomodate aspects of endogenous, structural mechanisms for contagion through the choice of risk measure. This is illustrated in Example 7 in Section 2.2.

The rest of the chapter is organized as follows: In Section 2.2, we provide an axiomatic definition of a systemic risk measure. In Section 2.2.1, we describe the structural decomposition of systemic risk, and in Section 2.2.2 we discuss a number of examples of systemic risk measures. In Section 2.3, we construct primal and dual variational representations for systemic risk measures. In Section 2.4, we discuss a systemic risk attribution scheme. In Section 2.5, we present extensions of our theory to homogeneous systemic risk measures. In the Online Supplement, we demonstrate a decentralized framework for systemic risk management, as well as provide proofs.

2.2 Model Formulation

We consider a one-period model consisting of a finite set of firms $\mathcal{F}$ and a finite set of future scenarios Ω. We define an *economy* by a matrix $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$. Here, the quantity $X_{i,\omega}$ is the loss (or, if negative, the profit) of firm i in scenario ω. We denote by $X_\omega \in \mathbb{R}^{|\mathcal{F}|}$ the column vector of outcomes in scenario ω across all firms; we refer to this as the *cross-sectional profile of losses* across firms of the economy X, in scenario ω. In some examples, we assume there is a probability distribution $p \in \mathbb{R}_+^{|\Omega|}$ over the space of scenarios Ω. In these cases, we can interprete the matrix $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ as a random vector which has outcome $X_\omega \in \mathbb{R}^{|\mathcal{F}|}$ with probability p_ω.

In the rest of the chapter, the following notation is helpful: the vector $\mathbf{1}_\Omega \in \mathbb{R}^{|\Omega|}$ denotes a unit loss of an individual firm in all scenarios, and vector $\mathbf{1}_\mathcal{F} \in \mathbb{R}^{|\mathcal{F}|}$ denotes a cross-sectional loss profile in a scenario where each firm has a unit loss, and the matrix $\mathbf{1}_\mathcal{E} \triangleq \mathbf{1}_\mathcal{F}\mathbf{1}_\Omega^\top \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ denotes an economy with a unit loss for every firm in every scenario. Similarly, the vectors $\mathbf{0}_\mathcal{F} \in \mathbb{R}^{|\mathcal{F}|}$, $\mathbf{0}_\Omega \in \mathbb{R}^{|\Omega|}$, and the matrix $\mathbf{0}_\mathcal{E} \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ correspond to cases with zero profit or loss for every firm in every scenario. Given an economy X, a cross-sectional loss profile x, and a scenario ω, the matrix $(x, X_{-\omega}) \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ denotes an economy with loss profile x in scenario ω, but where outcomes in all other scenarios are given by the corresponding columns in X. Inequalities between pairs of vectors and matrices are to be interpreted component-wise.

A *systemic risk measure* ρ is a summary statistic that quantifies the level of 'risk' associated with an economy X by a single real number $\rho(X)$. Given two economies X and Y, if $\rho(X) > \rho(Y)$ then we say that X is riskier than Y and thus less preferred. Hence, a systemic risk measure implicitly encodes the preferences of a regulator over the universe of possible economies.

We first review the axiomatic framework for coherent *single-firm*[1] *risk measures*[2] commonly used in the literature [Artzner et al., 1999].

Definition 1 (Single-Firm Risk Measure). *A single-firm risk measure is a function*[3] $\rho\colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ *that satisfies the following conditions 1–4, for all loss vectors* $x, y \in \mathbb{R}^{|\Omega|}$ *of a single firm:*

1. Monotonicity: *If* $x \geq y$*, then* $\rho(x) \geq \rho(y)$.

2. Positive homogeneity: *For all non-negative scalars* $\alpha \geq 0$, $\rho(\alpha x) = \alpha\rho(x)$.

3. Convexity: *Given a scalar* $0 \leq \alpha \leq 1$, $\rho\big(\alpha x + (1-\alpha y)\big) \leq \alpha\rho(x) + (1-\alpha)\rho(y)$.

4. Normalization: $\rho\big(\mathbf{1}_\Omega\big) = 1$.

If in addition, a single-firm risk measure satisfies the following condition 1, it is called coherent:

[1]In this chapter, we use the term 'single-firm' risk measure to refer to the risk measures for a single entity, i.e., an entity for which the outcome in every scenario of nature is a single real number. This is in contrast to the systemic risk measures which we will introduce shortly, where there is a vector of outcomes (one for each component of the system) in every scenario of nature. Note that a systemic risk measure could, for example, also be applied in the case of an individual firm, where the 'components' correspond to divisions of the firm that contribute to the overall risk.

[2]Our terminology is slightly non-standard here: for example, for single-firm risk measures, Schied [2006] defines a 'monetary measure of risk' as satisfying 1 and 1, a 'convex measure of risk' as satisfying 1 and 3–1, and a 'coherent measure of risk as satisfying 1–1.

[3]In what follows, we sometimes consider single-firm risk measures $\rho\colon \mathbb{R}_+^{|\Omega|} \to \mathbb{R}$ defined only on the positive orthant. In that case, we assume that conditions 1–4 are satisfied for all $x, y \in \mathbb{R}_+^{|\Omega|}$.

1. Cash invariance: *For all scalars* $\alpha \in \mathbb{R}$, $\rho(x + \alpha \mathbf{1}_\Omega) = \rho(x) + \alpha$.

The conditions for a single-firm risk measure can be motivated as follows: The monotonicity condition 1 reflects that, if one firm has greater losses in every scenario than another, it is less preferred. The positive homogeneity condition 2 requires that the risk increases in proportion to the scale of losses. The convexity condition 3 asserts that diversification reduces risk, i.e., the risk of a firm diversified between outcomes corresponding to x and y is less than the weighted risk of the component firms x and y. The normalization condition 4 fixes the multiplicative scaling[4] of the risk measure. The cash invariance condition 1 allows the interpretation of risk as a capital requirement: when a certain loss α is added to the outcome in every scenario, the risk of the firm increases by exactly α.

Building on the definition for a single-firm risk measure, we formally define a systemic risk measure as follows:

Definition 2 (Systemic Risk Measure)**.** *A systemic risk measure is a function* $\rho \colon \mathbb{R}^{|\mathcal{F}| \times |\Omega|} \to \mathbb{R}$ *that satisfies the following conditions, for all economies* $X, Y, Z \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$*:*

1. Monotonicity: *If* $X \geq Y$*, then* $\rho(X) \geq \rho(Y)$.

2. Positive homogeneity: *For all non-negative scalars* $\alpha \geq 0$, $\rho(\alpha X) = \alpha \rho(X)$.

3. Preference consistency: *Define a partial order* $\succeq_\rho$ *on cross-sectional profiles as follows: Given cross-sectional profiles* $x, y \in \mathbb{R}^{|\mathcal{F}|}$*, we say that* $x \succeq_\rho y$ *iff* $\rho\big(x \mathbf{1}_\Omega^\top\big) \geq \rho\big(y \mathbf{1}_\Omega^\top\big)$. *Suppose that, for every scenario* ω, $X_\omega \succeq_\rho Y_\omega$. *Then,* $\rho(X) \geq \rho(Y)$.

[4]Note that in the case of a coherent single-firm risk measure, the normalization condition 4 follows from positive homogeneity and cash invariance. In our work, it will be useful to consider non-coherent risk measures, hence we retain this as a separate condition.

4. Convexity:

(a) Outcome convexity: *Suppose* $Z = \alpha X + (1-\alpha)Y$, *for a given scalar* $0 \leq \alpha \leq 1$. *Then,* $\rho(Z) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$.

(b) Risk convexity: *Suppose* $\rho(Z_\omega \mathbf{1}_\Omega^\top) = \alpha\rho(X_\omega \mathbf{1}_\Omega^\top) + (1-\alpha)\rho(Y_\omega \mathbf{1}_\Omega^\top)$, $\forall\ \omega \in \Omega$, *for a given scalar* $0 \leq \alpha \leq 1$. *Then,* $\rho(Z) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$.

5. Normalization: $\rho(\mathbf{1}_{\mathcal{E}}) = |\mathcal{F}|$.

Our definition of a systemic risk measure is justified as follows: conditions 1–2 are similar to the corresponding conditions for a single-firm risk measure, and can be justified in a similar manner. The preference consistency condition 3, on the other hand, does not have an analog in the single-firm case. This condition defines an ordering (or, preference relationship) $\succeq_\rho$ on cross-sectional profiles $x, y \in \mathbb{R}^{|\mathcal{F}|}$ by comparing the systemic risk (according to ρ) of the constant economies[5] $x\mathbf{1}_\Omega^\top$ and $y\mathbf{1}_\Omega^\top$. If $x \succeq_\rho y$, we say that y is preferred to x. Preference consistency requires that if cross-sectional profiles in the economies X and Y are such that if, in every scenario ω, Y_ω is preferred to X_ω, the systemic risk of Y must be consistent with this preference and thus cannot be greater than the systemic risk of X. When the economy consists of a single firm, condition 3 follows from monotonicity.

The preference consistency condition implies independence from irrelevant alternatives [see, e.g., Kreps, 1988] as follows: suppose that $x, y \in \mathbb{R}^{|\mathcal{F}|}$ are cross-sectional loss profiles such that $x \succeq_\rho y$, i.e., y is preferred to x. Then, for any economy $Z \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$ and any scenario ω, define $(x, Z_{-\omega})$ to be the economy where outcomes for firms in scenario ω are given by x, and outcomes

[5]Note that, given $x \in \mathbb{R}^{|\mathcal{F}|}$, the economy $x\mathbf{1}_\Omega^\top$ has the same cross-sectional profile of losses x in every scenario. Hence, we call this a constant economy.

in all other scenarios are given by Z, and define $(y, Z_{-\omega})$ similarly. Preference consistency implies that $\rho(x, Z_{-\omega}) \geq \rho(y, Z_{-\omega})$. In other words, if y is preferred to x, then, all else being equal, any economy which realizes y in some scenario is less risky to an economy which realizes x in that scenario, independent of the scenario and of the outcomes in other scenarios. Thus, by imposing the preference consistency axiom, we assume that the systemic risk measure expresses a preference over cross-sectional profiles that is consistent across scenarios. Introducing the preference consistency condition is one of the major contributions of this chapter, in that it allows us to extend the single-firm risk measure to a systemic risk measure that captures the interaction of many firms.

The convexity conditions 4a and 4b are both concerned with the benefits of diversification. Condition 4a, labeled 'outcome convexity', is the usual notion of convexity: when the economy Z is a diversified mixture of two economies X and Y, the risk of Z is no greater than the weighted combination of the risk of economies X and Y. Outcome convexity is concerned with a 'portfolio' of economies X and Y, in that we are allowed to add the outcomes from the two economies and the risk reduction comes from the fact that outcomes X and Y are possibly correlated.

Condition 4b is concerned with convexity as it relates to risk aversion. The context of this condition is as follows. We have *two* stages of uncertainty. The outcome of the first stage is the economy X with probability α and the economy Y with probability $1-\alpha$. In the second stage, the scenario ω and the firm outcomes corresponding to ω and the economy selected in the first stage are revealed. Note that in this setting we do not have a 'portfolio' of economies. The economy Z in condition 4b is such that in every scenario $\omega \in \Omega$ the risk $\rho(Z_\omega \mathbf{1}_\Omega^\top)$ is a convex combination of the risk $\rho(X_\omega \mathbf{1}_\Omega^\top)$ and $\rho(Y_\omega \mathbf{1}_\Omega^\top)$, i.e., the outcomes of economy Z are *not* subject to the first stage randomness. Condition 4b states that the risk of the 'average' economy Z is at most the convex combination $\alpha\rho(X)+(1-\alpha)\rho(Y)$. The risk reduction in this case comes from the removing one stage of randomness. When the economy consists of a single firm, condition 4b is implied by cash invariance and outcome convexity.

Finally, the normalization condition 5 requires the risk of a unit loss by all firms with certainty to be the total loss, i.e., the number of firms $|\mathcal{F}|$. This is simply a convenient choice of scaling and is imposed without loss of generality.

Note that our definition of systemic risk does not contain a cash invariance condition, as required by a coherent single-firm risk measure. This is because we want to allow for systemic risk measures derived from deposit insurance that are incompatibile with cash invariance: if all outcomes in the future are reduced by a deterministic amount, this does not necessarily result in a commensurate reduction in the price of deposit insurance.

2.2.1 Structural Decomposition

In order to assess the systemic risk of an economy, a regulator is concerned with *both* the cross-sectional profile of losses across firms *and* the distribution of aggregate outcomes across scenarios. Thus, in order to define a risk preference over the universe of economies, one might seek to independently express these two types of preferences. We formalize this notion as follows:

Definition 3 (Aggregation Function). *A function $\Lambda\colon \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ over cross-sectional loss profiles of firms is an aggregation function if, for all cross-sectional loss profiles $x, y \in \mathbb{R}^{|\mathcal{F}|}$, it satisfies:*

1. Monotonicity: *If $x \geq y$, then $\Lambda(x) \geq \Lambda(y)$.*

2. Positive homogeneity: *For all $\alpha \geq 0$, $\Lambda(\alpha x) = \alpha\Lambda(x)$.*

3. Convexity: *For all* $0 \leq \alpha \leq 1$, $\Lambda\big(\alpha x + (1-\alpha y)\big) \leq \alpha \Lambda(x) + (1-\alpha)\Lambda(y)$.

4. Normalization: $\Lambda\big(\mathbf{1}_{\mathcal{F}}\big) = |\mathcal{F}|$.

An aggregation function provides a summary statistic that encapsulates a cross-sectional profile of losses across firms in a single scenario into a real number, thus expressing a preference over such profiles. The conditions 1–4 are analogous to the corresponding conditions for a systemic risk measure, and motivated by similar concerns. Subject to these conditions, the regulator has considerable freedom in specifying preferences over the distribution of losses across firms, and we will see a number of examples of aggregation functions in what follows.

Once the cross-sectional outcomes across firms are aggregated, the evaluation of systemic risk reduces to an evaluation of the profile of aggregated outcomes across scenarios. This can be accomplished by a single-firm risk measure $\rho_0 \colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ (Definition 1), which we call the *base risk measure*. The independent choice of an aggregation function and a base risk measure provides a clear way to specify preferences over the universe of economies. The following theorem, whose proof is deferred until Section 2.2.3, illustrates how these functions can be composed to yield a systemic risk measure, and that, in fact, *all* systemic risk measures admit such a decomposition.

Theorem 1. *1. A function* $\rho \colon \mathbb{R}^{|\mathcal{F}| \times |\Omega|} \to \mathbb{R}$ *is a systemic risk measure with image* $\operatorname{Im} \rho = \mathbb{R}$ *if and only if there exist an aggregation function* $\Lambda \colon \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ *with image* $\operatorname{Im} \Lambda = \mathbb{R}$ *and a coherent single-firm risk measure* $\rho_0 \colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$*, such that* ρ *is the composition of* ρ_0 *and* Λ*, i.e.,*

$$\rho(X) = (\rho_0 \circ \Lambda)(X) \triangleq \rho_0\left(\Lambda(X_1), \Lambda(X_2), \ldots, \Lambda(X_{|\Omega|})\right), \quad \forall\, X \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}.$$

2. *A function* $\rho\colon \mathbb{R}^{|\mathcal{F}|\times|\Omega|} \to \mathbb{R}$ *is a systemic risk measure with image* $\operatorname{Im}\rho = \mathbb{R}_+$ *if and only if there exist an aggregation function* $\Lambda\colon \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ *with image* $\operatorname{Im}\Lambda = \mathbb{R}_+$ *and a single-firm risk measure* $\rho_0\colon \mathbb{R}^{\Omega}_+ \to \mathbb{R}$*, such that* ρ *is the composition of* ρ_0 *and* Λ*, i.e.,*

$$\rho(X) = (\rho_0 \circ \Lambda)(X) \triangleq \rho_0\left(\Lambda(X_1), \Lambda(X_2), \ldots, \Lambda(X_{|\Omega|})\right), \quad \forall\, X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}.$$

Note that for a systemic risk measure ρ the positive homogeneity and the normalization conditions imply that $\operatorname{Im}\rho$ is either $\mathbb{R}$ or $\mathbb{R}_+$. Hence, the two parts in Theorem 1 state that, in all cases, the choice of a systemic risk measure is equivalent to the choice of a base risk measure and an aggregation function. Further, Theorem 1 does not guarantee the uniqueness of the base risk measure and aggregation function corresponding to a particular systemic risk measure.

As shown in Theorem 1, the key ingredient that bridges single-firm risk measures to systemic risk measures is the choice of aggregation function. An aggregation function allows us to measure the risk of aggregate positions as that of a single firm's positions. We emphasize that it is the preference consistency condition in Definition 2 that makes this structural decomposition possible. In fact, when the other conditions (including monotonicity, positive homogeneity, and convexity) are modified, a similar structural decomposition result continues to hold so long as preference consistency is satisfied. For example, in Section 2.5, we provide a structural decomposition when the convexity condition is dropped and the positive homogeneity condition is kept; similarly, a structural decomposition can also be constructed if the positive homogeneity condition is dropped and the convexity condition is kept. Simply put, the preference consistency condition connects a reasonable single-firm risk measure to a systemic version, and one has the freedom to choose other appropriate conditions for risk measure.

2.2.2 Applications

We now consider some examples to illustrate how the choice of an aggregation function and a base risk measure describes a systemic risk measure.

Example 1 (Systemic Expected Shortfall). *Consider the aggregation function*

$$\Lambda_{\mathsf{total}}(x) \triangleq \sum_{i \in \mathcal{F}} x_i = \mathbf{1}_{\mathcal{F}}^\top x, \quad \forall\, x \in \mathbb{R}^{|\mathcal{F}|}. \tag{2.1}$$

This aggregation function defines the aggregate loss of a cross-sectional profile to be the sum of the profits and losses of individual firms. Assume there is a given distribution $p \in \mathbb{R}_+^{|\Omega|}$ over the space of scenarios Ω, and define

$$\begin{aligned} \mathsf{CVaR}_\zeta(y) \triangleq \quad & \underset{q \in \mathbb{R}^{|\Omega|}}{\text{maximize}} && q^\top y \\ & \text{subject to} && \mathbf{0}_\Omega \le q \le p/\zeta, \\ & && \mathbf{1}_\Omega^\top q = 1, \end{aligned} \tag{2.2}$$

for all $y \in \mathbb{R}^{|\Omega|}$. Here, $\mathsf{CVaR}_\zeta(y)$ is the expected shortfall or conditional value at risk of the ζ-percentile of the aggregate loss vector y, where $0 < \zeta < 1$. By taking this as the base risk measure, we can define the systemic risk measure

$$\rho_{SES}(X) \triangleq \mathsf{CVaR}_\zeta\left(\mathbf{1}_{\mathcal{F}}^\top X_1, \ldots, \mathbf{1}_{\mathcal{F}}^\top X_{|\Omega|}\right), \quad \forall\, X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}.$$

This systemic risk measure is closely related[6] to the 'systemic expected shortfall' of the economy discussed by Acharya et al. [2010a]. Note that this choice of aggregation function treats losses and gains symmetrically. Further, it allows gains from one firm to cancel with losses of another firm. This approach might be undesirable if the regulator cannot subsidize the losses of some firms with the profits of others.

Example 2 (Deposit Insurance)**.** *Consider the aggregation function*[7]

$$\Lambda_{\mathsf{loss}}(x) \triangleq \sum_{i\in\mathcal{F}} x_i^+, \quad \forall\, x \in \mathbb{R}^{|\mathcal{F}|}. \tag{2.3}$$

This aggregation function considers only the losses of the firms. Assume there is a given distribution $p \in \mathbb{R}_+^{|\Omega|}$ over the space of scenarios Ω, and define the base risk measure to be the expectation

$$\mathsf{E}[y] = p^\top y, \quad \forall\, y \in \mathbb{R}^{|\Omega|}.$$

Then, we have

$$\rho_{DI}(X) = \mathsf{E}\left[\sum_{i\in\mathcal{F}} X_{i,\omega}^+\right], \quad \forall\, X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}.$$

In this example, the risk measure is the expected value of total losses only. When the expectation is taken over the risk neutral distribution, $\rho(X)$ equals the price of a 'deposit insurance' contract that pays out the losses of insolvent firms. This is similar in spirit to a number of proposed

[6]Strictly speaking, Acharya et al. [2010a] define a risk measure via preferences over the collection of *returns* of individual firms, while we express preferences over the losses, in absolute terms, experienced by individual firms. This difference is minor, however, and our risk measures could easily be defined in that setting.

[7]Given a scalar $z \in \mathbb{R}$, we define $z^+ \triangleq \max(z,0)$ to be the positive part of z, and $z^- \triangleq \max(-z,0)$ to be the negative part of z.

systemic risk measures [Lehar, 2005, Huang et al., 2009]. Note that the aggregation function Λ_{loss} treats losses and gains asymmetrically, and does not allow the gains of some firms to subsidize losses of other firms.

One feature common to both Examples 1 and 2 is that they are indifferent to *how* a large loss is spread out across firms in an economy. In particular, the aggregation functions Λ_{total} and Λ_{loss} assign the same aggregate outcome to a cross-sectional profile where one firm losses a lot of money and other firms have zero loss, or a profile where all firms lose an average amount of money. In practice, a regulator may have a preference over two such profiles. Through the design of appropriate aggregation functions, our framework is sufficiently rich to express such preferences.

Example 3 (Investing with Performance Fees). *Consider an investor with investments in a collection of hedge funds indexed by the set $\mathcal{F}$. Here, given a loss profile $x \in \mathbb{R}^{|\mathcal{F}|}$, we interpret x_i as the gross loss incurred by the investor due to the investment in the hedge fund i. Consider the following two cases: (a) the investor is a direct investor in the individual hedge funds; (b) the investor is indirectly invested in the individual hedge funds via a fund-of-funds. We assume that each hedge fund i charges a performance fee that is a fraction $\gamma_i \in [0,1]$ of the gross profits (if any) generated by the fund for the investor. The fund-of-funds charges a performance fee that is a fraction $\gamma \in [0,1]$ of the aggregate profits (if any) of the investor across all of the funds, net of performance fees paid to the individual funds. In case (a), the direct hedge fund investor can express preferences over loss profiles via the aggregation function*

$$\Lambda_{\mathsf{HF}}(x) \triangleq \sum_{i \in \mathcal{F}} \left(x_i + \gamma_i x_i^- \right), \quad \forall\, x \in \mathbb{R}^{|\mathcal{F}|}. \tag{2.4}$$

In case (b), the fund-of-funds investor can do so via the aggregation function

$$\Lambda_{\mathsf{FoF}}(x) \triangleq \sum_{i\in\mathcal{F}} \left(x_i + \gamma_i x_i^-\right) + \gamma \left(\sum_{i\in\mathcal{F}} \left(x_i + \gamma_i x_i^-\right) \right)^-, \quad \forall\, x \in \mathbb{R}^{|\mathcal{F}|}. \tag{2.5}$$

These aggregation functions consider the total profit or loss across all funds to the investor, net of all performance fees.

In the above example, we measure the systemic risk from a portfolio management viewpoint of an investor. Here, because of performance fees, losses and gains must be treated asymmetrically. Moreover, *dispersion risk* is important: holding the gross profit $\sum_{i\in\mathcal{F}} x_i$ fixed, the investor prefers to eschew profiles where the individual fund outcomes $\{x_i\}$ are dispersed, and the investor pays fees to the positively performing funds, but does not recover fees from the negatively performing funds. With modifications, more complicated performance fee structures or tax schemes imposed on profits can be captured by a similar aggregation functions. In these examples, the choice of the base risk measure is left up to the investor.

In the following examples, we illustrate systemic risk measures that are not restricted to financial applications. For complex systems with many interacting components, we can often design systemic risk measures with specialized structure appropriate for the application at hand. The examples we consider involve aggregation functions that are special cases of the following general class:

Example 4 (Optimization Aggregation Function)**.** *Given matrices* $A \in \mathbb{R}_+^{K\times|\mathcal{F}|}$, $B \in \mathbb{R}^{K\times N}$, *a vector*

$c \in \mathbb{R}^N_+$, and a convex cone $\mathcal{K} \subset \mathbb{R}^N$, define $\Lambda_{\mathsf{OPT}} : \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ by

$$\begin{aligned} \Lambda_{\mathsf{OPT}}(x) \triangleq \quad & \underset{y \in \mathcal{K}}{\text{minimize}} \quad c^\top y \\ & \text{subject to} \quad Ax \le By. \end{aligned} \tag{2.6}$$

for all loss profiles $x \in \mathbb{R}^{|\mathcal{F}|}$. If we assume that, for example,

$$\exists\, \bar{y} \in \mathcal{K} \text{ such that } B\bar{y} > \mathbf{0}, \qquad \exists\, \bar{z} \in \mathbb{R}^K_+ \text{ such that } B^\top \bar{z} = c,$$

then it is not difficult to see that the program (2.6) *is feasible and bounded for all x, and that Λ_{OPT} is monotonic, positively homogeneous, and convex. Further, if $\Lambda_{\mathsf{OPT}}(\mathbf{1}_{\mathcal{F}}) > 0$, c can be rescaled such that Λ_{OPT} is normalized as well, and thus is an aggregation function.*

Note that Λ_{OPT} captures a broad class of aggregation functions, including all of the previous examples in this section. In the following examples, we illustrate that optimization aggregation functions can be naturally applied in practical settings, including many non-financial applications.

Example 5 (Resource Allocation). *Consider a resource allocation setting, where $\mathcal{F}$ denotes a set of capacitated resources, and $\mathcal{A}$ denotes a collection of activities. Suppose activity a consumes resource $r \in \mathcal{F}$ at rate $b_{ra} \ge 0$ per unit of activity. Given a loss profile $x \in \mathbb{R}^{|\mathcal{F}|}$, we interpret each loss x_r as a shortage (or, if negative, the surplus) to the available supply of resource r, relative to* *a baseline utilized capacity. Define the aggregation function*

$$\Lambda_{\mathsf{RA}}(x) \triangleq \begin{array}{ll} \underset{u\in\mathbb{R}^{|\mathcal{A}|}}{\text{minimize}} & \sum_{a\in\mathcal{A}} c_a u_a \\ \text{subject to} & \sum_{a\in\mathcal{A}} b_{ra} u_a \geq x_r, \quad \forall\, r \in \mathcal{F}. \end{array} \tag{2.7}$$

Here, each decision variable u_a is the reduction (or, if negative, the increase) of the level of activity a. The constraint enforces the requirement that consumption of each resource r across activities be adjusted so as to accommodate the resource supply reduction x_r. The vector $c \geq \mathbf{0}$ specifies the cost associated with reductions in the level of each activity. Given a resource supply shock x, $\Lambda_{\mathsf{RA}}(x)$ measures the cost of the corresponding optimized reduction in activities. Hence, Λ_{RA} reflects the preferences of a system manager in a resource allocation setting.

Example 6 (Flow Network)**.** *Consider a network with vertices $\mathcal{V}$ and a set $\mathcal{F} \subset \mathcal{V} \times \mathcal{V}$ of directed edges. Each edge $(u,v) \in \mathcal{F}$ corresponds to a capacitated link, and the goal of the network manager is to direct maximal flow from a source $s \in \mathcal{V}$ to a destination $t \in \mathcal{V}$. Given a loss profile $x \in \mathbb{R}^{|\mathcal{F}|}$, we interpret each loss $x_{(u,v)}$ as a reduction (or, if negative, an increase) to the capacity of link $(u,v) \in \mathcal{F}$, relative to a baseline level of utilized capacity. Define the aggregation function*

$$\Lambda_{\mathsf{NF}}(x) \triangleq \begin{array}{llcll} \underset{f\in\mathbb{R}^{|\mathcal{F}|}}{\text{minimize}} & \sum_{v:\ (s,v)\in\mathcal{F}} f_{(s,v)} & & & \\ \text{subject to} & f_{(u,v)} & \geq & x_{(u,v)}, & \forall\,(u,v) \in \mathcal{F}, \\ & \sum_{u:\ (u,w)\in\mathcal{F}} f_{(u,w)} & = & \sum_{v:\ (w,v)\in\mathcal{F}} f_{(w,v)}, & \forall\, w \in \mathcal{V} \setminus \{s,t\}. \end{array} \tag{2.8}$$

Here, each decision variable $f_{(u,v)}$ represents the required reduction of flow along the link (u,v).

The first constraint enforces the requirement that flows be reduced so as to accommodate the capacity shock x, while the second constraint is a flow balance equation. Given a capacity shock x, $\Lambda_{\mathsf{NF}}(x)$ *measures the minimal necessary flow reduction, and hence reflects the preferences of a manager in a max-flow setting.*

In previous examples, we have viewed outcomes across firms and scenarios of nature as exogenously specified and did not consider structural mechanisms by which the loss of one firm can create losses at other firms, i.e., contagion. The following example illustrates that it is possible to introduce mechanisms for contagion, through the careful definition of the value function.

Example 7 (Contagion Model[8])**.** *Let* $\mathcal{F}$ *denote a collection of firms, each of whom has certain assets and obligations to each other. Let* Π_{ij} *denote the fraction of the total debt of firm i that is owed to firm j. Let* $x \in \mathbb{R}^{|\mathcal{F}|}$ *denote the loss profile in a particular scenario. Define the aggregation function*

$$\begin{aligned} \Lambda_{\mathsf{CM}}(x) \triangleq \quad & \underset{y\in\mathbb{R}_+^{|\mathcal{F}|},\ b\in\mathbb{R}_+^{|\mathcal{F}|}}{\text{minimize}} && \sum_{i\in\mathcal{F}} y_i + \gamma \sum_{i\in\mathcal{F}} b_i \\ & \text{subject to} && b_i + y_i \geq x_i + \sum_{j\in\mathcal{F}} \Pi_{ji} y_j, \quad \forall\, i \in \mathcal{F}. \end{aligned} \tag{2.9}$$

We interpret x as losses external to the obligations the firms have to each other. The loss x_i *must be covered either by firm i reducing the payments on its obligations to other firms by an amount* y_i*, or relying on an injection of external funds from the regulator in the amount* b_i*. The parameter* $\gamma > 1$ *balances the preferences of the regulator in trading off between, on the one hand, the aggregate shortfalls across the economy on inter-firm obligations* $\sum_{i\in\mathcal{F}} y_i$ *and, on the other hand, the cost* $\sum_{i\in\mathcal{F}} b_i$ *of injecting new capital to support the economy. The feasibility constraints*

[8] We thank an anonymous reviewer for suggesting this example.

reflects the fact that the reduction in payment y_i by firm i results in an additional loss of an amount $\Pi_{ij}y_i$ for firm j. The aggregation function Λ_{CM} measures the net systemic cost of the contagion. It is monotonic, positively homogeneous, and convex. Also, it can be normalized since $\Lambda_{\mathsf{CM}}(\mathbf{1}_{\mathcal{F}}) > 0$.

The constraint in (2.9) *is inspired by the structural contagion model of Eisenberg and Noe [2001]. However, in their model, the firms have limited liability, i.e., $y \leq \bar{p}$ for some vector $\bar{p} \in \mathbb{R}_+^{|\mathcal{F}|}$ of total liabilities, and there is no external injection of capital, i.e., $b \triangleq 0$. In that case, however, the aggregation function would not be positively homogeneous. Motivated by this example, an interesting possible extension of our work would be to consider such convex but not positively homogeneous systemic risk measures.*

2.2.3 Proof of Theorem 1

The two parts in this theorem have essentially one proof with minor differences. First, suppose that ρ is a systemic risk measure. For each loss profile $x \in \mathbb{R}^{|\mathcal{F}|}$, define

$$\Lambda(x) \triangleq \rho(x\mathbf{1}_\Omega^\top).$$

In other words, Λ computes the systemic risk of a constant economy represented by x. The monotonicity, convexity and positive homogeneity of Λ holds due to the monotonicity, convexity, and positive homogeneity of ρ. Also, $\Lambda\big(\mathbf{1}_{\mathcal{F}}\big) = |\mathcal{F}|$, since $\rho\big(\mathbf{1}_{\mathcal{E}}\big) = |\mathcal{F}|$. Let $Q \triangleq \mathrm{Im}(\Lambda)$ be the image of Λ. We know $|\mathcal{F}| \in Q$. By the positive homogeneity of Λ, we conclude that $\mathbb{R}_+ \subset Q$. Suppose there exists an economy X such that $\rho(X) < 0$. We can find a vector $x \in \mathbb{R}^{|\mathcal{F}|}$ such that $x\mathbf{1}_\Omega^\top \leq X$. So $\Lambda(x) = \rho(x\mathbf{1}_\Omega^\top) \leq \rho(X) < 0$. By the positive homogeneity of Λ, we conclude that $\mathbb{R}_- \subset Q$.

Thus, for part 1, $Q = \mathbb{R}$; for part 2, $Q = \mathbb{R}_+$. For each vector $z \in Q^\Omega$, define

$$\rho_0(z) \triangleq \rho(X),$$

where X is an economy that satisfies

$$\Lambda(X_\omega) = z_\omega, \quad \forall\, \omega \in \Omega.$$

First, we show ρ_0 is well-defined. Suppose two economies X, Y have that

$$\Lambda(X_\omega) = \Lambda(Y_\omega), \quad \forall\, \omega \in \Omega.$$

Since ρ has preference consistency across scenarios, we have that

$$\Lambda(X_\omega) = \rho(X_\omega \mathbf{1}_\Omega^\top) \geq \Lambda(Y_\omega) = \rho(Y_\omega \mathbf{1}_\Omega^\top), \quad \forall\, \omega \in \Omega \quad \Rightarrow \quad \rho(X) \geq \rho(Y),$$

$$\Lambda(X_\omega) = \rho(X_\omega \mathbf{1}_\Omega^\top) \leq \Lambda(Y_\omega) = \rho(Y_\omega \mathbf{1}_\Omega^\top), \quad \forall\, \omega \in \Omega \quad \Rightarrow \quad \rho(X) \leq \rho(Y).$$

Thus, we conclude that $\rho(X) = \rho(Y)$, and ρ_0 is well-defined. Clearly, ρ_0 is monotonic and positively homogeneous, from the monotonicity and positive homogeneity of Λ and ρ. We show that ρ_0 is convex. For two vectors $x, y \in Q^\Omega$, given a scalar $0 \leq \alpha \leq 1$, define $z = \alpha x + (1-\alpha) y$. Define vectors $\hat{X}, \hat{Y}, \hat{Z} \in \mathbb{R}^{|\Omega|}$ such that

$$\rho_0(x) = \rho(\hat{X}), \quad \rho_0(y) = \rho(\hat{Y}), \quad \rho_0(z) = \rho(\hat{Z}).$$

Then, for all scenarios $\omega \in \Omega$,

$$\Lambda(\hat{Z}_\omega) = z_\omega = \alpha x_\omega + (1-\alpha) y_\omega = \alpha \Lambda(\hat{X}_\omega) + (1-\alpha)\Lambda(\hat{Y}_\omega).$$

From the risk convexity of ρ, we have that

$$\rho_0(z) = \rho(\hat{Z}) \leq \alpha \rho(\hat{X}) + (1-\alpha)\rho(\hat{Y}) = \alpha \rho_0(x) + (1-\alpha)\rho_0(y).$$

This establishes the convexity of ρ_0. In addition,

$$\rho_0(\mathbf{1}_\Omega) = \rho(a\mathbf{1}_\Omega^\top),$$

where

$$\Lambda(a) = \rho(a\mathbf{1}_\Omega^\top) = 1.$$

It follows that $\rho_0(\mathbf{1}_\Omega) = 1$. For part 1, $-1 \in Q$, we can show $\rho_0(-\mathbf{1}_\Omega) = -1$ similarly. Now, we can show that for part 1, for a scalar $\alpha \in \mathbb{R}$, by the sub-additivity (as a result of convexity and positive homogeneity) of ρ_0, we have that

$$\rho_0(x + \alpha \mathbf{1}_\Omega) \geq \rho_0(x) + \alpha \rho_0(\mathbf{1}_\Omega) = \rho_0(x) + \alpha,$$

and

$$\rho_0(x - \alpha(-\mathbf{1}_\Omega)) \leq \rho_0(x) - \alpha \rho_0(-\mathbf{1}_\Omega) = \rho_0(x) + \alpha.$$

Hence, ρ_0 has the cash invariance property $\rho_0(x+\alpha\mathbf{1}_\Omega)=\rho_0(x)+\alpha$ for part 1.

To summarize, for part 1, we have shown that ρ_0 is a coherent single-firm risk measure; for part 2, we have shown that ρ_0 is a single-firm risk measure. From the definition of Λ and ρ_0, we have the structural decomposition,

$$\rho(X)=(\rho_0\circ\Lambda)(X)=\rho_0\left(\Lambda(X_1),\Lambda(X_2),\ldots,\Lambda(X_{|\Omega|})\right).$$

For the converse of the theorem, suppose that Λ is an aggregation function and ρ_0 is a base risk measure. Since Λ and ρ_0 are monotonic, convex and positively homogeneous, it is clear that ρ has the properties of monotonicity, convexity and positive homogeneity. The normalization condition is due to that of Λ and the fact that $\rho_0(\mathbf{1}_\Omega)=1$. To show the preference consistency of ρ, consider two economies $X,Y\in\mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ where, in every scenario $\omega\in\Omega$,

$$\rho\big(X_\omega\mathbf{1}_\Omega^\top\big)=(\rho_0\circ\Lambda)(X_\omega\mathbf{1}_\Omega^\top)\geq\rho\big(Y_\omega\mathbf{1}_\Omega^\top\big)=(\rho_0\circ\Lambda)(Y_\omega\mathbf{1}_\Omega^\top).$$

By the monotonicity of ρ_0, we have

$$\Lambda(X_\omega)\geq\Lambda(Y_\omega),\quad\forall\,\omega\in\Omega.$$

Then, by using the monotonicity of ρ_0 again, we conclude that

$$\rho\big(X\big)=(\rho_0\circ\Lambda)(X)\geq(\rho_0\circ\Lambda)(Y)=\rho\big(Y\big).$$

Now, we show the risk convexity of ρ. For any three economies $X, Y, Z \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ and any scalar $0 \leq \alpha \leq 1$, suppose for all scenarios $\omega \in \Omega$,

$$(\rho_0 \circ \Lambda)(Z_\omega \mathbf{1}_\Omega) = \alpha\rho(X_\omega \mathbf{1}_\Omega) + (1-\alpha)\rho(Y_\omega \mathbf{1}_\Omega) = \alpha(\rho_0 \circ \Lambda)(X_\omega \mathbf{1}_\Omega) + (1-\alpha)(\rho_0 \circ \Lambda)(Y_\omega \mathbf{1}_\Omega).$$

We know that for part 1, we have that $\rho_0(\pm\mathbf{1}_\Omega) = \pm 1$, and for part 2, we have $\rho_0(\mathbf{1}_\Omega) = 1$. Thus, we can simplify the above equation, for both part 1 and part 2, as

$$\Lambda(Z_\omega) = \alpha\Lambda(X_\omega) + (1-\alpha)\Lambda(Y_\omega), \quad \forall\, \omega \in \Omega.$$

Using the convexity of ρ_0, we conclude that, for all scenarios $\omega \in \Omega$,

$$\rho(Z) = (\rho_0 \circ \Lambda)(Z) \leq \alpha(\rho_0 \circ \Lambda)(X) + (1-\alpha)(\rho_0 \circ \Lambda)(Y) = \alpha\rho(X) + (1-\alpha)\rho(Y).$$

In addition, for part 1, there exists a vector $x \in \mathbb{R}^{|\mathcal{F}|}$ such that $\Lambda(x) < 0$. So we have $\rho(x\mathbf{1}_\Omega^\top) = \Lambda(x) < 0$ and $\rho\big(\mathbf{1}_\mathcal{E}\big) = |\mathcal{F}| > 0$. By positive homogeneity of ρ, we conclude that $\operatorname{Im}\rho = \mathbb{R}$. For part 2, $\Lambda(x) \geq 0$, for all $x \in \mathbb{R}^{|\mathcal{F}|}$. So $\rho(X) \geq \rho_0(\mathbf{0}_\mathcal{E}) = 0$, for all $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$. We also know $\rho\big(\mathbf{1}_\mathcal{E}\big) = |\mathcal{F}| > 0$. By positive homogeneity of ρ, we conclude that $\operatorname{Im}\rho = \mathbb{R}_+$.

2.3 Variational Representations

In this section, we develop two variational representations for systemic risk measures. In Section 2.3.1, we introduce a primal representation, where the systemic risk is the value of an optimization problem over a set of 'acceptable' economies. In Section 2.3.2, we develop a dual representation, where the systemic risk is the worst-case scaled expected loss across the economy. The dual representation provides a 'shadow price' to capture the systemic risk externality of the decision making of individual firms. This suggests a decentralized framework for systemic risk based decision-making that is further explored in the Online Supplement.

2.3.1 Acceptance Sets and Primal Representation

For the case of coherent single-firm risk measures, Artzner et al. [1999] describe a representation for the risk as the minimum quantity of cash that needs to be injected into each scenario such that the collection of outcomes is contained in a set of 'acceptable' outcomes. Motivated by this, we wish to construct a similar primal representation for systemic risk measures. In order to do this, we need the following definition.

Definition 4 (Acceptance Set)**.** *Consider a finite set of entities $\mathcal{N}$. An acceptance set over $\mathcal{N}$ is a set $\mathcal{S} \subset \mathbb{R} \times \mathbb{R}^{|\mathcal{N}|}$, that is a non-empty closed convex cone, and that satisfies:*

1. Monotonicity: *If $(m, x_1) \in \mathcal{S}$, $x_2 \in \mathbb{R}^{|\mathcal{N}|}$, and $x_1 \geq x_2$, then $(m, x_2) \in \mathcal{S}$.*

2. Epigraph property: *If $(m_1, x) \in \mathcal{S}$, $m_2 \in \mathbb{R}$, and $m_2 \geq m_1$, then $(m_2, x) \in \mathcal{S}$.*

We take the set of entities $\mathcal{N}$ in Definition 4 either to be the collection of firms $\mathcal{F}$ or scenarios Ω. In the former case, when $\mathcal{N} = \mathcal{F}$, $(m, x) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|}$ is contained in the acceptance set if the cross-sectional profile x is considered acceptable at a given level of 'risk exposure' m. The monotonicity property suggests that, at a fixed level of risk exposure, loss profiles that are dominated by an acceptable profile are also acceptable. The epigraph property suggests that if a loss profile is acceptable at a certain level of risk exposure, it is also acceptable at higher levels of risk exposure. Similarly, when $\mathcal{N} = \Omega$, the acceptance set captures sets of risk exposures across scenarios, in addition to an overall risk measure, such that the per scenario risk exposures are acceptable relative to the overall risk measure. The properties of acceptance sets follow from the underlying properties of aggregation functions and base risk measures; in fact, we will shortly see that acceptance sets are epigraphs of these objects. Note that, relative to the case considered by Artzner et al. [1999], we require an additional dimension corresponding to the level of risk exposure of the regulator. If a cash invariance assumption held as in Artzner et al. [1999], this extra dimension could be eliminated, but in the present context it is necessary.

The following theorem provides a primal representation to a systemic risk measure, as the value of an optimization problem over a feasible set defined through acceptable sets:

Theorem 2 (Primal Representation)**.** *Suppose* ρ *is a systemic risk measure. Then, there exist acceptance sets* $\mathcal{A} \subset \mathbb{R} \times \mathbb{R}^{|\Omega|}$ *and* $\mathcal{B} \subset \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|}$ *over scenarios and firms, respectively, such that, for all economies* $X \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$, $\rho(X)$ *can be expressed as the value of the optimization problem*

$$\begin{aligned}
\rho(X) = \quad \underset{m,\ell}{\text{minimize}} \quad & m \\
\text{subject to} \quad & (m, \ell) \in \mathcal{A}, \\
& (\ell_\omega, X_\omega) \in \mathcal{B}, \qquad \forall\, \omega \in \Omega, \\
& m \in \mathbb{R},\ \ell \in \mathbb{R}^{|\Omega|}.
\end{aligned} \tag{2.10}$$

Further, if ρ *is characterized by a base risk measure* ρ_0 *and an aggregation function* Λ*, i.e.,* $\rho =$

$\rho_0 \circ \Lambda$, *then the acceptance sets can be taken as the epigraphs of* ρ_0 *and* Λ, *i.e.,*

$$\mathcal{A} \triangleq \left\{(m,z) \in \mathbb{R} \times \mathbb{R}^{|\Omega|} \,:\, m \geq \rho_0(z)\right\}, \quad \mathcal{B} \triangleq \left\{(\ell,x) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|} \,:\, \ell \geq \Lambda(x)\right\}. \tag{2.11}$$

Proof. Given ρ, by Theorem 1 a base risk measure ρ_0 and an aggregation function Λ exist such that $\rho = \rho_0 \circ \Lambda$. Define $\mathcal{A}$ and $\mathcal{B}$ to be their epigraphs through (2.11). From the properties of ρ_0 and Λ, it is clear that these are acceptance sets.

Moreover, we have the epigraph representations

$$\begin{aligned} \rho_0(\ell) = \; & \underset{m \in \mathbb{R}}{\text{minimize}} && m \\ & \text{subject to} && (m,\ell) \in \mathcal{A}. \end{aligned} \qquad\qquad \begin{aligned} \Lambda(x) = \; & \underset{\ell \in \mathbb{R}}{\text{minimize}} && \ell \\ & \text{subject to} && (\ell,x) \in \mathcal{B}. \end{aligned}$$

for all $\ell \in \mathbb{R}^{|\Omega|}$, $x \in \mathbb{R}^{|\mathcal{F}|}$. Using the fact that $\rho = \rho_0 \circ \Lambda$, and the epigraph representation of ρ_0, we have for all $X \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$,

$$\begin{aligned} \rho(X) = \; & \underset{m \in \mathbb{R}}{\text{minimize}} && m \\ & \text{subject to} && \big(m, \Lambda(X_1), \ldots, \Lambda(X_{|\Omega|})\big) \in \mathcal{A}. \end{aligned}$$

Using the fact that $\mathcal{A}$ is monotonic, and applying the epigraph representation of Λ, the result follows. ■

The primal program (2.10) is easily interpreted: the vector of decision variables ℓ defines the regulator's minimal risk exposure in each scenario given the corresponding cross-sectional loss profile, while the scalar decision variable m is overall systemic risk given the vector ℓ of risk exposures across scenarios.

2.3.2 Dual Representation

In this section, we define a dual representation for systemic risk measures. This variational representation provides an alternative way to compute systemic risk measures and an alternative interpretation of their meaning. Moreover, it provides certain computational and operational advantages. In Section 2.4, we show that the dual representation also provides the basis of a risk attribution rule.

To begin, suppose $\rho = \rho_0 \circ \Lambda$ is a systemic risk measure. As in Theorem 2, take the epigraphs of ρ_0 and Λ as the acceptance sets $\mathcal{A}$ and $\mathcal{B}$, respectively. Define

$$\mathcal{A}^* \triangleq \left\{ (\pi_0, \hat{\pi}) \in \mathbb{R} \times \mathbb{R}^{|\Omega|} \;:\; \pi_0 m - \hat{\pi}^\top \ell \geq 0,\ \forall\ (m, \ell) \in \mathcal{A} \right\}, \tag{2.12}$$

$$\mathcal{B}^* \triangleq \left\{ (\xi_0, \hat{\xi}) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|} \;:\; \xi_0 \ell - \hat{\xi}^\top x \geq 0,\ \forall\ (\ell, x) \in \mathcal{B} \right\}. \tag{2.13}$$

Up to a sign change, $\mathcal{A}^*$ and $\mathcal{B}^*$ are the dual cones to $\mathcal{A}$ and $\mathcal{B}$, respectively. Then, the following theorem, whose full proof can be found in the Online Supplement, holds.

Theorem 3. *Suppose $\rho = \rho_0 \circ \Lambda$ is a systemic risk measure characterized by an aggregation function Λ and a base risk measure ρ_0. Then, for all economies $X \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$, $\rho(X)$ can be expressed* *as the value of the optimization problem*

$$\begin{aligned}
\rho(X) = \quad \underset{\bar{\pi},\Xi}{\text{maximize}} \quad & \sum_{i\in\mathcal{F}}\sum_{\omega\in\Omega} \Xi_{i,\omega}X_{i,\omega} \\
\text{subject to} \quad & (1,\bar{\pi}) \in \mathcal{A}^*, \\
& (\bar{\pi}_\omega, \Xi_\omega) \in \mathcal{B}^*, \qquad \forall\, \omega \in \Omega, \\
& \bar{\pi} \in \mathbb{R}^{|\Omega|},\ \Xi \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}.
\end{aligned} \tag{2.14}$$

In addition, feasible points $(\bar{\pi},\Xi)$ *for this problem must satisfy*

$$\bar{\pi} \geq \mathbf{0}_\Omega, \quad \mathbf{1}_\Omega^\top \bar{\pi} \leq 1, \quad \Xi \geq \mathbf{0}_{\mathcal{E}}, \quad \mathbf{1}_{\mathcal{F}}^\top \Xi \leq |\mathcal{F}|\bar{\pi}^\top. \tag{2.15}$$

In order to interpret the dual problem (2.14), observe that (2.15) implies that, for feasible $(\bar{\pi},\Xi)$, $\bar{\pi}$ is a sub-stochastic vector. This can be interpreted as a probability distribution over the augmented set of scenarios $\Omega\cup\{\omega_0\}$, where ω_0 is an additional, artificial scenario in which every firm has a 0 outcome. Define the matrix $\hat{\Xi} \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ by

$$\hat{\Xi}_{i,\omega} \triangleq \begin{cases} \Xi_{i,\omega}/\pi_\omega & \text{if } \pi_\omega > 0, \\ 0 & \text{otherwise,} \end{cases}$$

for each firm i and scenario ω. Then, the objective in (2.14) becomes

$$\sum_{\omega\in\Omega} \bar{\pi}_\omega \sum_{i\in\mathcal{F}} \hat{\Xi}_{i,\omega} X_{i,\omega},$$

where (2.15) implies that

$$\hat{\Xi} \geq \mathbf{0}_{\mathcal{E}}, \quad \sum_{i \in \mathcal{F}} \hat{\Xi}_{i,\omega} \leq |\mathcal{F}|, \; \forall\, \omega \in \Omega.$$

In other words, the dual objective is the *worst-case expected loss*, over some set of feasible probability distributions $\hat{\pi}$ and scaling functions $\hat{\Xi}$, of a scaled economy in which the participation of the $|\mathcal{F}|$ firms in the economy in scenario ω is rescaled according to the vector $\hat{\Xi}_{1,\omega}, \ldots, \hat{\Xi}_{|\mathcal{F}|,\omega}$. This is analogous to the robust interpretation of a single-firm coherent risk measure as a worst-case expected loss.

The following is an immediate corollary of Theorem 3:

Corollary 1. *Suppose that* ρ *is a systemic risk measure with dual representation* (2.14). *Given an economy* X*, if* $(\bar{\pi}^*, \Xi^*)$ *is a dual optimal solution, then* Ξ^* *is a subgradient of* ρ *at* X*.*

Corollary 1 suggests another interpretation of the optimal dual solution Ξ^* for an economy X. The quantity $\Xi^*_{i,\omega}$ is the minimal marginal increase in systemic risk as function of a marginal increase in the losses of firm i in scenario ω. In other words, $\Xi^*_{i,\omega}$ captures the externalities imposed by the decision-making of a firm on the system regulator, and hence is a *shadow price for systemic risk*. Note that these shadow prices can vary both by scenario — incremental losses in some scenarios may have a much larger impact than in other scenarios — and by the identity of the firm. These shadow prices could be used to coordinate decision-making by individual firms with the goals of the regulator. For example, it is possible to design tax schemes based on these prices, in the spirit of Acharya et al. [2010a], such that individual firms optimize their portfolios to distribute profits and losses across scenarios in a way that is aligned with the concerns of the regulator. This topic is further explored in the Online Supplement.

The dual optimization problem (2.14) may also lead to useful decentralized schemes for computing systemic risk. Here, a centralized regulator can seek to choose optimal values for the dual variables $(\bar{\pi}, \Xi)$, while relying on constituent firms to compute their individual weighted profits and losses, scaled according to each putative choice of dual variables. The utility of the dual representation from analytical, operational, and computational perspectives is illustrated by the following examples.

Example 8 (Total P&L)**.** *Consider the total profit and loss aggregation function* Λ_{total}*, defined by* (2.1)*. For this aggregation function, it is easy to see that*

$$\mathcal{B}^* = \left\{ (\xi_0, \xi_0 \mathbf{1}_{\mathcal{F}}) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|} \;:\; \xi_0 \in \mathbb{R}_+ \right\}.$$

Then, the dual representation (2.14) *takes the simplified form*

$$\begin{aligned} \rho(X) = \quad & \underset{\bar{\pi} \in \mathbb{R}^{|\Omega|}}{\text{maximize}} \quad \sum_{i \in \mathcal{F}} \sum_{\omega \in \Omega} \bar{\pi}_\omega X_{i,\omega} \\ & \text{subject to} \quad (1, \bar{\pi}) \in \mathcal{A}^*. \end{aligned}$$

In Example 8, the base risk measure (and thus the constraint set $\mathcal{A}^*$) has not been specified. However, independent of this choice, given an optimal dual solution $\bar{\pi}^*$, a shadow price for systemic risk for each firm i in a scenario ω is given by $\bar{\pi}^*_\omega$ and is independent of the identity of the firm. This is consistent with the choice of aggregation function: the impact of a marginal increase in the loss of any firm is the same, since the sum total of all profits and losses is of concern.

Example 9 (Total Loss)**.** *Consider the total loss aggregation function* Λ_{loss}*, defined by* (2.3)*. Then,* *we have that*

$$\mathcal{B}^* = \left\{ \left(\xi_0, \hat{\xi}\right) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|} \;:\; \mathbf{0}_{\mathcal{F}} \leq \hat{\xi} \leq \xi_0 \mathbf{1}_{\mathcal{F}} \right\}.$$

Thus, the dual representation (2.14) *takes the simplified form*

$$\begin{aligned} \rho(X) = \quad & \underset{\bar{\pi} \in \mathbb{R}^{|\Omega|}}{\text{maximize}} \quad \sum_{i \in \mathcal{F}} \sum_{\omega \in \Omega} \bar{\pi}_\omega \left(X_{i,\omega}\right)^+ \\ & \text{subject to} \quad (1, \bar{\pi}) \in \mathcal{A}^*. \end{aligned}$$

In Example 9, given an optimal dual solution $\bar{\pi}^*$, a shadow price for systemic risk for firm i in scenario ω is given by $\bar{\pi}^*_\omega$ if $X_{i,\omega} \geq 0$, and is 0 otherwise. This is because a firm can only marginally impact the systemic risk in scenarios where it is not profitable.

Example 10 (CVaR)**.** *Suppose, given* $0 < \zeta < 1$*, the aggregation function is taken to be*

$$\Lambda_{\mathsf{CVaR}}(x) \triangleq \inf_{\ell \in \mathbb{R}} \; |\mathcal{F}|\ell + \tfrac{1}{\zeta} \sum_{i \in \mathcal{F}} (x_i - \ell)^+.$$

This corresponds to the aggregate total profits and losses of the worst ζ*-percentile of firms in the cross-sectional profile, i.e., it is analogous to the* CVaR_ζ *risk measure of* (2.2)*, but taken across firms rather than scenarios. Then, we have that*

$$\mathcal{B}^* = \left\{ \left(\xi_0, \hat{\xi}\right) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{F}|} \;:\; \mathbf{0}_{\mathcal{F}} \leq \hat{\xi} \leq \tfrac{\xi_0}{\zeta} \mathbf{1}_{\mathcal{F}},\; \mathbf{1}_{\mathcal{F}}^\top \hat{\xi} = \xi_0 |\mathcal{F}|. \right\}.$$

Thus, the dual representation (2.14) *takes the simplified form*

$$\begin{aligned} \rho(X) = \quad \underset{\bar{\pi},\Xi}{\text{maximize}} \quad & \sum_{i\in\mathcal{F}} \sum_{\omega\in\Omega} \Xi_{i,\omega} X_{i,\omega} \\ \text{subject to} \quad & (1,\bar{\pi}) \in \mathcal{A}^*, \\ & \mathbf{0}_{\mathcal{E}} \le \Xi \le \tfrac{1}{\zeta}\bar{\pi}\mathbf{1}_{\Omega}^{\top}, \\ & \mathbf{1}_{\mathcal{F}}^{\top}\Xi = |\mathcal{F}|\bar{\pi}, \\ & \bar{\pi} \in \mathbb{R}^{|\Omega|},\ \Xi \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}. \end{aligned}$$

In Example 10, a shadow price for systemic risk $\Xi^*_{i,\omega}$, in general, depends both on the identity of the firm i and the particular scenario ω.

2.4 Risk Attribution

In this section, we consider the problem of attributing or allocating the systemic risk across the firms that compose the economy. The spirit here is to identify systemically risky institutions, and quantify their overall impact on the risk in the economy. Motivated by the discussion of shadow prices in Section 2.3.2, consider the following definition:

Definition 5 (Risk Attribution)**.** *Suppose ρ is a systemic risk measure, with dual decomposition* (2.14)*. For each economy $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$, define $\mathcal{M}(X) \subset \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ to be the set of dual optimal solutions for $\rho(X)$. Given a shadow price for systemic risk $\Xi^* \in \mathcal{M}(X)$, we define a vector*

$y^*(X, \Xi^*) \in \mathbb{R}^{|\mathcal{F}|}$, *with component*

$$y_i^*(X, \Xi^*) \triangleq \sum_{\omega \in \Omega} \Xi_{i,\omega}^* X_{i,\omega}$$

as the systemic risk attributable to firm i. We define set of all attribution vectors

$$\mathcal{Y} \triangleq \left\{ y^*(X, \Xi^*) \colon \Xi^* \in \mathcal{M}(X) \right\}.$$

Note that the attribution rule is unique if the dual optimal solution for $\rho(X)$ is unique at X.

This definition allocates systemic risk to each firm according its entire profile of profits and losses across scenarios, where each profit or loss is valued according to the appropriate shadow price for systemic risk. Note that the risk allocation is an immediate by-product of the dual representation, and hence requires no computation if the dual solution is available.

The allocation of Definition 5 has a number of desirable properties. First, since Theorem 3 guarantees that the dual optimum equals $\rho(X)$, it is immediate that

$$\rho(X) = \sum_{i \in \mathcal{F}} y_i^*, \quad \forall\, y^* \in \mathcal{Y}$$

In order words, the individual risk attributions add up to the total systemic risk. Second, following Corollary 1, the sensitivity of the attribution y_i of firm i to a change in the loss $X_{i,\omega}$ in some scenario ω is precisely the shadow price for systemic risk. Hence, the local incentives created by this allocation are aligned with the systemic risk objective. Finally, the risk attribution that we propose has the following fairness property:

Theorem 4. *Fix a systemic risk measure* ρ. *Let* $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ *denote a given economy.*

For a vector $\alpha \in \mathbb{R}_{+}^{\mathcal{F}}$, *define* $r(\alpha)$ *to be the systemic risk associated with an economy* $\alpha * X$ *that has outcomes for firm i in scenario* ω *given by* $\alpha_i X_{i,\omega}$. *Then, for any risk attribution* $y^* \in \mathcal{Y}$, $\alpha^\top y^* \leq r(\alpha)$.

Proof. From the dual representation in Theorem 3, we have that

$$\begin{aligned} r(\alpha) = \quad & \underset{\bar{\pi},\Xi}{\text{maximize}} \quad \sum_{i\in\mathcal{F}} \alpha_i \sum_{\omega\in\Omega} \Xi_{i,\omega} X_{i,\omega} \\ & \text{subject to} \quad (1,\bar{\pi}) \in \mathcal{A}^*, \\ & \qquad\qquad\quad (\bar{\pi}_\omega, \Xi_\omega) \in \mathcal{B}^*, \qquad \forall\, \omega \in \Omega, \\ & \qquad\qquad\quad \bar{\pi} \in \mathbb{R}^{|\Omega|},\ \Xi \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}. \end{aligned} \tag{2.16}$$

Suppose y^* is obtained by a dual optimal solution Ξ^*. Since Ξ^* is a feasible solution of the dual representation of $\rho(X)$ in (2.14), Ξ^* is also a feasible solution of the dual representation of $r(\alpha)$ in (2.16), for any α. The objective value achieved by Ξ^* in (2.16) is $\alpha^\top y^*$, which can be no greater than the optimal value $r(\alpha)$, i.e., $\alpha^\top y^* \leq r(\alpha)$. ■

Theorem 4 is a 'no-undercut' result, in the spirit of Denault [2001]. Here, $r(\alpha)$ is the systemic risk associated with an economy where each firm i participates proportionally to the factor $\alpha_i \geq 0$. The result states that, if a fractional coalition of firms specified by the vector α form a new economy $\alpha * X$, the systemic risk of that economy $r(\alpha) \triangleq \rho(\alpha * X)$ is at least as large as the weighted sum of risk attributed to the firms in the original economy. Thus, the risk attribution is fair: the risk attributed to any fractional coalition is no greater than it would incur as a standalone economy.

The risk attribution we propose is closely related to prices of Aumann and Shapley [1974] for allocating the cost in a fractional coalition game. If we assume that r is differentiable at the point $\alpha = \mathbf{1}_{\mathcal{F}}$, then by positive homogeneity, it is differentiable on the ray $\{t\mathbf{1}_{\mathcal{F}} \ : \ t \geq 0\}$. The Aumann-Shapley prices are then defined by

$$y^{\mathsf{AS}} \triangleq \int_0^1 \nabla r\big(t\mathbf{1}_{\mathcal{F}}\big)\, dt = \nabla r\big(\mathbf{1}_{\mathcal{F}}\big).$$

The last equality follows from the fact that r is positively homogeneous, and is sometimes referred to as the Euler allocation rule or gradient allocation rule [Denault, 2001, Buch and Dorfleitner, 2008]. In fact, when these latter attribution rules are well-defined, they correspond with our notion of risk allocation:

Theorem 5. *Given a systemic risk measure ρ and an economy $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$, if ρ is differentiable at X, then the risk attribution y^* is unique and coincides with the Aumann-Shapley prices y^{AS}.*

Proof. Under the hypothesis, $r(\alpha)$ defined in Theorem 4 is differentiable at $\alpha = \mathbf{1}_{\mathcal{F}}$. Observing that the constraint set in (2.16) is compact, we can apply Danskin's theorem [Bertsekas, 1999] to (2.16) to obtain

$$\frac{\partial r(\alpha)}{\partial \alpha_i} = \sum_{\omega \in \Omega} \Xi^*_{i,\omega} X_{i,\omega},$$

for all firms i , where $\Xi^* \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$ is the unique dual optimal solution for $\rho(X)$. Therefore, $y^{AS} = y^*$. ■

Note that the work of Denault [2001], Fischer [2003], and Buch and Dorfleitner [2008] suggests the gradient allocation rule for risk attribution in a portfolio setting; that setting is a special case of systemic risk measure corresponding to the total P&L aggregation function Λ_{total} of (2.1). In that case, the gradient allocation rule is identical to our dual risk attribution y^*. Our dual risk attribution y^*, however, requires no differentiability assumption, and can apply to a more general class of aggregation functions. When ρ is not differentiable at X, several attribution rules y^* are possible and how to choose among them may require further investigation. Related discussion on risk attribution can be found in the work of Tsanakas [2009] and Cherny and Orlov [2011].

2.5 Homogeneous Systemic Risk Measures

In this section, we extend our analysis to value-at-risk-like measures that are monotonic, positively homogeneous but non-convex. Value-at-risk (VaR) [see, e.g., Jorion, 2006] is a single-firm risk measure of particular importance, because it is extensively used in the practice of financial risk management. Originally developed by practitioners in the financial industry, it is widely employed both by firm managers and regulators to compute and manage market risk — in fact, VaR is the preferred measurement of market risk of the Basel II regulatory regime. The popularity of VaR as a single-firm risk measure has motivated a number of VaR-based measures of systemic risk, such as the CoVaR measure proposed by Adrian and Brunnermeier [2009].

The VaR at a confidence level $\zeta \in (0,1)$ is defined as follows: suppose $x \in \mathbb{R}^{|\Omega|}$ is a vector of losses across scenarios Ω, and that $p \in \mathbb{R}_+^{|\Omega|}$ with $\mathbf{1}_\Omega^\top p = 1$ is a probability distribution over Ω. Then, the VaR of the random loss x is the minimum loss threshold value ℓ such that the probability of the loss exceeding ℓ is at most $1-\zeta$, i.e.,

$$\mathsf{VaR}_\zeta(x) \triangleq \inf \left\{ \ell \in \mathbb{R} : \sum_{\omega \in \Omega:\ x_\omega > \ell} p_\omega \leq 1 - \zeta \right\}. \tag{2.17}$$

From the definition of VaR, it is clear that this risk measure is positively homogeneous, monotonic, normalized, and cash invariant, i.e., it satisfies conditions 1, 2, 4, and 1 of Definition 1. However, it is *not* convex.

The lack of convexity is the principal difference between homogeneous risk measures and the (convex) risk measures defined in Section 2.2. Aside from their wide use in practice, homogeneous risk measures have also generated some interest in the literature [e.g., Kou et al., 2009, Cerreia-Vioglio et al., 2010]. Our goal here is to illustrate the impact of the absence of this axiom on our framework for systemic risk.

In what follows, we investigate the impact of dropping the convexity requirement for systemic risk measures. In Section 2.5.1, we give a complete structural decomposition for homogeneous risk measures. The benefit from this analysis is two-fold. First, as in the case with convex systemic risk measures discussed in the previous sections, the characterization gives us a rule for constructing homogeneous systemic risk measures from a homogeneous base risk measure and homogeneous aggregation functions. The second and equally important benefit is that the characterization elucidates the implicit assumptions that are being made when one combines single-firm homogeneous risk measures to create a systemic risk measure. As a by-product of our characterization, in Section 2.5.2 we show that a homogeneous systemic measures have a convex representation. Finally, in Section 2.5.4 we consider some examples of homogeneous systemic risk measures.

2.5.1 Structural Decomposition

Motivated by the discussion above, we define homogeneous systemic risk measures as follows:

Definition 6 (Homogeneous Systemic Risk Measure). *A homogeneous systemic risk measure is a function* $\rho\colon \mathbb{R}^{|\mathcal{F}|\times|\Omega|} \to \mathbb{R}$ *that satisfies the following conditions, for all economies* $X, Y \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$*:*

1. Monotonicity: $X \geq Y$ *implies* $\rho(X) \geq \rho(Y)$.

2. Positive homogeneity*: For all* $\alpha \geq 0$ *and* $\rho(\alpha X) = \alpha\rho(X)$.

3. Preference consistency: $\rho\big(X_\omega \mathbf{1}_\Omega^\top\big) \geq \rho\big(Y_\omega \mathbf{1}_\Omega^\top\big)$, $\forall\, \omega \in \Omega$*, implies* $\rho\big(X\big) \geq \rho\big(Y\big)$.

4. Normalization: $\rho\big(\mathbf{1}_\mathcal{E}\big) = |\mathcal{F}|$.

We define homogeneous single-firm risk measures and homogeneous aggregation functions as follows:

Definition 7 (Homogeneous Single-Firm Risk Measure). *A homogeneous single-firm risk measure is a function*[9] $\rho\colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ *that, for all loss vectors* $x, y \in \mathbb{R}^{|\Omega|}$ *of a single firm, satisfies the following conditions:*

1. Monotonicity: *If* $x \geq y$*, then* $\rho(x) \geq \rho(y)$.

2. Positive homogeneity: *For all non-negative scalars* $\alpha \geq 0$, $\rho(\alpha x) = \alpha\rho(x)$.

3. Normalization: $\rho\big(\mathbf{1}_\Omega\big) = 1$.

Definition 8 (Homogeneous Aggregation Function). *A function* $\Lambda\colon \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ *over cross-sectional loss profiles of firms is a homogeneous aggregation function if, for all cross-sectional loss profiles* $x, y \in \mathbb{R}^{|\mathcal{F}|}$*, it satisfies:*

[9] As was the case with Definition 1, we sometimes consider homogeneous single-firm risk measures $\rho\colon \mathbb{R}_+^{|\Omega|} \to \mathbb{R}$ defined only on the positive orthant. In that case, we assume that conditions 1–2 are satisfied for all $x, y \in \mathbb{R}_+^{|\Omega|}$.

1. Monotonicity: *If* $x \geq y$*, then* $\Lambda(x) \geq \Lambda(y)$*.*

2. Positive homogeneity: *For all* $\alpha \geq 0$*,* $\Lambda(\alpha x) = \alpha\Lambda(x)$*.*

3. Normalization: $\Lambda\big(\mathbf{1}_{\mathcal{F}}\big) = |\mathcal{F}|$.

The (convex) single-firm risk measures and the (convex) aggregation functions defined earlier in Definitions 1 and 3 are in fact also homogeneous single measures and aggregation functions, respectively; in addition, those functions are also convex.

homogeneous systemic risk measures admit a structural decomposition analogous to that of Theorem 1, as follows:

Theorem 6. *1. A function* $\rho\colon \mathbb{R}^{|\mathcal{F}|\times|\Omega|} \to \mathbb{R}$ *is a homogeneous systemic risk measure with image* $\operatorname{Im}\rho = \mathbb{R}$ *if and only if there exist a homogeneous aggregation function* $\Lambda\colon \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ *and a homogeneous single-firm risk measure* $\rho_0\colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ *with* $\rho_0(\pm\mathbf{1}_\Omega) = \pm 1$ *such that* ρ *is the composition of* ρ_0 *and* Λ*, i.e.,*

$$\rho(X) = (\rho_0 \circ \Lambda)(X) \triangleq \rho_0\left(\Lambda(X_1), \Lambda(X_2), \ldots, \Lambda(X_{|\Omega|})\right).$$

2. A function $\rho\colon \mathbb{R}^{|\mathcal{F}|\times|\Omega|} \to \mathbb{R}$ *is a homogeneous systemic risk measure with image* $\operatorname{Im}\rho = \mathbb{R}_+$ *if and only if there exist a homogeneous aggregation function* $\Lambda\colon \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}$ *with* $\operatorname{Im}\Lambda = \mathbb{R}_+$ *and a homogeneous single-firm risk measure* $\rho_0\colon \mathbb{R}_+^{|\Omega|} \to \mathbb{R}$ *such that* ρ *is the composition of* ρ_0 *and* Λ*, i.e.,*

$$\rho(X) = (\rho_0 \circ \Lambda)(X) \triangleq \rho_0\left(\Lambda(X_1), \Lambda(X_2), \ldots, \Lambda(X_{|\Omega|})\right).$$

Proof. The proof is a simplified version of the proof of Theorem 1, since, in this case, it is not necessary to establish convexity. It is thus omitted. ■

Observe that, comparing Theorem 1 and Theorem 6, preference consistency condition 3 is key to establishing this structural decomposition. The other conditions, namely homogeneity and monotonicity, imply these same properties for the single-firm risk measure ρ_0 and the aggregation function Λ.

2.5.2 Convex Representation

In this section, we develop a convex representation for homogeneous (non-convex) systemic risk measures as the pointwise minima of a collection of convex risk functions. To begin, consider the following lemma, a proof of which that follows the argument of Castellani [2000] is provided in the Online Supplement:

Lemma 1. *A function $g\colon \mathbb{R}^n \to \mathbb{R}$ is positively homogeneous and monotonic if and only if there exists an index set $\mathcal{S}$ where, for each $s \in \mathcal{S}$, $g^{(s)}\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a positively homogeneous, monotonic, and convex extended real-valued function, such that*

$$g(x) = \underset{s\in\mathcal{S}}{\text{minimize}}\ g^{(s)}(x), \quad \forall\, x \in \mathbb{R}^n.$$

The following is a corollary of Lemma 1. It establishes a representation for homogeneous systemic risk measures in terms of *convex* single-firm risk measures and aggregation functions.

Corollary 2. *Suppose ρ is a homogeneous systemic risk measure with* $\operatorname{Im}\rho = \mathbb{R}$. *For all economies*

$X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$, $\rho(X)$ can be expressed as the value of the optimization problem

$$\rho(X) = \underset{t\in\mathcal{T},\ s_1\in\mathcal{S},\ \ldots,\ s_{|\Omega|}\in\mathcal{S}}{\text{minimize}}\ \rho^{(t)}\big(\Lambda^{(s_1)}(X_1),\ldots,\Lambda^{(s_{|\Omega|})}(X_{|\Omega|})\big). \tag{2.18}$$

Here, $\mathcal{S}$ and $\mathcal{T}$ are index sets. For each $s \in \mathcal{S}$, the aggregation function $\Lambda^{(s)} : \mathbb{R}^{|\mathcal{F}|} \to \mathbb{R}\cup\{\infty\}$ satisfies conditions 1–3 of Definition 3 (i.e., monotonicity, positive homogeneity, convexity). For each $t \in \mathcal{T}$, the single-firm risk measure $\rho_0^{(t)} : \mathbb{R}^{|\Omega|} \to \mathbb{R}\cup\{\infty\}$ satisfies conditions 1–3 of Definition 1 (i.e., monotonicity, positive homogeneity, convexity).

Proof. The result follows by first applying Theorem 6 to obtain the representation $\rho = \rho_0 \circ \Lambda$ in terms of a homogeneous single-firm base risk measure ρ_0 and a homogeneous aggregation function Λ, and then applying Lemma 1 to ρ_0 and Λ. ■

As an example of this construction, consider the following:

Example 11 (VaR)**.** *Define the single-firm base measure $\rho_0(x) = \mathsf{VaR}_\zeta(x)$ to be the value-at-risk defined in (2.17), given a uniform probability distribution $p \triangleq \mathbf{1}_\Omega/|\Omega|$ and a confidence level $\zeta \in (0,1)$. In this case, we have that $\rho_0(x) = x_{[k^*]}$, where $k^* \triangleq \lceil\zeta|\Omega|\rceil$ and, for each $1 \le k \le |\Omega|$, $x_{[k]}$ is the kth order statistic of the vector x. Define $q^{(k)} \in \mathbb{R}^{|\Omega|}$ to be a vector with the first $|\Omega| - k$ components equal to 1, and the rest equal to 0, and define $\mathcal{T} \subset \mathbb{R}^{|\Omega|\times|\Omega|}$ to be the set of $|\Omega|\times|\Omega|$ permutation matrices. Then, we have that*

$$\rho_0(x) = \sum_{k=k^*}^{|\Omega|} x_{[k]} - \sum_{k=k^*+1}^{|\Omega|} x_{[k]} = \max_{P\in\mathcal{T}} x^\top P q^{(k^*)} - \max_{P\in\mathcal{T}} x^\top P q^{(k^*+1)} = \min_{P\in\mathcal{T}} \rho^{(P)}(x),$$

where, for each $P \in \mathcal{T}$,

$$\rho^{(P)}(x) \triangleq \max_{Q \in \mathcal{T}} x^\top \left(Qq^{(k^*)} - Pq^{(k^*+1)} \right)$$

is a convex base risk measure.

2.5.3 Risk Attribution

Next, we discuss the issue of attributing the total risk $\rho(X)$ across the $|\mathcal{F}|$ firms in the economy. We show below that a good risk attribution rule exists for a subset of homogeneous systemic risk measures that includes VaR.

For a systemic risk measure ρ, an economy X, and a vector $\alpha \in \mathbb{R}_+^{|\mathcal{F}|}$, define $r_\rho(\alpha) \triangleq \rho(\alpha * X)$, where $\alpha * X$ is defined by setting the outcomes for firm i in scenario ω to $\alpha_i X_{i,\omega}$. In this setting, a *risk allocation function* $\Psi^{|\mathcal{F}|}$ takes as inputs two arguments, namely the function $r_\rho \colon \mathbb{R}_+^{|\mathcal{F}|} \to \mathbb{R}$ and a vector of fractional participation $\alpha \in \mathbb{R}_+^{|\mathcal{F}|}$, and returns a risk allocation to each firm. The following are certain desirable properties for Ψ that are typically assumed in the literature [e.g., Billera and Heath, 1982]. Suppose we have any two systemic risk measures ρ_1 and ρ_2.

1. *Full risk allocation:* $\alpha^\top \Psi^{|\mathcal{F}|}(r_\rho, \alpha) = r_\rho(\alpha)$.

2. *Additivity:* $\Psi^{|\mathcal{F}|}(r_{\rho_1}, \alpha) + \Psi^{|\mathcal{F}|}(r_{\rho_2}, \alpha) = \Psi^{|\mathcal{F}|}(r_{\rho_1} + r_{\rho_2}, \alpha)$.

3. *Monotonicity:* If $r_{\rho_1}(\alpha) \geq r_{\rho_2}(\alpha)$, for all $\alpha \in \mathbb{R}_+^{|\mathcal{F}|}$, then

$$\Psi^{|\mathcal{F}|}(r_{\rho_1}, \alpha) \geq \Psi^{|\mathcal{F}|}(r_{\rho_2}, \alpha), \quad \forall\, \alpha \in \mathbb{R}_+^{|\mathcal{F}|}.$$

4. *Rescaling invariance:* For a vector $\beta \in \mathbb{R}_+^{|\mathcal{F}|}$, define $\beta * \alpha \triangleq (\beta_1\alpha_1, \beta_2\alpha_2, \ldots, \beta_{|\mathcal{F}|}\alpha_{|\mathcal{F}|})$. If

$r_{\rho_1}(\alpha) = r_{\rho_2}(\beta * \alpha)$, then

$$\Psi_i^{|\mathcal{F}|}(r_{\rho_1}, \beta^{-1} * \alpha) = \beta_i \Psi_i^{|\mathcal{F}|}(r_{\rho_2}, \alpha), \quad \forall\, i \in \mathcal{F}.$$

5. *Consistency:* If there is a function $\bar{r}$ such that $r(\alpha) = \bar{r}(\mathbf{1}_{\mathcal{F}}^\top \alpha)$, then

$$\Psi_i^{|\mathcal{F}|}(r, \alpha) = \Psi^1(\bar{r}, \mathbf{1}_{\mathcal{F}}^\top \alpha), \quad \forall\, i \in \mathcal{F}.$$

Here, Ψ^1 is a single-firm cost allocation function.

Note that for the general class of differentiable risk functions r, Billera and Heath [1982] show that Aumann-Shapley prices are the only attribution rule that satisfies all five properties. For piecewise linear risk functions, Haimanko [2001] shows that the Mertens [1988] mechanism is the unique cost allocation scheme that satisfies the five properties. Specifically, the Mertens mechanism Ψ_M is given by

$$\Psi_{M,i}^{|\mathcal{F}|}(r, \alpha) \triangleq \mathsf{E}\left[\int_0^1 D_{\alpha * C, e_i} r(t\alpha)\, dt\right], \tag{2.19}$$

where $e_i \in \mathbb{R}^{|\mathcal{F}|}$ is the ith unit vector, the expectation is taken over a random vector $C \in \mathbb{R}^{|\mathcal{F}|}$ of independent standard Cauchy random variables, and

$$D_u r(\alpha) \triangleq \lim_{\varepsilon \to 0^+} \frac{r(\alpha + \varepsilon u) - r(\alpha)}{\varepsilon}, \qquad D_{u,v} r(\alpha) \triangleq \lim_{\varepsilon \to 0^+} \frac{D_{u+\varepsilon v} r(\alpha) - D_u r(\alpha)}{\varepsilon}, \tag{2.20}$$

are directional derivatives of $r(\alpha)$, given directions $u, v \in \mathbb{R}^{|\mathcal{F}|}$.

We propose the Mertens mechanism for the attribution rules in the context of piecewise linear systemic risk measures. The following lemma gives a sufficient condition for a homogeneous systemic risk measures ρ to be piecewise linear. Recall that ρ is generated by index sets $\mathcal{S}$ and $\mathcal{T}$ in Corollary 2.

Lemma 2. *Suppose index sets $\mathcal{S}$ and $\mathcal{T}$ are finite, $\Lambda^{(s)}$ is a piecewise linear continuous function for all $s \in \mathcal{S}$, and $\rho^{(t)}$ is a piecewise linear continuous function for all $t \in \mathcal{T}$. Then the homogeneous systemic risk measure ρ generated by index sets $\mathcal{S}$ and $\mathcal{T}$ is piecewise linear and continuous.*

Proof. It is immediate from the representation in Corollary 2. ∎

Suppose the hypothesis of Lemma 2 is satisfied. Then, $r_\rho(\alpha)$ is clearly piecewise linear. Thus the Mertens mechanism can be used for risk attribution. Note that this attribution scheme does not have the 'no-undercut' property introduced in Theorem 4. Further, risk attribution for general non-differentiable systemic risk measures beyond this piecewise linear class is an area for future investigation.

2.5.4 Examples

In this section, we describe examples of homogeneous systemic risk measures.

Example 12 (VaR). *Consider the aggregation function $\Lambda_{\mathsf{total}}(x) \triangleq \mathbf{1}_{\mathcal{F}}^\top x$ of* (2.1)*, i.e., the total profit and loss across all firms. Given a probability distribution p over the scenarios Ω and a confidence level ζ, consider the value-at-risk function VaR_ζ of* (2.17) *as a base risk measure. Then the homogeneous systemic risk measure*

$$\rho_{\mathsf{VaR}}(X) \triangleq (\mathsf{VaR}_\zeta \circ \Lambda_{\mathsf{total}})(X) = \mathsf{VaR}_\zeta\left(\mathbf{1}_{\mathcal{F}}^\top X_1, \ldots, \mathbf{1}_{\mathcal{F}}^\top X_{|\Omega|}\right),$$

is simply the value-at-risk of the aggregated outcomes. Note that Λ_{total} *and* VaR_ζ *are piecewise linear and continuous.*

When the probability measure p over the scenarios is arbitrary, we can consider conditional *probability measures of the form* $\mathbb{P}(\omega|A)$ *where* $A \subset \Omega$ *denotes a set of stress scenarios. This* conditional *variation of the value-at-risk is in the same spirit as the CoVaR measure defined by Adrian and Brunnermeier [2009]. Unlike them, we only condition on subsets of the exogenously defined scenarios. This restriction is required in order to be able compare different sets of outcomes on the same set of scenarios.*

In this example, we can also illustrate the Mertens mechanism for risk attribution. For ease of exposition, assume that the probability distribution over scenarios is given by the uniform distribution, i.e., $p \triangleq \mathbf{1}_\Omega/|\Omega|$*. As in Example 11, the risk* $\rho_{\mathsf{VaR}}(X)$ *takes the value* $(1_{\mathcal{F}}^\top X)_{[k^*]}$*, i.e., the* k^**th order statistic of the aggregated losses across all firms. Now, assume that this value is achieved by a unique scenario* ω^**, so that* $\rho_{\mathsf{VaR}}(X) = \mathbf{1}_{\mathcal{F}}^\top X_{\omega^*}$*. Then, in fact, for all* $t > 0$*,* $r_{\mathsf{VaR}}(t\mathbf{1}_{\mathcal{F}}) = \rho_{\mathsf{VaR}}(tX) = t\mathbf{1}_{\mathcal{F}}^\top X_{\omega^*}$*. Also, in this case,* r_{VaR} *will be a linear function in the neighborhood of each* $\alpha = t\mathbf{1}_{\mathcal{F}}$*. Thus, we have the directional derivatives*

$$D_u r_{\mathsf{VaR}}(t\mathbf{1}_{\mathcal{F}}) = u^\top X_{\omega^*}, \qquad D_{u,e_i} r_{\mathsf{VaR}}(t\mathbf{1}_{\mathcal{F}}) = X_{i,\omega^*},$$

for all directions u. Then, the risk attribution to firm i according to the Mertens mechanism (2.19) *simplifies to become*

$$\Psi_{M,i}^{|\mathcal{F}|}(r_{\mathsf{VaR}}, \mathbf{1}_{\mathcal{F}}) = X_{i,\omega^*}.$$

In other words, the risk attribution of firm i will be the loss incurred by the firm in the critical sce-

nario ω^*. *More generally, if the value-at-risk is achieved in multiple scenarios, the risk attribution for each firm will be an average of losses across these scenarios.*

Example 13 (Comonotonic Risk Measures)**.** *Following Kou et al. [2009], one can defined homogeneous aggregation functions of the form*

$$\Lambda(x) \triangleq \sum_{k=1}^{|\mathcal{F}|} \gamma_k x_{[k]},$$

where $\gamma \in \mathbb{R}_+^{|\mathcal{F}|}$, $\mathbf{1}_{\mathcal{F}}^\top \gamma = |\mathcal{F}|$, *and* $x_{[k]}$ *denotes the kth order statistic of the vector x. Note that the value-at-risk and the median are special cases of this risk function. Fix a probability measure p over the scenarios. Combining this aggregation function with the homogeneous single-firm risk measure*

$$\rho_0(x) \triangleq p^\top x,$$

we get the homogeneous systemic risk measure

$$\rho(X) \triangleq (\rho_0 \circ \Lambda)(X) = \sum_{\omega \in \Omega} p_\omega \Lambda(X_\omega).$$

Note that Λ *and* ρ_0 *are piecewise linear and continuous. Since all comonotonic risk measures are homogeneous risk measures, it follows that all our results in this section apply to such risk measures.*

2.6 Decentralized Implementation

In this section, we explore the decentralization of systemic risk management. In particular, we establish that the dual representation of Section 2.3 and the risk attribution rule of Section 2.4 provide the basis of a tax scheme for internalizing the systemic risk into the decisions of individual firms.

To this end, for each firm $i \in \mathcal{F}$, let $\mathcal{T}^{(i)} \subset \mathbb{R}^{|\Omega|}$ denote the convex feasible set of possible outcomes for firm i over the set of scenarios Ω. The set $\mathcal{T}^{(i)}$ can be interpreted as the set of possible investment opportunities available to firm i. Denote by $\mathcal{T} \triangleq \mathcal{T}^{(1)} \times \mathcal{T}^{(2)} \times \ldots \times \mathcal{T}^{(|\mathcal{F}|)} \subset \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$ the resulting set of possible economies. Let $U_i \colon \mathcal{T}^{(i)} \to \mathbb{R}$ denote the utility function of firm i. We assume that U_i is strictly concave and differentiable. Given a systemic risk measure $\rho \colon \mathbb{R}^{|\mathcal{F}| \times |\Omega|} \to \mathbb{R}$, we make the following definition:

Definition 9 (Social Optimality). *An economy $\bar{X} = \left(\bar{X}^{(1)}; \bar{X}^{(2)}; \ldots; \bar{X}^{(|\mathcal{F}|)}\right) \in \mathcal{T}$ is socially optimal if it maximizes risk-adjusted welfare according to the optimization problem*

$$\bar{X} \in \underset{X \in \mathcal{T}}{\operatorname{argmax}} \left\{ \sum_{i \in \mathcal{F}} U_i(X^{(i)}) - \tau \rho(X) \right\}. \tag{2.21}$$

Here, $\tau > 0$ is a parameter that captures the impact of the systemic risk externality.

A regulator or central planner wishes to drive individual firms to make decisions so that, collectively, these decisions results in a socially optimal economy. However, the regulator is not able to directly control the outcomes of each firm; it is only able to influence investment decisions indirectly via incentives. In particular, suppose the regulator imposes a tax $t_i(X^{(i)})$ on firm i given outcomes $X^{(i)} \in \mathcal{T}^{(i)}$. The firm would then choose outcomes so as to optimize its tax-adjusted utility, i.e., it would solve the optimization problem

$$\underset{X^{(i)} \in \mathcal{T}^{(i)}}{\text{maximize}} \left\{ U_i(X^{(i)}) - t_i(X^{(i)}) \right\}.$$

Motivated by the dual representation of Section 2.3 and the risk attribution rule of Section 2.4, the following theorem suggests a taxation scheme to implement any socially optimal economy:

Theorem 7. *Suppose that $\bar{X} \in \mathcal{T}$ is a socially optimal economy. There exists $\Xi^* \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$ that is an optimal solution to the dual problem* (2.14) *for the systemic risk $\rho(\bar{X})$ so that if we define, for each firm i, the tax function*

$$t_i(X^{(i)}) \triangleq \tau \sum_{\omega \in \Omega} \Xi^*_{i,\omega} X_{i,\omega},$$

then, $\bar{X}^{(i)}$ is an optimal outcome for firm i, i.e.,

$$\bar{X}^{(i)} \in \underset{X^{(i)} \in \mathcal{T}^{(i)}}{\operatorname{argmax}} \; U_i(X^{(i)}) - \tau \sum_{\omega \in \Omega} \Xi^*_{i,\omega} X_{i,\omega}. \tag{2.22}$$

Proof. Since $\bar{X}$ is a socially optimal economy, first order conditions for optimality for (2.21) imply that there must be a subgradient $\Xi^* \in \mathbb{R}^{|\mathcal{F}| \times |\Omega|}$ of ρ at $\bar{X}$ so that

$$\sum_{i \in \mathcal{F}} \sum_{\omega \in \Omega} \left(\nabla_{i,\omega} U_i(X^{(i)}) - \tau \Xi^*_{i,\omega} \right) (X_{i,\omega} - \bar{X}_{i\omega}) \leq 0, \tag{2.23}$$

for all $X \in \mathcal{T}$. Any subgradient Ξ^* is clearly a dual optimal solution to (2.14), according to Danskin's theorem [Bertsekas, 1999].

Now, given $i \in \mathcal{F}$, we can take $X^{(j)} = \bar{X}^{(j)}$ for all $j \neq i$ in (2.23). We obtain that, for all $i \in \mathcal{F}$,

$$\sum_{\omega \in \Omega} \left(\nabla_{i,\omega} U_i(X^{(i)}) - \tau \Xi^*_{i,\omega} \right) (X_{i,\omega} - \bar{X}_{i\omega}) \leq 0, \tag{2.24}$$

for all $X^{(i)} \in \mathcal{T}^{(i)}$. Note that (2.24) is the first order optimality condition for (2.22), and thus $\bar{X}^{(i)}$ is an optimal solution to (2.22). ■

Theorem 7 demonstrates that the objective of the regulator can be aligned with the incentives of individual firms through taxation schemes. Moreover, the tax paid by each firm is (up to the constant τ) determined by the risk attribution of that firm, c.f. Definition 5.

Since the taxation scheme in Theorem 7 is based on a subgradient Ξ^* at the social optimal economy $\bar{X}$, it appears that, in order to compute this taxation scheme, the regulator would need to solve the centralized problem (2.21). In other words, the regulator would need to know the utilities U_i and investment opportunities $\mathcal{T}^{(i)}$ of each firm i. While it is reasonable to assume that the regulator knows the risk function ρ and that it can demand that the firms reveal their losses X across scenarios, it is unlikely that the firms would reveal their utility functions or investment opportunities. This obstacle can be overcome via an iterative scheme provided we assume that each U_i is a smooth, strictly concave function, and that

1. the firms are myopic, i.e., the optimize their loss profiles with the given taxation scheme without considering the impact of their decisions on future taxes, and

2. the firms report their loss profiles truthfully.

The second condition can be achieved in a regulatory regime with suitable penalties. The first can be justified on the basis of bounded rationality on the part of the firms. Consider the following scheme, in each iteration k:

1. Each firm i computes

$$X_k^{(i)} \in \underset{X \in \mathcal{T}^{(i)}}{\operatorname{argmax}} \left\{ U_i(X^{(i)}) - \tau(\Xi_i^{(k)})^\top (X^{(i)} - X_{k-1}^{(i)}) - \frac{\nu_k}{2} \|X^{(i)} - X_{k-1}^{(i)}\|_2^2 \right\},$$

   where $X_{k-1}^{(i)}$ denotes the loss profile of the firm in iteration prior $k-1$, and $\nu_k \|X^{(i)} - X_{k-1}^{(i)}\|_2^2$ is a proximal term that prevents large changes in $X^{(i)}$ in each iteration. The optimization problem above can be interpreted as a utility maximization problem with convex taxes.

2. The regulator receives the outputs $X_k = (X_k^{(i)}, \ldots, X_k^{(|\mathcal{F}|)})$ and computes a subgradient Ξ_k of ρ at X_k.

There exists a suitable choice for the sequence of regularization parameters $\{\tau_k\}$ such that proximal subgradient scheme above converges to a socially optimal economy $\bar{X}$ and the subgradient Ξ_k converge to a subgradient $\bar{\Xi}$ of ρ at $\bar{X}$ such that the first order optimality condition (2.24) hold for the pair $(\bar{X}, \bar{\Xi})$ [e.g., Bertsekas, 1999]. Note that each iteration of the above subgradient scheme, the regulator only requires the firms to truthfully communicate the current iterate X_k and not their utility functions.

2.7 Discussion

There are a number of exciting directions of research that are worth mentioning: One important issue is understanding how mergers or spin-offs in the economy affect the systemic risk. In the case where the aggregation function is the sum of profits or losses of individual firms, the systemic risk does not change when firms choose to combine or subdivide. However, with general aggregation functions, the systemic risk measure may change. A related point is understanding how a regulator can express preferences for the distribution of losses across the economy. Is it better to have a single firm lose a large some of money, or have a loss equally divided amongst many firms? Existing systemic risk measures are largely indifferent between these two cases, while a regulator may not be. Our framework is sufficiently expressive to allow systemic risk measures that express such preferences. Another direction is related to the issue of firms which are 'too-big-to-fail'. In this chapter, we primarily considered positively homogeneous risk measures, which penalize the scale of losses linearly. One might alternatively consider super-linear systemic risk measures, which assign increasing marginal risk to the scale of an economy. Such systemic risk measures could be defined by considering a set of axioms which include convexity but not positive homogeneity.

Issues related to the strategic behavior of firms with respect to systemic risk measures are also worth consideration. This topic would be of great importance, for example, if a regulator wanted to implement a tax scheme according the risk attribution of a systemic risk measure.

Chapter 3

Asset-based Contagion Models for Systemic Risk

3.1 Introduction

To identify the key ingredients of systemic risk that stem from the interactions of components of systems, this chapter proposes a structural model which highlights contagion effects. In particular, we investigate the financial network that emphasizes asset-firm portfolio holding and re-balancing. The main contributions of this paper are as follows:

- **We provide a structural model for a financial network consisting of a set of firms holding common assets.**

  In the model, endogenous asset prices are captured by the marketing clearing condition when the economy is in equilibrium. The key ingredients in the financial market that are captured in this model include the general portfolio choice flexibility of firms given posted asset prices and economic states, and the mark-to-market wealth of firms.

- **We provide a price sensitivity analysis for the financial network.**

  The analysis characterizes the mechanism of network contagion when facing various forms of exogenous shocks. In particular, we identify two key components of the chain effects in the network. One is direct effects, and the other is network effects. Direct effects, as the immediate response under shocks, allows for the modeling flexibility of external shocks of different types. On the other hand, network effects highlights the network contagion by connecting direct effects to price changes. The network contagion can be viewed as the process where the direct effects, generated immediately under shocks, are propagated through network effects to form final price changes. Another interesting result is that the contagion can be viewed as an aggregate effect of all paths through the bipartite network of asset-firm holding. The intensity of the effect of each path is proportional to the holding portfolio between assets and firms.

- **We characterize the key features of financial holding networks that minimize systemic risk, as a function of overall leverage.**

  Given primitives including asset prices, wealth of firms, and leverage of firms, we solve the optimization problem of designing the best holding networks that minimize the network amplifier under given shocks. We identify two economic regimes that are characterized by systemic leverage of the financial system. Surprisingly, we found that the structures of the optimal networks for these two regimes are completely different, with even opposite features. For a low-leverage economy, the optimal network is the mutual-fund network where every firm holds a common market portfolio. However, for a high-leverage economy, the optimal networks have a structural tendency to drive the holding network into partitions, where each asset is isolated and only held by a subgroup of firms. Meanwhile, firms tend to invest all the leveraged wealth into only one asset. Consequently the connections across assets via common holding firms are kept minimum. As comparison, the low-leverage economy favors diversification of investment across assets, but the high-leverage economy favors diversity of investment across firms. The systemic leverage becomes a defining quantity that distinguishes these two economic regimes. This provides a managerial guidance for the regulator to monitor the systemic risk of the financial system and possibly to design regulatory policy to drive the system to less risky states.

There is a growing literature on structural models for systemic risk and in particular contagion through financial networks. Some of the pioneers who worked on this topic include Allen and Gale [2000] who modeled cross-firm connectivity through banks holding interregional claims on other banks, and Eisenberg and Noe [2001] who studied interbank lending networks and provided a debt clearing mechanism. Some of the variations and generalization of Eisenberg and Noe [2001] include Acemoglu et al. [2013] who characterize the optimal interbank lending networks, and N. Chen and Yao [2014] argues the amplification of shocks result from effects of networks and liquidity. Other models of financial networks include Gai and Kapadia [2010] who exploit a model of epidemics for shock propagation. Another recent work is from Glasserman and Young [2014] who compute the probability of systemic failure by simulation. The above-mentioned articles emphasizes the interactions among institutions, with some elements of assets that can be liquidated for debt obligations. This chapter put the entire focus on the explicit connection between firms and assets through portfolio holdings and re-balancing. The cross-firm relationship in our model is implicit since two firms are connected through the network of asset-firm holdings. One important work that share similar features is from Brunnermeier and Pedersen [2009] who model the traders' funding liquidity and an asset's market liquidity to explain empirical features of market liquidity. Another work by Greenwood et al. [2012] models asset-based contagion by assuming leverage targeting and a reduced-form price impact. As a comparison, this chapter models an endogenous price for portfolio re-balancing.

There is one result in this chapter that relates well to the existing literature. It is the tradeoff between diversification and diversity. This form of tradeoff shares similar intuition with some published work. For example, Wagner [2011] presents similar phenomenon through an equilibrium model with the explicit modeling of the joint liquidation cost in a specific portfolio choice model, while the model in this chapter accommodates general portfolio choices and we consider the optimal reduction of systemic risk from the perspective of the regulator. Elliott et al. [2013] found the tradeoff between diversification, i.e., more counterparties in their model, and integration, i.e., deeper relationship with each counterparty. Despite of the fact that their model does not involve tradable assets, on the higher level of the source of contagion in financial networks, we share the same viewpoint that the conventional diversification argument in the risk management lack of some key considerations for systemic risk. We need to point out that the fundamental difference between Elliott et al. [2013] and this chapter is that the concept of "assets" Elliott et al. [2013] propose is through the counterparty cross-holding among firms, but we explicitly model the portfolio holding relationship between firms and tradable assets.

This chapter is organized as follows: in section 3.2, we establish the settings of our model that highlights the endogenous asset prices via the market clearing condition. In section 3.3, we conduct the price sensitivity analysis by applying the implicit function theorem, where we identify key components in the asset-based contagion. In section 3.4, we introduce the network amplifier as a measurement of the contagion. In section 3.5, we investigate how the network configuration affects contagion. The low-leverage and high-leverage cases are solved respectively.

3.2 Model Formulation

In this section, we will establish a model for the financial network of interrelated assets and firms, which endogenously captures asset prices through a market clearing condition. In particular, let $\mathcal{A}$ denote the set of assets, and $\mathcal{F}$ denote the set of financial firms. We assume the initial endowments for all of the firms are denoted by (θ^0, Θ), where $\theta^0 \in \mathbb{R}^{|\mathcal{F}|}$ are the endowments in the risk-free asset for all of the firms, and $\Theta \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}$ are the endowments in the risky assets. In particular, Θ_{ih} is the number of shares of asset i initially held by firm h. For any posted asset prices denoted by $q \in \mathbb{R}^{|\mathcal{A}|}$, the wealth of firms is given by the vector $w = \theta^0 + \Theta^\top q \in \mathbb{R}^{|\mathcal{F}|}$. Finally, define $D \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ to be a diagonal matrix where $D_{ii} = \sum_{h\in\mathcal{F}} \Theta_{ih}$ is the total number of outstanding shares for asset i.

The portfolio chosen by the firms is denoted by the matrix $\Pi \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}$. Specifically, Π_{ih} is the percentage of the wealth of firm h invests in asset i. For each $h \in \mathcal{F}$, the hth column of the matrix Π is denoted by a vector $\Pi_h \in \mathbb{R}^{|\mathcal{A}|}$, which represents the portfolio of firm h. In our model, we allow the portfolio choice Π_h to depend on:

- q, the vector of posted prices.
- w_h, the wealth of firm h.
- $x \in \mathbb{R}$, which is a risk factor.

The risk factor x captures an underlying variable that, in what follows, will be subject to a shock and whose downstream effects we wish to understand. Depending on the context, for example, x might capture an economic factor, beliefs of investors, and changes in regulation.

In other words, Π_h takes the functional form $\Pi_h(q, w_h, x)$. Here we wish to allow x to be flexible in terms of modeling other factors that affect portfolio choices. For example, x can be the beliefs of firms for asset returns. Also, we allow for the possibility that initial cash endowment θ_0 depends on x. Namely θ_0 is a function of x.

In our model, Π can capture a variety of portfolio choice mechanisms or preset re-balancing rules. In the following example, we present a concrete form of Π as a function of asset prices q and other factors.

Example 14 (CRRA Investors). *Suppose that firm h has the belief that the payoff of the risky assets are a random vector $p_h \in \mathbb{R}^{|\mathcal{A}|}$ which follows a jointly log-normal distribution, i.e., $\log(p_{ih}) \sim N(\mu_h, \Sigma_h)$. For a fixed asset price q_i, the returns of the risky assets, under the belief of firm h, is denoted by $R_h = \left\{ \frac{p_{ih}}{q_i} \right\}_i \in \mathbb{R}^{|\mathcal{A}|}$, which also follows a jointly log normal distribution. Given the belief, the vector of log-returns takes the form $r_{ih} = \log \frac{p_{ih}}{q_i} = \log(p_{ih}) - \log(q_i)$, and thus $r_h \sim N(\hat{\mu}_h, \hat{\Sigma}_h)$, which follows a joint normal distribution where $\hat{\mu}_{ih} = \mu_{ih} - \log(q_i)$ and $\Sigma_h = \hat{\Sigma}_h$.*

Assume that firm h chooses its portfolio $\Pi_h \in \mathbb{R}^{|\mathcal{A}|}$ in the risky assets and $\pi_h \in \mathbb{R}$ in the risk-free asset, to maximizes its expected utility. The utility function U_h for firm h has the constant *relative risk aversion (CRRA) characterized by β_h. So the optimization problem firm h solves is the following.*

$$\underset{\mathbf{1}^\top \Pi_h + \pi_h^0 = 1}{\text{maximize}} \; E[\frac{\tilde{w}_h^{1-\beta_h}}{1-\beta_h}], \tag{3.1}$$

where $\tilde{w}_h = w_h^0 \sum_{i\in\mathcal{A}} \Pi_{ih} e^{r_{ih}}$ is the random wealth of firm h. According to the derivation by Campbell and Viceira [2002], we have an approximately optimal solution as follows:

$$\Pi_h = \frac{1}{\beta_h} \Sigma_h^{-1} \left(\mu_h - \log(q) - r_f \mathbf{1} + \frac{1}{2} \operatorname{diag}\{\Sigma_h\} \right). \tag{3.2}$$

We notice that in this example the function of portfolio Π depends on assets prices q, risk aversion factor β, beliefs μ and Σ, but not wealth.

Example (14) is one of many portfolio choice rules that can be captured in the model. It shows that the portfolio matrix Π can be the solution of utility maximization problems. In other cases, Π can possibly be preset re-balancing rules. The model allows for very general portfolio choices that depend on asset prices and other factors.

3.2.1 Market Equilibrium

In this section, we will define our concept of market equilibrium, which captures a set of asset prices and portfolios consistent with the investors' portfolio choice rules and where supply and demand are balanced. We will further see that our equilibrium is equivalent to a simpler market clearing condition. To begin, we consider the following definition.

Definition 10 (Equilibrium)**.** *Suppose we are given initial endowments (Θ, θ^0) and the risk factor x.*

An equilibrium is the tuple $(q, \hat{\Theta}, \hat{\theta}^0)$*, where* $q \in \mathbb{R}^{|\mathcal{A}|}$ *are posted asset prices,* $\hat{\Theta} \in \mathbb{R}^{|\mathcal{F}| \times |\mathcal{A}|}$ *is the portfolio of risky assets held by firms, and* $\hat{\theta}^0 \in \mathbb{R}^{|\mathcal{A}|}$ *is the vector for risk-free assets held by firms. We require that* $(q, \hat{\Theta}, \hat{\theta}^0)$ *satisfy:*

- *Budget Balance.*

$$\sum_{i \in \mathcal{A}} \left(\hat{\Theta}_{ih} - \Theta_{ih}\right) q_i + \left(\hat{\theta}^0_h - \theta^0_h\right) = 0, \quad \forall\, h \in \mathcal{F}. \tag{3.3}$$

- *Share Balance.*

$$D_{ii} = \sum_{h \in \mathcal{F}} \Theta_{ih} = \sum_{h \in \mathcal{F}} \hat{\Theta}_{ih}, \quad \forall\, i \in \mathcal{A}. \tag{3.4}$$

- *Portfolio Consistency.*

$$\hat{\Theta}_{ih} q_i = \Pi_{ih}(q, \theta^0 + \Theta^\top q, x) \left(\hat{\theta}^0_h + \sum_{j \in \mathcal{A}} \hat{\Theta}_{jh} q_j\right), \quad \forall\, i \in \mathcal{A},\ h \in \mathcal{F}. \tag{3.5}$$

In this definition of equilibrium, the budget balance condition in equation (3.3) requires that at posted asset prices q, the portfolio re-balancing from (Θ, θ^0) to $(\hat{\Theta}, \hat{\theta}^0)$ does not change the wealth of each firm. This condition immediately implies that wealth is invariant under re-balancing, i.e., $w = \theta^0 + \Theta^\top q = \hat{\theta}^0 + \hat{\Theta}^\top q$. The share balance condition in equation (3.4) requires that at posted asset prices q, the number of shares of each asset is not changed by the portfolio re-balancing from (Θ, θ^0) to $(\hat{\Theta}, \hat{\theta}^0)$. The portfolio consistency condition in equation (3.5) indicates that Π and Θ are consistent ways of expressing the portfolio of firms. Π_{ih} denotes the percentage of wealth of firm h invested in asset i, but Θ_{ih} represents the portfolio in terms of the number of shares that firm h holds in asset i.

In the following theorem, we introduce a simpler characterization of an equilibrium. The proof of the theorem can be found in Appendix A.2.

Theorem 8. *Suppose we are given initial endowments* (Θ, θ^0) *and the risk factor x. For a vector of posted asset prices q, there exists an equilibrium with prices q if and only if q satisfies the following* ***market clearing condition****:*

$$Dq = \Pi(q, \theta^0 + \Theta^\top q, x) \left(\theta^0 + \Theta^\top q\right). \tag{3.6}$$

The market clearing condition in equation (3.6) captures the key ingredients of how firms are connected with assets through the asset holding relationships. We notice that the left-side of equation (3.6) is the vector for market capitalization, and the right-side is a vector whose element is the total invested wealth for each asset. One way to interpret the market clearing condition is that the market capitalization equals the total invested wealth. Intuitively, the total invested wealth $\Pi\left(\theta^0 + \Theta^\top q\right)$ is the dollar demand of investment in the market, while the market capitalization Dq is the dollar supply for investment in the market.

In order to interpret Theorem 8, it is helpful to consider the following definition.

Definition 11. *Suppose we are given initial endowments* (Θ, θ^0) *and the risk factor x. For a vector of posted asset prices q, firms choose portfolio* Π*. The* ***excess dollar demand*** *is defined as follows:*

$$EDD = \Pi(q, \theta^0 + \Theta^\top q, x) \left(\theta^0 + \Theta^\top q\right) - Dq. \tag{3.7}$$

EDD characterizes a mismatch in the market. It is the difference between the total wealth that is to be invested and the market capitalization. It is clear that q satisfies the market clearing condition in equation (3.6) if and only if the excess dollar demand is zero.

3.3 Asset Price Contagion

In this section, we wish to understand how various forms of exogenous shocks are propagated to generate endogenous asset price changes. Through a price sensitivity analysis, we will try to identify key components that contribute to price movement and characterize the contagion effects. In general, we wish to capture exogenous shocks through a risk factor denoted by x. A shock is defined as a change in some exogenous variables defined by x. With respect to changes in x, we characterize $\frac{dq}{dx} \in \mathbb{R}^{|\mathcal{A}|}$, the vector describing price sensitivities.

First, we introduce some notation. Define $\bar{\Theta} \triangleq D^{-1}\Theta \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ to be a version of Θ normalized by the number of shares. Define a new matrix $H \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ as

$$H \triangleq \begin{bmatrix} (G^1 w)^\top \\ (G^2 w)^\top \\ \vdots \\ (G^n w)^\top \end{bmatrix}, \tag{3.8}$$

where for each i, $G^i \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}$ is defined as

$$G^i \triangleq \begin{bmatrix} \frac{\partial \Pi_{i1}}{\partial q_1} & \frac{\partial \Pi_{i2}}{\partial q_1} & \cdots & \frac{\partial \Pi_{im}}{\partial q_1} \\ \frac{\partial \Pi_{i1}}{\partial q_2} & \frac{\partial \Pi_{i2}}{\partial q_2} & \cdots & \frac{\partial \Pi_{im}}{\partial q_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial \Pi_{i1}}{\partial q_n} & \frac{\partial \Pi_{i2}}{\partial q_n} & \cdots & \frac{\partial \Pi_{im}}{\partial q_n} \end{bmatrix}. \tag{3.9}$$

We can interpret H as the wealth-weighted cross-asset portfolio sensitivity matrix. That is, the entry H_{ij} describes a certain linkage between asset j and asset i. Specifically, we can write out

$$H_{ij} = [(G^i w)^\top]_j = \sum_{h=1}^{|\mathcal{F}|} \frac{d\Pi_{ih}}{dq_j} w_h. \tag{3.10}$$

It represents the wealth-weighted change of Π due to the change in prices q. Another new matrix is $\Omega \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}$ defined as

$$\Omega \triangleq \begin{bmatrix} \frac{\partial \Pi_{11}}{\partial w_1} w_1 & \frac{\partial \Pi_{12}}{\partial w_2} w_2 & \cdots & \frac{\partial \Pi_{1|\mathcal{F}|}}{\partial w_{|\mathcal{F}|}} w_{|\mathcal{F}|} \\ \frac{\partial \Pi_{21}}{\partial w_1} w_1 & \frac{\partial \Pi_{22}}{\partial w_2} w_2 & \cdots & \frac{\partial \Pi_{2|\mathcal{F}|}}{\partial w_{|\mathcal{F}|}} w_{|\mathcal{F}|} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial \Pi_{|\mathcal{A}|1}}{\partial w_1} w_1 & \frac{\partial \Pi_{|\mathcal{A}|2}}{\partial w_2} w_2 & \cdots & \frac{\partial \Pi_{|\mathcal{A}||\mathcal{F}|}}{\partial w_{|\mathcal{F}|}} w_{|\mathcal{F}|} \end{bmatrix}. \tag{3.11}$$

Ω is the wealth-weighted firm-asset portfolio sensitivity matrix: the entry Ω_{ih} specifies the portfolio sensitivities of the position of firm h in asset i with respect to wealth change of firm h.

The main result of this section is the following theorem, whose proof can be found in Appendix A.2.

Theorem 9. *Suppose (q,Θ,θ^0) is a given equilibrium, and the risk factor is x_0. Assume that $\Pi(q(x),w(\theta^0(x),q(x)),x)$ is a continuously differentiable function of q, w, and x, and $\theta^0(x)$ is a continuously differentiable function of x. If the matrix $\left[I-(\Pi+\Omega)\bar{\Theta}^\top-HD^{-1}\right]$ is invertible at $x=x_0$, then there exists a neighborhood $N(x_0)$ of x_0 and a unique price function $q(x)$ on $N(x_0)$ such that for all $x\in N(x_0)$, $q(x)$ satisfies the market clearing condition,*

$$Dq(x) \quad = \quad \Pi(q(x),w(\theta^0(x),q(x)),x)\left(\theta^0(x)+\Theta^\top q(x)\right), \tag{3.12}$$

and we have that

$$D\frac{dq}{dx}=\left[I-(\Pi+\Omega)\bar{\Theta}^\top-HD^{-1}\right]^{-1}\left[(\Pi+\Omega)\frac{d\theta^0}{dx}+\frac{\partial\Pi}{\partial x}w\right]. \tag{3.13}$$

The main result of Theorem 9 is to characterize $\frac{dq}{dx}$, the price sensitivity with respect to a shock. The detailed interpretation is in the following section.

3.3.1 Direct Effects and Network Effects

In this section, we try to interpret Theorem 9 to form a deeper understanding of the formation of price changes. First, we define two concepts.

Definition 12 (Direct Effects)**.** *The direct effects are defined as the vector*

$$[DE]\triangleq(\Pi+\Omega)\frac{d\theta^0}{dx}+\frac{\partial\Pi}{\partial x}w. \tag{3.14}$$

Definition 13 (Network Effects)**.** *The network effects are defined as the matrix*

$$[NE] \triangleq \left[I - (\Pi + \Omega)\,\bar{\Theta}^{\top} - HD^{-1}\right]^{-1}. \tag{3.15}$$

Given these definitions, we can re-write the price sensitivity equation as follows:

$$D\frac{dq}{dx} = [NE] \cdot [DE]. \tag{3.16}$$

In what follows, we will see that the direct effects capture the immediate response from the financial system when facing the external shocks, while the network effects characterize the follow-on effects through the interactions within the financial network to eventually produce the endogenous price changes.

In particular, we can interpret the direct effects as follows. Suppose we keep the posted price q fixed when we shock the risk factor according to $x' = x + \Delta x$, where Δx is a small external shock on x. Suppose $w = \theta^0(x) + \Theta^{\top} q$ is the wealth vector. When the conditions in Theorem 9 are satisfied, the excess dollar demand is given by

$$\begin{aligned}
EDD(q,x') &= \Pi(q, w(\theta^0(x')), x')\left(\theta^0(x') + \Theta^{\top} q\right) - Dq \\
&= \left[(\Pi + \Omega)\,\tfrac{d\theta^0}{dx} + \tfrac{\partial \Pi}{\partial x} w\right](x' - x) + \Pi(q, w(\theta^0(x)), x)\left(\theta^0(x) + \Theta^{\top} q\right) \\
&\quad - Dq + o(\Delta x) \\
&= \left[(\Pi + \Omega)\,\tfrac{d\theta^0}{dx} + \tfrac{\partial \Pi}{\partial x} w\right]\Delta x + o(\Delta x) \\
&= [DE]\Delta x + o(\Delta x),
\end{aligned}$$

where for the third equality we use the fact $EDD(q,x) = 0$ as market clears in equilibrium. Therefore, for a small shock Δx, if the posted prices don't change, there would be an excess dollar demand in the system proportional to the vector of direct effects. Intuitively, the direct effects quantifies that immediate change in the assets across the market, thus resulting in an imbalance in the market clearing condition. If there exists an outside investor who manages to absorb this excess dollar demand, the posted prices could indeed stay unchanged.

In the following examples, we describe the vector of direct effects $[DE]$ in the context of a variety of shocks

Example 15 (Cash Shock)**.** *Suppose firm g has a wealth shock on its initial cash position, i.e., the risk factor takes the form* $x = \theta_g^0$*. In this case,* $[DE] = \Pi_g$*. In other words, the direct effects of this shock correspond to the portfolio of firms.*

Example 16 (Changes in Beliefs)**.** *Suppose firm g has a belief in the distribution of the future payoff of asset k. The mean of the distribution is denoted by* μ_{kg}*. If we take the risk factor to be* $x = \mu_{kg}$*, then when firm g changes its belief in* σ_{kg}*, this is equivalent to a shock on x. In this case,* $[DE] = \frac{\partial \Pi_g}{\partial \mu_{kg}} w_g$*. We can interpret* $[DE]$ *as the change, caused by the change of the belief, in the dollar positions of firm g assuming wealth* w_g *is fixed.*

Example 17 (Changes in Regulation)**.** *Suppose there is a regulation that the level of leverage for firms is limited to be below a fixed parameter* $L \in \mathbb{R}_+$*. When firms choose their portfolios, they must satisfy this leverage constraint. If we take the risk factor to be* $x = L$*, then when there is a change in the regulation on the level of L, it is equivalent to a shock on x. In this case,* $[DE] = \sum_{g \in \mathcal{F}} \frac{\partial \Pi_g}{\partial L} w_g$*.*

We can interpret $[DE]$ as the change, caused by the change of the regulation, in the dollar position of firm g assuming wealth w_g is fixed.

While the direct effects capture the excess dollar demand of a shock, the network effects describe the way that these demands must subsequently propagate in order to re-equilibrate the system. In other words, asset-firm interactions and contagion within the system are characterized in the formula (3.15) of $[NE]$. We identify three key components, $\Pi\bar{\Theta}^\top$, $\Omega\bar{\Theta}^\top$ and HD^{-1}.

- $\Pi\bar{\Theta}^\top \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ characterizes the holding-induced cross-asset interaction through Π and $\bar{\Theta}$, with Π been fixed. It indicates how the excess demand of asset i is affected by the excess demand of asset j through the fixed portfolios of all firms. In particular,

$$(\Pi\bar{\Theta}^\top)_{ij} = \sum_{h\in\mathcal{F}} \Pi_{ih}\bar{\Theta}_{jh}.$$

  This component emphasizes the effect from portfolio re-balancing.

- $\Omega\bar{\Theta}^\top \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ characterizes how the excess demand of asset i is affected by the excess demand of asset j through the portfolio sensitivity with respect to wealth. In particular,

$$(\Omega\bar{\Theta}^\top)_{ij} = \sum_{h\in\mathcal{F}} \frac{\partial\Pi_{ih}}{\partial w_h}\bar{\Theta}_{jh}w_h.$$

  This component emphasizes the wealth effect that brought by the change in wealth.

- $HD^{-1} \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ provides the normalized wealth-weighted cross-asset portfolio elasticity. It characterizes how the excess demand of asset i is affected by the excess demand of asset j

through the portfolio sensitivity with respect to price. In particular,

$$(HD^{-1})_{ij} = \sum_{h \in \mathcal{F}} \frac{\partial \Pi_{ih}}{\partial q_j} \frac{w_h}{D_{jj}}.$$

In the following sections, we will conduct more investigation on $[NE]$ through network interpretations, quantitative measure of the effects, and optimal design for the effects.

3.3.2 Bipartite Structure of Financial Networks

In order to gain better intuition for network effects, we will focus on the first component $\Pi\bar{\Theta}^\top$ in the above discussion. This component relates to the interconnectedness of the financial networks. We will show that the contagion effects can be characterized by the connectivity of the network.

Specifically, assume that $\Omega = \mathbf{0}$ and $H = \mathbf{0}$, i.e., there is no wealth effect and no cross-asset portfolio elasticity. Then the network effects matrix takes the form

$$[NE] = [I - \Pi\bar{\Theta}^\top]^{-1}.$$

From classic linear algebra, when the spectral radius $\rho(\Pi\bar{\Theta}^\top) < 1$ for the square matrix $\Pi\bar{\Theta}^\top$, we can expand the network effects into a infinite power series.

$$[NE] = \sum_{t=0}^{\infty} \left(\Pi\bar{\Theta}^\top\right)^t = I + \Pi\bar{\Theta}^\top + \left(\Pi\bar{\Theta}^\top\right)^2 + \dots. \tag{3.17}$$

This power series can be interpreted in the context of the holding network described as follows:

Definition 14. *The contagion graph has a bipartite structure. We have two types of nodes, $\mathcal{V} = \mathcal{A} + \mathcal{F}$, where assets are denoted by set $\mathcal{A}$ and firms are denoted by set $\mathcal{F}$. We have two types of directed edges, $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$, where the asset-to-firm edges are denoted by $\mathcal{E}_1$ and the firm-to-asset edges are denoted by $\mathcal{E}_2$. The weights of these links are specified as follows:*

- *For each edge $(j,g) \in \mathcal{E}_1$ that links asset j to firm h, its weight is $\bar{\Theta}_{jg}$.*
- *For each edge $(h,i) \in \mathcal{E}_2$ that links firm h to asset i, its weight is Π_{ih}.*

With the above definition, a directed path from asset j to asset i in the contagion graph is a sequence of edges that connect a sequence of nodes, with the starting node being asset j and the ending node being asset i. The length of a directed path is the number edges in the path. A path between two assets in the contagion graph represents a pathway of cascading flow, through which shocks can be propagated. The contribution of a particular path to overall contagion is the product of the weights of all edges along the path. We notice that the terms of the power series in equation (3.17) can be interpreted as follows:

- I is the identity matrix, which captures the direct effects of the shock.
- $\Pi\bar{\Theta}^\top$ is the first-level cross-asset impact matrix. Each element $\left(\Pi\bar{\Theta}^\top\right)_{ij}$ characterizes the impact that the excess demand of asset i has on asset j via all firms whose initial endowments consist of asset i and invest in asset j.
- $\left(\Pi\bar{\Theta}^\top\right)^2$ is the second-level cross-asset impact matrix. Each element $\left(\Pi\bar{\Theta}^\top\right)^2_{ij}$ characterizes the impact that the excess demand of asset i has on asset j via all via all possible intermediate assets. Now the contagion has a second-level impact.

- $\left(\Pi\bar{\Theta}^{\top}\right)^{t}$ is the t-th-level cross-asset impact matrix for any t. Each element $\left(\Pi\bar{\Theta}^{\top}\right)_{ij}^{t}$ characterizes the impact that the excess demand of asset i has on asset j via all via all possible paths with length $2t$. Now the contagion has a t-th-level impact.

To summarize, for any two assets in the financial network, the contagion effect from one to the asset can be characterized by the aggregate impact of all possible paths with all possible lengths in the network between these two assets. The intensity of the contagion is the sum of all the impact matrices of all levels according to equation (3.17).

In the literature, there are many articles that investigate related financial networks. For example, Eisenberg and Noe [2001] discuss the interbank lending networks. Elliott et al. [2013] provide a model for cross-holdings among institutions. While most of the existing literature emphasizes cross-firm interactions through debt obligation or institution-wise cross-holdings, the financial networks discussed in this chapter characterize the bipartite connection between firms and tradable assets via holding portfolios of assets by firms and portfolios re-balancing.

For the rest of this chapter, we will continue to focus on the structure of the asset-firm holding network as in Section 3.3.2. So we keep the assumption that $H = 0$ and $\Omega = 0$. This is a reasonable assumption when firms are executing a portfolio tracking strategy. In other words, $H = 0$ and $\Omega = 0$ when firms maintain a fixed Π.

3.4 Network Amplifier, Holding Network, and Leverage

In this section, we introduce a new concept, the network amplifier, as a quantitative approach to measure the scale of the contagion effects. In particular, when $\Omega = \mathbf{0}$ and $H = \mathbf{0}$ (as in Section 3.3.2), the network effects matrix takes the form

$$[NE] = [I - \Pi\bar{\Theta}^{\top}]^{-1}.$$

We will quantify the scale of these effects as follows:

Definition 15 (Maximal Network Amplifier)**.** *Suppose (q, Θ, θ^0) is a given equilibrium, and the risk factor is x_0. Assume that $\Pi(x)$ is a continuously differentiable function of x, and the matrix $\left[I - \Pi\bar{\Theta}^{\top}\right]$ is invertible at $x = x_0$. The network amplifier is defined by*

$$\mathit{MNA} \triangleq \rho([NE]) = \rho([I - \Pi\bar{\Theta}^{\top}]^{-1}), \tag{3.18}$$

where $\rho(\cdot)$ is the spectral radius of a matrix.

In what follows, we will see that under mild technical conditions, $[NE]$ has real eigenvalues. From straightforward linear algebra, it follows that

$$\mathsf{MNA} = \max_{\|[DE]\|_2 \leq 1} \|[NE] \cdot [DE]\|_2.$$

This implies that

$$\|[NE] \cdot [DE]\|_2 \leq \mathsf{MNA} \|[DE]\|_2.$$

The above relationship motivates our definition for the network amplifier. It is clear that the definition of MNA gives the worst-case amplification caused by shocks. We notice that for any given vector of direct effects $[DE]$, the amplification that the system can produce is bounded by the network amplifier $\rho([NE])$. The network amplifier offers a bound on the scale of the network amplification between direct effects and price changes.

3.4.1 Holding Network and Leverage

In order to understand the key drivers of contagion, we will decompose the structure of an economy into two components. One is the leverage of firms, and the other is the configuration of the interactions of firms and assets. The goal is to separately describe the dependency of contagion effects on network configuration and leverage.

We assume that the proportion of net wealth in risky asset for firm h is b_h. In other words, b_h is the ratio between the wealth invested in risky asset by firm h and the net wealth of firm h. We require that $b_h > 0$ so that firm h has a positive position in risky assets. When $b_h > 1$, firm h is borrowing cash to invest in risky asset. Thus, firm h is leveraged with a b_h-to-1 ratio. The leverage vector across firms is denoted by $b \in \mathbb{R}^{|\mathcal{F}|}$. The leverage vector relates to the portfolio choice Π as

$$b_h = \sum_{i \in \mathcal{A}} \Pi_{ih}, \quad \forall\, h \in \mathcal{F}.$$

We introduce a **holding network** matrix $X \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{F}|}$, which encodes the relationship between each firm and risky asset holdings. In particular,

$$X_{ih} = \frac{\Pi_{ih}}{b_h}, \quad \forall\, i \in \mathcal{A},\; h \in \mathcal{F}.$$

Each element X_{ih} is the fraction of all risky assets that is held in asset i by firm h, and $\sum_{i \in \mathcal{A}} X_{ih} = 1$

for any h. Now, we can re-write Π.

$$\Pi = X\,\mathrm{diag}(b). \tag{3.19}$$

Now we have two ways to represent the portfolio choices of firms. One is through Π matrix. The other is through X matrix and the leverage vector b.

In this setting, we will simplify the expression for the maximal network amplifier. Specifically, we assume that the initial endowment is in equilibrium. The condition we assume is that Π and Θ_{ih} describe the same portfolio at prices q, so the following condition holds.

$$\Theta_{ih} q_i = \Pi_{ih} w_h, \quad \forall\, i\, \forall\, h, \tag{3.20}$$

where we recall that $w \triangleq \theta^0 + \Theta^\top q$.

Now we present the following theorem, whose proof can be found in the Appendix A.2.

Theorem 10. *Suppose (q, Θ, θ^0) is a given equilibrium, with the risk factor $x = x_0$. Assume that the conditions of Theorem 9 hold, and $H = \mathbf{0}$ and $\Omega = \mathbf{0}$. If the initial endowment is in equilibrium and satisfies equation* (3.20)*, then we have the following results.*

1. $$\Pi\bar{\Theta}^\top = Y(X),$$

   where $Y(X) \triangleq X\,\mathrm{diag}(b)\,\mathrm{diag}(w)\,\mathrm{diag}(b) X^\top M^{-1}$, where $M = D\,\mathrm{diag}(q)$ is a diagonal matrix, where the diagonal entries are the total market chaptalization of each asset.

2. *The eigenvalues of $Y(X)$ are real and non-negative.*

3. The maximal network amplifier is given by

$$\mathsf{MNA} = \max_{1 \le i \le |\mathcal{A}|} |1 - \lambda_i(Y(X))|^{-1}, \tag{3.21}$$

where each λ_i is the ith eigenvalue of the matrix.

Theorem 10 links the magnitude of the network effects to the eigenvalues of a matrix related to the holding network X, the leverage b and the wealth w. In the following section, we will investigate how these two components contribute to contagion through the analysis of the maximal network amplifier.

3.5 Optimal Network Design

Given an economy consisting of a set of firms and a set of assets, this section investigates the optimal structure of holding networks form a perspective of mitigating contagion and systemic risk. The concrete problem we are interested in is to decide which holding network that is feasible for a given economy has the smallest MNA. It is worth mentioning that there are two holding networks with extreme structures. One is the mutual fund economy where all firms invest in the same market portfolio. In this case, firms benefit from diversification across the assets that provides a buffer for shocks. The other case is the isolated economy where firms hold different portfolio with each other. This structure limits market activities and contagion. It turns out that these two extreme structures play critical roles in deciding the optimal holding networks. In what follows, we will show that if the economy has overall low leverage, the mutual fund economy is the optimal holding network when for a high-leverage economy, the isolated structure provides the optimal holding network.

3.5.1 Economies and Feasible Holding Networks

In our model, we identify four quantities that characterize an economy in the following definition.

Definition 16 (Economy)**.** *Suppose $D \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}$ is a diagonal matrix where diagonal entries are the numbers of shares for assets, $q \in \mathbb{R}^{\mathcal{A}}$ denotes asset prices, $w \in \mathbb{R}^{\mathcal{F}}$ denotes wealth of firms, and $b \in \mathbb{R}^{\mathcal{A}}$ denotes leverage of firms. The quantity (D,q,w,b) defines an economy if*

$$\mathbf{1}_{\mathcal{A}}^{\top} Dq = w^{\top} b.$$

For the condition in Definition 16, we notice that $\mathbf{1}_{\mathcal{A}}^{\top} Dq$ is the total market capitalization for risky assets, and $w^{\top} b$ is the total leveraged wealth of firms that is invested in risky assets. The constraint is motivated by the requirement that these two quantities are the same. Another observation is that an economy, according to Definition 16, does not specify the holding relationship between assets and firms in the financial system. The following definition characterizes the feasible holding networks for a given economy.

Definition 17 (Feasible Holding Networks)**.** *Suppose we are given an economy (D,q,w,b). We denote the set of feasible holding networks by $\mathcal{X}$, defined as*

$$\mathcal{X} = \left\{ X \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|} : Dq = X \operatorname{diag}(b) w,\ \mathbf{1}_{\mathcal{A}}^{\top} X = \mathbf{1}_{\mathcal{F}}^{\top}, X \geq 0 \right\}.$$

We notice that in the above definition, the condition $Dq = X\,\mathrm{diag}(b)w$ is just a re-formulation of the market clearing condition. This requires the balance between the supply and demand of fund in the market. Namely the market capitalization Dq, through a holding network X, should match the wealth invested in risky assets $\mathrm{diag}(b)w$.

One quick observation is that $\mathbf{1}_{\mathcal{A}}^\top Dq = w^\top b$ if and only if $\mathcal{X}$ is non-empty. To see this, note that if $\mathcal{X}$ is non-empty, clearly $\mathbf{1}_{\mathcal{A}}^\top Dq = \mathbf{1}_{\mathcal{A}}^\top X\,\mathrm{diag}(b)w = \mathbf{1}_{\mathcal{F}}^\top\,\mathrm{diag}(b)w = w^\top b$. In fact, the converse is also true. When $\mathbf{1}_{\mathcal{A}}^\top Dq = w^\top b$, define $m = Dq$. Also we define the total volume of the market as $v = 1_{\mathcal{A}}^\top Dq = w^\top b$. Take $X = \frac{m\mathbf{1}_{\mathcal{F}}^\top}{v}$ as one feasible holding network.

3.5.2 Optimal Holding Networks

Suppose we are given an economy. We are interested in the optimal structure of the holding network that keeps a minimum level of systemic risk and contagion. The concrete problem is to minimize MNA by choosing the holding network X from the feasible set $\mathcal{X}$. This problem, following Theorem 10, can be formulated as an optimization problem

$$\min_{X\in\mathcal{X}} \mathsf{MNA}(X) = \min_{X\in\mathcal{X}} \max_{i\in\mathcal{A}} |1-\lambda_i(Y(X))|^{-1}. \tag{3.22}$$

For the purpose of solving the above problem, we introduce two quantities, $\underline{\lambda_{max}}$ and $\overline{\lambda_{min}}$, defined as

$$\underline{\lambda_{max}} \triangleq \min_{X\in\mathcal{X}} \lambda_{max}(Y(X)), \tag{3.23}$$

$$\overline{\lambda_{min}} \triangleq \max_{X\in\mathcal{X}} \lambda_{min}(Y(X)). \tag{3.24}$$

The following theorem relates these quantities:

Theorem 11.

$$\overline{\lambda_{min}} \leq \underline{\lambda_{max}}. \tag{3.25}$$

The proof can be found in the Appendix A.2.

Based on these two quantities, $\underline{\lambda_{max}}$ and $\overline{\lambda_{min}}$, we can define two economic regimes that present different features of contagion. One regime is what we call low-leverage economies and the other is high-leverage economies. Specifically, we have the following definition:

Definition 18 (Leverage Regimes)**.** *Suppose we are given an economy* (D,q,w,b)*. It is defined as a low-leverage economy if* $\underline{\lambda_{max}} < 1$*. It is defined as a high-leverage economy if* $\overline{\lambda_{min}} > 1$*.*

Theorem 11 ensures that the set of low-leverage economies is disjoint from the set of high-leverage economies. The reason why these two economic regimes are named low-leverage and high-leverage respectively will become clear after we introduce a quantity systemic leverage, as we will show that the definitions are closely related to the scale of leverage of the economy.

Definition 19 (Systemic Leverage)**.** *Suppose we are given an economy* (D,q,w,b)*. We define the systemic leverage of the economy as*

$$\gamma = \frac{\sum_{h \in \mathcal{F}} w_h b_h^2}{\sum_{h \in \mathcal{F}} w_h b_h}.$$

If we view the normalized total wealth as weights, $\bar{w}_h = \frac{w_h}{\sum_{h \in \mathcal{F}} w_h}$, then $\mathbf{1}_{|\mathcal{F}|}^{\top} \bar{w} = 1$ and $\bar{w} \geq 0$. It is clear that $\bar{w}$ can be interpreted as a probability distribution. We can re-write γ as

$$\gamma = \frac{\mathsf{E}_{\bar{w}}[b_h^2]}{\mathsf{E}_{\bar{w}}[b_h]} = \mathsf{E}_{\bar{w}}[b_h](1 + \frac{\mathrm{Var}_{\bar{w}}[b_h]}{\mathsf{E}_{\bar{w}}[b_h]^2}).$$

Under the probability distribution $\bar{w}$, $\mathsf{E}_{\bar{w}}[b_h]$ is the mean of the leverage, and $\frac{\mathrm{Var}_{\bar{w}}[b_h]}{\mathsf{E}_{\bar{w}}[b_h]^2}$ is the squared coefficient of variation of the leverage. This relation implies that the systemic leverage γ is larger than the average wealth-weighted leverage, by an amount that characterizes the variability of leverages across firms. In other words, γ is large if the average wealth-weighted leverage is high. When the average leverage is fixed, if the coefficient of variation of the wealth-weighted leverage is higher, γ is larger.

Now we have the following observation that links $\underline{\lambda_{max}}$ with γ. The proof of Theorem 12 can be found in Appendix A.2.

Theorem 12. *In any economy,*

$$\underline{\lambda_{max}} = \gamma. \tag{3.26}$$

The result in Theorem 12 indicates that when the systemic leverage $\gamma < 1$ the economy is a low-leverage one. This fact provides a more convenient way to test if the state of an economy is low-leverage, since the γ can be computed by only the primitives the wealth vector w and the leverage vector b. This fact motivates Definition 18 for a low-leverage economy.

In what follows, we will demonstrate analytical results for both low-leverage and high-leverage economies. For the economic regimes that are neither low-leverage nor high-leverage economies, however, there are extra technical challenges that our current model can not accommodate. The difficulties include the fact that the minimum level of the maximal network amplifier cannot be simplified as what we will show for low-leverage and high-leverage economies. The main purpose of this model is not trying to provide analytical results that cover the entire spectrum of possible economic states. Instead we wish to show the counterintuitive fact that the optimal structures of financial networks present completely different features for the low-leverage and high-leverage economies.

Specifically, in the following section, we will try to solve the minimization problem for the maximal network amplifier as in equation (3.22) for the low-leverage economy and the high-leverage economy. For each case, we will characterize what the optimal structure of holding network X is.

3.5.3 Low-Leverage Economy

For a low-leverage economy, the problem of minimizing the maximal network amplifier in equation (3.22) can be simplified to

$$\min_{X \in \mathcal{X}} \mathsf{MNA}(X) = \frac{1}{1 - \underline{\lambda_{max}}}. \tag{3.27}$$

The following theorem characterizes the optimal network configuration that minimizes the maximal network amplifier for the low-leverage economy. The proof is in the Appendix A.2.

Theorem 13. *Suppose we are given a low-leverage economy. The minimum value of the maximal network amplifier can be achieved by a mutual-fund economy. Namely,*

$$X^* \triangleq x^* \mathbf{1}_{\mathcal{F}}^\top \in \operatorname*{argmin}_{X \in \mathcal{X}} MNA(X),$$

where

$$x^* = \frac{Dq}{\mathbf{1}_{\mathcal{A}}^{\top} Dq}.$$

For the holding network X^*, every firm invests in the same market portfolio, where assets are held in proportion to the market capitalization of assets.

Therefore, one of the optimally constructed holding networks for the leverage economy is where each firm invests into a common market portfolio. In this case, the optimal network drives all the firms to diversify their investment across assets proportional to the market capitalization of assets.

Now we consider another special economy that presents interesting features. When firms have uniform leverage levels, the following theorem provides an optimal holding network. The proof is in the Appendix A.2.

Theorem 14. *For a low-leverage economy, suppose every firm has a constant leverage level $0 < \bar{b} < 1$, and $b = \bar{b}\mathbf{1}_{\mathcal{F}}$. Suppose the holding network $X \in \mathcal{X}$ is a non-negative matrix. Then the maximal network amplifier does not depend on the configuration of the holding network X.*

3.5.4 High-Leverage Economy

For a high-leverage economy, the problem of minimizing the maximal network amplifier in equation (3.22) can be simplified to

$$\min_{X \in \mathcal{X}} \mathsf{MNA}(X) = \frac{1}{\overline{\overline{\lambda_{min}}} - 1}. \tag{3.28}$$

The following theorem establishes a bound for the solution. The proof can be found in the Appendix A.2.

Theorem 15. *Suppose we are given a high-leverage economy. The minimum value of the maximal network amplifier is bounded below by* $\frac{1}{\gamma-1}$.

$$\min_{X \in X} MNA(X) = \frac{1}{\overline{\lambda_{min}} - 1} \geq \frac{1}{\gamma - 1}.$$

In the following part, we will demonstrate that the lower bound of the maximal network amplifier is achievable for some cases. For those cases when the bound is achievable, we will prove that isolated economies have the optimal structure as holding networks, where firms all invest different portfolio with each other.

Completely Symmetric Case

In this section, we examine a special case where firms are symmetric and assets are also symmetric. Again, our objective is to solve for the optimal network X that minimizes the maximal network amplifier. The following theorem solves the optimization problem. The proof can be found in the Appendix A.2.

Theorem 16. *Suppose the leverage is* $b = \bar{b}\mathbf{1}_{\mathcal{F}}$ *where* $\bar{b} > 1$*, wealth is* $w = \bar{w}\mathbf{1}_{\mathcal{F}}$*, and the market capitalization is* $m = \bar{m}\mathbf{1}_{\mathcal{A}}$*. We also assume that* $|\mathcal{F}| = n|\mathcal{A}|$*, where* $n \in N$*. The following network*

X^* minimizes the maximal network amplifier.

$$X^* = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \tag{3.29}$$

Specifically,

$$X^*_{ih} = \begin{cases} 1 \quad , & \text{if } (i-1)n+1 \le h \le in, \\ 0 \quad , & \text{otherwise}. \end{cases} \tag{3.30}$$

Remarks. Intuitively, this solution X^* defined by equation (3.30) has a special structure. It fills up the market capitalization of each asset with the full leveraged wealth of firms. In X^*, each firm only invests in one asset. So the holding network is partitioned into separate components, where each component corresponds to a group of firms that invest in a common asset. For high-leverage economy, this isolated holding network achieves the minimum level of the maximal network amplifier. We name the holding network X^* defined by equation (3.30) as a **fully isolated holding network**.

General Case

In the previous section, the fully isolated holding network defined by equation (3.30) identifies that the isolation of assets is of great importance to minimize the maximal network amplifier. However, for general parameters D, q, w, and b, it is usually impossible to construct a fully isolated holding network. In this section, we propose a construction of the holding network similar to a fully isolated holding network. We show that this construction provides an asymptotically optimal solution that minimizes the maximal network amplifier.

The problem we consider is to assign the leveraged wealth of firms to match up with the market capitalization of assets. In the construction defined by equation (3.30), we observe that there are firms that invest their full leveraged wealth in only one asset. Let $\mathcal{F}_i$ denote the set of firms that invest only in asset i. By this construction, we know that $\mathcal{F}_i \cap \mathcal{F}_j = \phi$ for any i and j in $\mathcal{A}$. In the general case, it is impossible to obtain a partition such that $\mathcal{F} = \cup_{i \in \mathcal{A}} \mathcal{F}_i$ while keeping it a feasible holding network. So for this section, we define $\mathcal{R} = \mathcal{F} - \cup_{i \in \mathcal{A}} \mathcal{F}_i$, which is the remaining firms that are available for the assignment of investment. The firms in $\mathcal{R}$ can contribute to match the total market capitalization of assets and total leveraged wealth available for investment. Consequently, we now have constructed a partition of $\mathcal{F}$, denoted by $\mathcal{P} = \{\mathcal{F}_i\}_{i \in \mathcal{A}} + \mathcal{R}$. The introduction of $\mathcal{R}$ generalizes equation (3.30). We will soon show, however, the impact of the remaining firms in $\mathcal{R}$ vanishes in an asymptotic regime.

Let $m_i = D_{ii} q_i$ denote the market capitalization of asset i for all $i \in \mathcal{A}$. Given a partition $\mathcal{P}$ of $\mathcal{F}$, we can calculate the mismatch between the market capitalization and the assigned investment for each asset, which is denoted by the vector $m' \in \mathcal{A}$. Specifically, $m'_i = m_i - \sum_{h \in \mathcal{F}_i} w_h b_h$, which is the gap between the market capitalization m_i of asset i and the part that is covered by firms in $\mathcal{F}_i$. To cover this mismatch m', firms in $\mathcal{R}$ take a common portfolio which is proportional to m'. Now we can obtain the following holding network.

$$X(\mathcal{P}) = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 & \alpha m_1' & \alpha m_1' & \dots & \alpha m_1' \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & \alpha m_2' & \alpha m_2' & \dots & \alpha m_2' \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & \alpha m_{|\mathcal{A}|}' & \alpha m_{|\mathcal{A}|}' & \dots & \alpha m_{|\mathcal{A}|}' \end{pmatrix}, \tag{3.31}$$

where $\alpha = \frac{1}{\mathbf{1}_{|\mathcal{A}|}^{\top} m'}$. Specifically,

$$X_{ih}(\mathcal{P}) = \begin{cases} 1 & , \quad \text{if } h \in \mathcal{F}_i, \\ \alpha m_i' & , \quad \text{if } h \in \mathcal{R}, \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{3.32}$$

We notice that although m' may have negative elements, but the quantity

$$\mathbf{1}_{|\mathcal{A}|}^{\top} m' = \sum_{i \in \mathcal{A}} m_i - \sum_{i \in \mathcal{A}} \sum_{h \in \mathcal{F}_i} w_h b_h \geq \sum_{i \in \mathcal{A}} m_i - \sum_{h \in \mathcal{F}} w_h b_h = 0,$$

and it is always non-negative.

It can be easily verified that the construction of $X(\mathcal{P})$ satisfies that $Dq = X \operatorname{diag}(b) w$, and $\mathbf{1}_{\mathcal{A}}^{\top} X(\mathcal{P}) = \mathbf{1}_{\mathcal{F}}^{\top}$. But whether $X(\mathcal{P}) \geq 0$ depends on the choice of the partition $\mathcal{P}$. In general, $X(\mathcal{P})$ may not be a feasible holding network, and it may not minimize the maximal network amplifier. In the following, however, we wish to show that there exists an asymptotic regime in which asymptotically $X(\mathcal{P})$ is a feasible and an optimal solution to the problem of minimizing the maximal network amplifier.

Now we consider a setting where the parameters that describe the economy are randomly drawn from given distributions. Suppose for each firm h, the leverage and the wealth of the firm (b_h, w_h) is an i.i.d. random variable drawn from a joint distribution for (b_h, w_h). Suppose the means $\bar{b} = E[b_h]$ and $\bar{w} = E[w_h]$ are both finite, and the second moments $E[b_h^2]$, $E[w_h^2]$, and $E[w_h b_h]$ are also finite. Also, the joint distribution of (b_h, w_h) has a finite third and fourth moment. In particular, $\sigma^2_{b_h w_h}$ is the variance of $b_h w_h$. In addition, we assume that $\bar{\gamma} = \frac{E[w_h b_h^2]}{E[w_h b_h]} > 1$.

Suppose we are given the proportions of the market capitalization of each asset i, denoted by $\bar{m}_i$. We require that $\sum_{i\in\mathcal{A}} \bar{m}_i = 1$ and $\bar{m}_i \geq 0$ for any $i \in \mathcal{A}$. We can calculate the market capitalization of each asset as

$$m_i = \bar{m}_i \sum_{h\in\mathcal{F}} w_h b_h, \quad \forall\, i \in \mathcal{A}.$$

In the following theorem, we will show that the construction of $X(\mathcal{P})$ inspired by the isolated holding network provides an optimal solution to minimize the maximal network amplifier asymptotically. Let $A = |\mathcal{A}|$ and $F = |\mathcal{F}|$. The complete proof of Theorem 17 can be found in Appendix A.2.

Theorem 17. *Suppose A is fixed and we take $F \to +\infty$. For any given F, we generate a set of firms $\mathcal{F}$ where $|\mathcal{F}| = F$. The leverage and wealth of $\mathcal{F}$ satisfy the setting in this section. Indexed by F, we can construct a sequence of sets of feasible holding networks $X^{(F)}$. Then, indexed by F, there exists a sequence of partitions $\mathcal{P}^{(F)}$ of $\mathcal{F}$ that defines a sequence of holding networks $X(\mathcal{P}^F)$ according to equation* (3.32) *such that the sequence of $X(\mathcal{P}^{(F)})$ is asymptotically the optimal solution to the problem of minimizing the maximal network amplifier. Namely, as $F \to +\infty$, $X(\mathcal{P}^F)$*

is asymptotically a feasible solution.

$$X(\mathcal{P}^F) \in \mathcal{X}^{(F)}, \quad \textit{for all but finitely many } F\textit{, almost surely}.$$

Also, $X(\mathcal{P}^F)$ *provides the optimality asymptotically.*

$$\mathsf{MNA}(X(\mathcal{P}^F)) - \min_{X \in \mathcal{X}^{(F)}} \mathsf{MNA}(X) \to_P 0.$$

Proof Sketch. To prove the theorem, we will construct the sequence of partitions $\mathcal{P}^F$ of $\mathcal{F}$. This theorem has two parts to prove. The first part is to prove that the sequence of $X(\mathcal{P}^F)$ asymptotically belongs to feasible holding networks $\mathcal{X}^{(F)}$. The second part is to prove that the sequence of $X(\mathcal{P}^F)$ asymptotically produces the optimal solution.

For a given F, we construct a partition $\mathcal{P}^F$ of $\mathcal{F}$. The partition consists of non-intersecting subsets of $\mathcal{F}$ specified by $\mathcal{P}^F = \{\mathcal{F}_i\}_{i\in\mathcal{A}} + \mathcal{R}$. Here $\mathcal{F}_i \subseteq \mathcal{F}$ denotes firms which invest in each asset i, and $\mathcal{R} \subseteq \mathcal{F}$ denotes the remaining firms. In this construction, only the cardinality $|\mathcal{F}_i|$ of $\mathcal{F}_i$ matters for this proof. So we first decide the cardinality of $\mathcal{F}_i$ for each $i \in \mathcal{A}$.

$$|\mathcal{F}_i| = \max\{\lceil \bar{m}_i F \rceil - 1 - \lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor, 0\}. \tag{3.33}$$

where $\varepsilon \in (0, \frac{1}{2})$ is a small number that is arbitrarily chosen and fixed. It is clear that $|\mathcal{F}_i| < \bar{m}_i F$, so $\sum_{i\in\mathcal{A}}|\mathcal{F}_i| \leq F$. For each $i \in \mathcal{A}$, we define

$$\begin{aligned} \mathcal{F}_i &= \{h \in \mathcal{F} : \textstyle\sum_{j=1}^{i-1}|\mathcal{F}_j| + 1 \leq h \leq \sum_{j=1}^{i}|\mathcal{F}_j|\}, \\ \mathcal{R} &= \mathcal{F} \setminus \textstyle\bigcup_{i\in\mathcal{A}}\mathcal{F}_i. \end{aligned} \tag{3.34}$$

Now based on the partition $\mathcal{P}^F$, we can construct $X(\mathcal{P}^F)$ according to equation (3.32). With this particular construction, we show that $|\mathcal{F}_i|$ approaches to $\bar{m}_i F$. Then it can be shown that $\frac{|\mathcal{R}|}{F} \to 0$. With some technical derivation, we can prove the theorem. More details can be found in the Appendix A.2. ∎

Remarks. In equation (3.33), $|\mathcal{F}_i|$ defines the number of firms that invest all their leveraged wealth into asset i. The remaining firms $\mathcal{R}$ are defined accordingly in equation (3.34). In the proof, we notice that for this construction $\mathcal{R}$ has a negligible portion of all the firms since $\frac{|\mathcal{R}|}{F} \to 0$. Therefore, the construction defined by equation (3.33) asymptotically becomes a fully isolated holding network. This asymptotically optimal solution shares the similar feature of isolated investment with the complete symmetric case in Section 3.5.4. This indicates that an asymptotically isolated holding network provides the optimal network for the general case.

3.5.5 Diversification versus Diversity

In previous sections, we have solved the optimal holding network that minimizes the maximal network amplifier, for the low-leverage case and high-leverage case respectively. The optimal solutions in the two cases demonstrate very different features. First, for the low-leverage case, the optimal solution aligns with the mutual fund economy where each firm invests in a common portfolio. This structure of network agrees with the conventional diversification argument in risk management. Namely, firms should diversify investment in the market portfolio to benefit from the reduction in the idiosyncratic risk across assets. From our analysis for the low-leverage case, it is also the optimal from the perspective of the benefit of the financial system. On the other hand, however, for the high-leverage case, the optimal network is no longer the mutual fund economy. Instead, the isolated holding network becomes the best to the financial system. In the isolated structure, firms are partitioned into groups, where each group i invests all of their leveraged wealth into only asset i. In this case, the investment behavior of each group of firms tend to be as different from other groups as possible. This diversity feature of the isolated holding network is complete opposite of the diversity feature in the mutual fund economy. Now we identify that the low-leverage case and the high-leverage case present completely opposite natures in terms of reducing the contagion effects. In other words, we wish to point out that the diversification across assets is more valued in the low-leverage case, but the diversity across firms is more important in the high-leverage case.

From the regulator's perspective, it is critical to determine whether the current state of the economy is low-leverage and high-leverage, since for these two states the driving forces behind systemic risk demonstrate completely different and even opposite features. By monitoring wealth and leverage of all the firms, the regulator can compute the systemic leverage γ. According to the analysis in this section, γ provides a good guideline to deciding which state the economy is. We point out that this type of monitoring does not require the knowledge of detailed portfolio positions of each firm. Instead, only two numbers, wealth and leverage, need to be reported to the regulator. In fact, this information is often available to the public. Thus each investor can also calculate the systemic leverage γ and decide the state of the economy. With the understanding of systemic risk caused by contagion, investors can take preventive actions before an outbreak of systemic events.

Chapter 4

Factor Decomposition for Risk Measures

4.1 Introduction

This chapter addresses another interesting topics of systemic risk: how to estimate risk measures based on risk factors. Since the true value of a particular risk measure is difficult to calculate, since the computation requires the complete knowledge of random outcomes of future scenarios. However, based on a set of risk factors that have been well understood and analyzed, it is possible to construct approximation of the original risk measures. This question leads to several interesting research problems. One is that given the realizations of risk factors, what types of functions of factor-based outcomes can we use to construct a risk measure as the estimate of the original measure? Another question is that given risk factors and the original risk measure, how we can construct factor-based risk measures that provides tightest approximations, including the best under-estimator and the best over-estimator. In this chapter, we have the following main contributions.

- **We provide a framework to analyze risk measures based on factors.**

  In this framework, we define a form of factor-separable risk measures. With factor-separable risk measures, the acceptance set of the original risk measure is connected with the acceptance sets of the factor-separable risk measures. We allow for approximation of risk measures with given factors.

- **We provide tight bounds for factor-separable coherent risk measures.**

  In particular, through the analysis of acceptance sets of factor-separable risk measures, we can provide closed form expressions for upper and lower bound of the original risk measure. We show that for factor-separable coherent risk measures, we can construct a tight upper bound. For lower bounds, we can provide a Pareto optimal construction.

The framework of coherent risk measures has been introduced by Artzner et al. [1999]. This chapter connects risk factors with the framework of risk measures. The introduction of risk factors closely connects to the field of dynamic risk measures since revealing values of factors is a type of information revealing in a dynamic setting. In the literature, Roorda et al. [2005] introduces an axiom of consistency for a multi-period setting. This axiom, in the setting of risk factors, becomes a factor consistency properties for risk measures. The detailed discussion of consistency properties can be found in Penner [2007] and Acciaio and Penner [2011]. Another perspective that has been studies is risk factor contributions. For example, Rosen and Saunders [2010] apply regression techniques to analyze risk factors. The article that is the closest to the results in this chapter is an independent work by Iancu et al. [2011], who provides tight approximations for dynamic risk measures. It can be shown that the tight upper bound they independently construct is the same as the one we provide. But we will demonstrate the construction of tight lower bounds which have not been discovered in the literature.

4.2 Factor-Separable Risk Measures

Suppose Ω is the set of finite future scenarios. For each scenario $\omega \in \Omega$, v_ω is the probability of scenario ω happening. Let $f \in \mathbb{R}^{|\Omega|}$ denote future outcomes of a risk factor. For simplicity, we only introduce one risk factor for the setting, but you will soon find out that all the results of this paper can be easily extended to a setting with a set of risk factors. There are $m \leq |\Omega|$ values that a risk factor f can take. These values are $\{v_i\}_{i=1}^m$. According to the values of risk factors the set of future scenarios Ω can be partitioned into subsets as follows.

$$\Omega = \cup_{i=1}^m \Omega_i,$$

where

$$\Omega_i = \{\omega \in \Omega \colon f_\omega = v_i\}, \quad \forall\, i = 1, 2, \ldots, m.$$

For each i, let $u_i \in \mathbb{R}^{|\Omega_i|}$ denote the projection of the probability measure v onto Ω_i. Namely for $\omega \in \Omega_i$,

$$u_i(\omega) = \frac{v_\omega}{\sum\limits_{\omega' \in \Omega_i} v_{\omega'}}.$$

For each i, $\mathbf{1}_{\Omega_i} \in \mathbb{R}^{|\Omega|}$ is a vector whose element $\mathbf{1}_{\Omega_i}(\omega) = 1$ if $\omega \in \Omega_i$, and $\mathbf{1}_{\Omega_i}(\omega) = 0$ otherwise. Define $M_{\Omega_i} \in \mathbb{R}^{|\Omega| \times |\Omega_i|}$ as a matrix whose element $M_{\Omega_i}(\omega, i) = 1$ if $\omega - i + 1 \in \Omega_i$, and $M_{\Omega_i}(\omega, i) = 0$ otherwise. So matrix M_{Ω_i} can map a vector in $\mathbb{R}^{|\Omega_i|}$ to a vector in $\mathbb{R}^{|\Omega|}$.

Suppose $\rho \colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ is a coherent single-firm risk measure. The acceptance set of ρ is $A = \left\{X \in \mathbb{R}^{|\Omega|} \colon \rho(X) \leq 0\right\}$ (See Artzner et al. [1999]). Now we introduce factor-separable risk measures as follows.

Definition 20 (Factor-separable Risk Measure)**.** *Given the risk factor f and a coherent risk measure ρ. Then ρ^f is a factor-separable risk measure if*

$$\rho^f(X) = \rho(\sum_{i=1}^{m} \rho_i(X)\mathbf{1}_{\Omega_i}), \tag{4.1}$$

where ρ_i is a coherent risk measure for each i.

According to the construction in equation (4.1), it can be easily shown that ρ^f is a coherent risk measure. Our definition of factor-separable risk measures is related to dynamic risk measures recently introduced in the literature. While a dynamic risk measure emphasizes the relationship among the risk measurement at differen time periods, a factor-separable risk measure deals with how risk measures given different factor values are related. We can interpret a factor-separable risk measure as a dynamic risk measures with only two periods. The first period is when factor values are not revealed, and the second period is when factor values are revealed. One important concept in the literature of dynamic risk measures is the time consistency property. In the setting with risk factors, the time consistency property, for a two period model with factor values revealing, becomes the "factor" consistency. In fact, it is clear that the definition of factor-separable risk measures satisfies the time consistency. See Penner [2007] and Acciaio and Penner [2011].

Suppose we are given the risk factor f and the risk measure ρ, our goal is to construct a factor-separable risk measure ρ^f such that it is either an overestimate or an underestimate of ρ. In other words, we require ρ^f to satisfy one of the following conditions.

- The condition for overestimation:

$$\rho^f(X) \geq \rho(X), \quad \forall X \in \mathbb{R}^{|\Omega|},$$

- The condition for underestimation:

$$\rho^f(X) \leq \rho(X), \quad \forall X \in \mathbb{R}^{|\Omega|}.$$

In the following section, we will present the constructions for both cases.

4.3 Acceptance Sets with Risk Factors

Recall that a risk measure can be characterized by its acceptance set introduced by Artzner et al. [1999]. In this section, we will work in the domain of acceptance sets since a better geometric interpretation can be provided for our analysis. For a given convex risk measure ρ, let A denote its acceptance set. The relationship between ρ and A is the following.

$$A = \{x : \rho(x) \leq 0\}, \qquad \rho(x) = \inf\{t : x - t\mathbf{1}_\Omega \in A\}. \tag{4.2}$$

For a factor-separable risk measure, its acceptance set takes a special form. Let $A_1(i)$ denote the acceptance set of ρ_i for $i = 1, 2, \ldots, m$, $A_1(i)$. Define an acceptance set A_0 as follows.

$$A_0 = \left\{ X = \sum_{i=1}^{m} t_i \mathbf{1}_{\Omega_i} : \ \rho(X) \le 0 \right\}.$$

Since ρ is a coherent risk measure, it is clear that A_0 is an acceptance set since it satisfies all the properties of acceptance sets, including monotonicity, convexity, and positive homogeneity. Then we can define a risk measure ρ_0 whose acceptance set is A_0.

Theorem 18. *Suppose we are given a factor-separable risk measure ρ^f, which has the following form.* ***TODO:*** *X_i?*

$$\rho^f(X) = \rho(\sum_{i=1}^{m} \rho_i(X) \mathbf{1}_{\Omega_i}).$$

Then the acceptance set of ρ^f has the following form.

$$A^f = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m).$$

Proof. According to equation (4.2), the acceptance set of ρ^f is

$$B^f = \left\{ x : \ \rho^f(x) \le 0 \right\}.$$

To prove the theorem, we will show that for any $a \in A^f$, $a \in B^f$, and for any $b \in B^f$, $b \in A^f$. First, for any $a \in A^f$, there exists $a^0 \in A_0$ and $a^i \in A_1(i)$ for any $i = 1, 2, \ldots, m$, such that

$$a = a^0 + \sum_{i=1}^{m} M_{\Omega_i} a^i.$$

For each $j = 1, 2, \ldots, m$,

$$\rho_j(a_j) = a_j^0 + \rho_j(a_j) \leq a_j^0.$$

The above inequality is because $a_j \in A_1(j)$ is equivalent to $\rho_j(a_j) \leq 0$. Therefore,

$$\rho^f(a) = \rho(\sum_{j=1}^{m} \rho_j(a_j) \mathbf{1}_{\Omega_j}) \leq \rho(\sum_{j=1}^{m} M_{\Omega_j} a_j^0) = \rho(a^0) \leq 0.$$

In the above derivation, the first inequality is because of the monotonicity of coherent risk measures. The last inequality is because $a^0 \in A_0$ implies $\rho(a_0) \leq 0$. Hence $a \in B^f$.

Second, for any $b \in B^f$, take

$$b^0 = \sum_{j=1}^{m} \rho_j(b_j) \mathbf{1}_{\Omega_j}.$$

Since $b \in B^f$, we have $\rho(b^0) \leq 0$. It follows that $b^0 \in A_0$. Take $b^i = b_i - b_i^0$. We know that

$$\rho_i(b^i) = \rho_i(b_i - b_i^0) = \rho_i(b_i - \rho_i(b_i) \mathbf{1}_{\Omega_i}) = \rho_i(b_i) - \rho_i(b_i) = 0.$$

In the above derivation, the second last equality is because of the cash invariance property of coherent risk measures. So we have obtained $b^0 \in A_0$ and $b^i \in A_1(i)$ for any $i = 1, 2, \ldots, m$, such that

$$b = b^0 + \sum_{i=1}^{m} M_{\Omega_i} b^i.$$

Hence $b \in A^f$.

Therefore $A^f = B^f$. ■

Theorem 18 states that the special structure of a factor-separable risk measure can be translated to the structure of its acceptance set. In the following, we show that the overestimation or underestimation of risk measures can also be translated to relationships between acceptance sets.

Theorem 19. *Suppose* $\rho_\alpha \colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ *and* $\rho_\beta \colon \mathbb{R}^{|\Omega|} \to \mathbb{R}$ *are two coherent risk measures. The acceptance sets of* ρ_α *and* ρ_β *are* A_α *and* A_β *respectively. Then*

$$\rho_\alpha(X) \le \rho_\beta(X), \quad \forall\, X \in \mathbb{R}^{|\Omega|},$$

if and only if

$$A_\alpha \supseteq A_\beta.$$

For a factor-separable risk measure ρ^f to be an overestimate or an underestimate of ρ, we have the following corollaries.

Corollary 3. *Suppose we are given a factor-separable risk measure* ρ^f*, which has the following form.*

$$\rho^f(X) = \rho(\sum_{i=1}^{m} \rho_i(X)\mathbf{1}_{\Omega_i}).$$

Then we have

- ρ^f *is an overestimate of* ρ *if and only if*

$$A^f = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m) \subseteq A, \tag{4.3}$$

- ρ^f *is an underestimate of* ρ *if and only if*

$$A^f = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m) \supseteq A, \tag{4.4}$$

Proof. The corollary immediately follows from Theorem 18 and Theorem 19. ■

The above corollaries have established important geometric facts that we can apply in the following analysis. The corollary 3 translates the upper and lower bound conditions for risk measures to a subset or superset condition in the domain of acceptance sets. Consequently, the construction of upper bounds and lower bounds for given risk measures leads naturally to the construction of subsets and supersets for given acceptance sets.

In the following section, for the analysis of an overestimate and an underestimate of ρ, we use a different set of notations with extra script. In particular, the acceptance set of a factor-separable risk measure ρ^f that is an upper bound of ρ is the following.

$$A^U = A_0 + A_1^U(1) \times A_1^U(2) \times \ldots \times A_1^U(m).$$

The the acceptance set of a factor-separable risk measure ρ^f that is a lower bound of ρ is the following.

$$A^L = A_0 + A_1^L(1) \times A_1^L(2) \times \ldots \times A_1^L(m).$$

4.4 Upper Bounding Risk Measures

In this section, we will construct a factor-separable risk measure that is an overestimate for a given risk measure. To construct an upper bound, we define the acceptance set as follows.

$$A_1^U(i) = \left\{ x_i \in \mathbb{R}^{|\Omega_i|} : x_i \mathbf{1}_{\Omega_i} \in A \right\}, \quad \forall\, i. \tag{4.5}$$

Then we can define the risk measure ρ_i^U according the acceptance set in equation (4.5).

$$\rho_i^U(X) = \inf\left\{ t \in \mathbb{R} :\ X - t\mathbf{1}_{\Omega_i} \in A_1^U(i) \right\}, \quad \forall\, i. \tag{4.6}$$

In the above construction, it is clear that we can simplify ρ_i^U as follows.

$$\rho_i^U(X) = \inf\{ t \in \mathbb{R} :\ \rho((X-t)\mathbf{1}_{\Omega_i}) \le 0 \}, \quad \forall\, i. \tag{4.7}$$

The above formulation is a primal representation that connects the original risk measure ρ to the newly constructed measure. In fact, there is also a dual representation that makes the connection. Specifically, the dual representations for ρ_i^U and ρ are as follows.

$$\rho_i^U(x) = \sup_{q_i \in Q_i} \mathsf{E}_{q_i}[x_i], \quad \rho(x) = \sup_{q \in Q} \mathsf{E}_q[x],$$

for some sets of probability measures Q_i and Q. By using the dual relationship between acceptance sets and risk measures, Q_i connects to Q through the following relationship.

$$Q_i = \left\{ p_i \in \mathbb{R}_+^{|\Omega_i|} : \exists p \in Q, \text{ s.t. } p_i(\omega) = \frac{p(\omega)}{\sum_{\omega \in \Omega_i} p(\omega)}, \ \forall \omega \in \Omega_i \right\}.$$

Namely, Q_i is the set of conditional probability measures from Q. Now we are ready to construct an upper bound for ρ.

Theorem 20 (Upper Bound). *Suppose we construct a factor-separable risk measure as follows.*

$$\rho^U(X) = \rho_0(\sum_{i=1}^{m} \rho_i^U(X)\mathbf{1}_{\Omega_i}), \tag{4.8}$$

where for any i, ρ_i^U *is defined by equation* (4.6)*. Then for any outcome* $X \in \mathbb{R}^{|\Omega|}$*, we have*

$$\rho^U(X) \geq \rho(X).$$

Proof. We wish to show $A^U = A_0 + A_1^U(1) \times A_1^U(2) \times \ldots \times A_1^U(m) \subseteq A$, which is equivalent to the result. For any $x \in A^U$, there exist $x_0 \in A_0$ and $x_i \in A_1^U(i)$ for all i, such that $x = x_0 + \sum_i x_i \mathbf{1}_{\Omega_i}$. From the definition of A_0^U, we know that $x_0 \in A$. Also, from the definition of $A_1^U(i)$, we know that $M_{\Omega_i} x_i \in A$ for all i. Since set A is a cone, the additivity of elements of a cone implies that $x \in A$. ■

The above theorem 20 establishes that ρ^U is an upper bound for ρ. The following theorem shows that ρ^U gives the best upper bound.

Theorem 21 (Best Upper Bound). *Suppose* ρ^U *is defined according to equation* (4.8)*. For another* *group of risk measures* ρ_i, *we construct a factor-separable risk measure as follows.*

$$\rho'(X) = \rho_0(\sum_{i=1}^{m} \rho_i(X)\mathbf{1}_{\Omega_i}).$$

If $\rho'(X)$ *produces an upper bound for* ρ, *then*

$$\rho'(X) \geq \rho^U(X), \quad \forall X \in \mathbb{R}^{|\Omega|}. \tag{4.9}$$

Proof. We prove this theorem by contradiction. Assume that there exists an outcome X that breaks equation 4.9.

Then there exists an element $x = X - \rho'(X)\mathbf{1}_\Omega$ such that

$$x \notin A^U = A_0 + A_1^U(1) \times A_1^U(2) \times \ldots \times A_1^U(m),$$

but

$$x \in A' = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m),$$

where $A_1(i)$ is the acceptance set for ρ_i. We notice that A^U and A' share the common cone A_0 by construction. The existence of such x implies that, there exists an i and an a_i such that $a_i \notin A_1^U(i)$ but $a_i \in A_1(i)$.

Since ρ_i gives an upper bound for ρ according to Theorem 20, we have that $A' = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m) \subseteq A$.

Take the $\mathbf{0}$ element from set $A_1(j)$ for $j \neq i$, and take a_i from $A_1(i)$. We have $a_i \mathbf{1}_{\Omega_i} \in A'$ and thus $a_i \mathbf{1}_{\Omega_i} \in A$.

According to the definition for $A_1^U(i)$ in equation (4.5), we obtain that $a_i \in A_1^U(i)$. This leads to the contradiction.

■

4.4.1 Maximum Loss

Besides the best upper bound we have constructed, other forms of ρ_i can also produce an upper bound for ρ. Among them, one simple construction is the maximum loss function max. Specifically, we can define

$$\rho_i^{max}(x_i) = \max_j \{x_i(j)\}.$$

For each i, the acceptance set of ρ_i^{max} is

$$A_1^{max}(i) = \left\{ x_i \in \mathbb{R}^{|\Omega_i|} : x_i \leq \mathbf{0} \right\}. \tag{4.10}$$

We can define a factor-separable risk measure based on the maximum loss function.

Definition 21 (Maximum Loss Risk Estimator). *Suppose risk measure ρ_i^{max} is defined by its acceptance set in equation 4.10 for each i. The maximum loss risk estimator is the following factor-separable risk measure.*

$$\rho^{MAX} = \rho(\sum_{i=1}^{m} \rho_i^{max}(X) \mathbf{1}_{\Omega_i}).$$

Because of monotonicity of ρ and the acceptance set A, it is clear that equation (4.3) holds and

ρ^{MAX} produces an upper bound for ρ. Namely

$$\rho^{MAX} = \rho(\sum_{i=1}^{m} \rho_i^{max}(X)\mathbf{1}_{\Omega_i}).$$

For any outcome $X \in \mathbb{R}^{|\Omega|}$, we have

$$\rho^{MAX}(X) \geq \rho(X).$$

We notice that this simple construction of ρ_i^{max} is different from the best upper bound ρ_i^U. However, in some cases, these two may coincide. The following theorem gives a condition that makes it happen.

Theorem 22 (Maximum Loss). *Suppose for a given $i \in \{1,2,\ldots,m\}$, $\rho_i^U(x_i) \leq 0$ implies that $x_i \leq \mathbf{0}$. Then we have that*

$$\rho_i^U(x_i) = \rho_i^{max}(x_i), \quad \forall\, x_i \in \mathbb{R}^{|\Omega_i|}.$$

Proof. Since $\rho_i^U(x_i) \leq 0$ implies that $x_i \leq \mathbf{0}$, and $x_i \leq \mathbf{0}$ also implies $\rho_i^U(x_i) \leq 0$ due to monotonicity of the coherent risk measure ρ_i^U, we obtain the acceptance set of ρ_i^U as follows.

$$A_1^U(i) = \left\{ x_i \in \mathbb{R}^{|\Omega_i|} : x_i \leq \mathbf{0} \right\}$$

This is exactly the acceptance set of ρ_i^{max}. Thus, $\rho_i^U = \rho_i^{max}$. ■

When the hypothesis in Theorem 22 is satisfied, the construction of the risk measures for the upper bound degenerates to the maximum loss function. This can be interpreted as a robust representation of the original risk measures since the new construction picks the worst case scenario as the representative of a particular realization of the risk factor.

Example 18 (CVaR)**.** *Suppose we are given a coherent risk measure* $\rho = CVaR_\alpha$*, the conditional value at risk at* α *level. It is also known as expected shortfall, or average value at risk. For* $CVaR_\alpha$ *we can construct our factor-separable risk measure* ρ^U *through a group of risk measures* ρ_i^U*. In general,* ρ_i^U *is different from* ρ_i^{max}*.*

But for typical values of α *can be* 5%*,* 2.5%*, or* 1%*, we also have the following typical condition.*

$$P(\Omega_i) \leq 1 - \alpha, \quad \forall\, i.$$

This is usually the case since in one single scenario Ω_i*, the probability of it happening is relatively small, while the right-side of the inequality is a number above* 90%*. By the construction of* ρ_i^U *in equation* (4.6)*, it is clear that* $\rho_i^U(x_i) \leq 0$ *implies that* $x_i \leq \mathbf{0}$ *since the acceptance set of* $CVaR_\alpha$ *contains only non-positive elements. According to Theorem 22, we obtain that*

$$\rho_i^U(x_i) = \rho_i^{max}(x_i), \quad \forall\, x_i \in \mathbb{R}^{|\Omega_i|}.$$

In other words, for typical cases when using $\rho = CVaR_\alpha$*, the maximum loss function provides the best upper bound.*

4.5 Lower Bounding Risk Measures

In this section, we will construct a factor-separable risk measure that is an overestimate for a given risk measure. To construct an upper bound, we define the acceptance set as follows.

Through the factor decomposition, we propose that ρ_i^L is defined by its acceptance set as follows.

$$\rho_i^L(X) = \inf\left\{t \in \mathbb{R} : \ X - t\mathbf{1}_{\Omega_i} \in A_1^L(i)\right\}, \quad \forall\, i. \tag{4.11}$$

For any $i = 1, 2, \ldots, m$, the acceptance sets $A_1^L(i)$ is defined by the following recursive formula.

$$\begin{aligned} A_1^L(i) &= Cone \quad \left\{a_i \in \mathbb{R}^{|\Omega_i|} : \ \exists a_{-i} \in B_i, \ s.t. \ a_{-i} + a_i\mathbf{1}_{\Omega_i} \in A\right\}, \\ B_i &= \quad \left\{b \in A_0 + A_1^L(1) \times \ldots \times A_1^L(i-1) \times \mathbf{0} \ldots \times \mathbf{0} : \right. \\ & \quad\quad \left. b + \varepsilon\mathbf{1}_{\Omega_1} \notin A_0 + A_1^L(1) \times \ldots \times A_1^L(i-1) \times \mathbf{0} \ldots \times \mathbf{0}, \ \forall \varepsilon > 0\right\}. \end{aligned} \tag{4.12}$$

We notice that the boundary case for $i = 1$ is the following.

$$B_1 \quad = \quad \{b \in A_0 : b + \varepsilon\mathbf{1}_{\Omega_1} \notin A_0, \ \forall \varepsilon > 0\}. \tag{4.13}$$

In the above formula, $\{B_i\}_{i=1}^m$ works as a series of auxiliary sets for the construction of A_i^L. To better understand the construction of $A_1^L(i)$, we have the following lemma. For any k, define

$$C_k = \left\{c_k \in A : c_k = \sum_{i=1}^{k} M_{\Omega_i} x_i + \sum_{i=k+1}^{m} t_i\mathbf{1}_{\Omega_i}, \text{ where } x_i \in \mathbb{R}^{|\Omega_i|}, t_i \in \mathbb{R}, \forall i\right\}.$$

Note that C_k is a subset of A, which has deterministic values on certain pieces of scenarios.

Lemma 3. *For any $k = 1, 2, \ldots, m$,*

$$C_k \subseteq A_0 + A_1^L(1) \times \ldots \times A_1^L(k) \times \mathbf{0} \ldots \mathbf{0}.$$

Proof. We prove this lemma by induction. First, we want to show that $C_1 \subseteq A_0 + A_1^L(1) \times \mathbf{0} \ldots \mathbf{0}$. For any $c_1 = M_{\Omega_1} x_1 + \sum_{i=2}^{m} t_i \mathbf{1}_{\Omega_i} \in C_1$, there exists $s_1 \in \mathbb{R}$ such that

$$z_1 = c_1 - (x_1 - s_1)\mathbf{1}_{\Omega_1} \ \in B_1 \subseteq A_0.$$

Let $a_1 = (x_1 - s_1)$, we have $c_1 = z_1 + a_1 \mathbf{1}_{\Omega_1}$. So $a_1 \in A_1$. Thus, $C_1 \subseteq A_0 + A_1^L(1) \times \mathbf{0} \ldots \mathbf{0}$.

Now, suppose that

$$C_{k-1} \subseteq A_0 + A_1^L(1) \times \ldots \times A_1^L(k-1) \times \mathbf{0} \ldots \mathbf{0}, \tag{4.14}$$

we want to show that

$$C_k \subseteq A_0 + A_1^L(1) \times \ldots \times A_1^L(k) \times \mathbf{0} \ldots \mathbf{0},$$

For any $c_k = \sum_{i=1}^{k} M_{\Omega_i} x_i + \sum_{i=k+1}^{m} t_i \mathbf{1}_{\Omega_i} \in C_k \subseteq A$, there exists $s_k \in \mathbb{R}$ such that

$$z_k = c_k - (M_{\Omega_k} x_k - s_k \mathbf{1}_{\Omega_k}) \ \in B_k \subseteq A_0 + A_1^L(1) \times \ldots \times A_1^L(k-1) \times \mathbf{0} \ldots \mathbf{0}.$$

The existence of such s_k first exploits the fact that $z_k \in C_{k-1}$ when s_k is small enough, e.g., $s_k =$

$\min x_k$ by monotonicity. Then using relationship 4.14, s_k can be modified (increased) so that $z_k \in B_k$.

Let $a_k = (x_k - s_k)$, we have $c_k = z_k + M_{\Omega_k} a_k$. So $a_k \in A_k$. Thus,

$$C_k \subseteq A_0 + A_1^L(1) \times \ldots \times A_1^L(k) \times \mathbf{0} \ldots \mathbf{0}.$$

■

Now we are ready to prove that our construction provides a lower bound for ρ.

Theorem 23 (Lower Bound)**.** *Suppose we construct a factor-separable risk measure as follows.*

$$\rho^L(X) = \rho_0(\sum_{i=1}^{m} \rho_i^L(X) \mathbf{1}_{\Omega_i}), \tag{4.15}$$

where for any i, ρ_i^L *is defined by equation* (4.11)*. Then for any outcome* $X \in \mathbb{R}^{|\Omega|}$*, we have*

$$\rho^L(X) \leq \rho(X).$$

Proof. We wish to show $A^L = A_0 + A_1^L(1) \times A_1^L(2) \times \ldots \times A_1^L(m) \supseteq A$, which is equivalent to the result. According to lemma 3, take $k = m$, this is immediate. ■

The above theorem 23 shows that ρ^L is a lower bound for ρ. The following part of the section will establish that ρ^L provides a "Pareto" optimal lower bound for ρ.

Lemma 4. *Suppose for a given group of* $\{A_1(i)\}_i$, *we have that*

$$A \subseteq A' = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m),$$

and

$$A_1(i) \subseteq A_1^L(i), \quad \forall i = 1, 2, \ldots, m.$$

This implies that

$$A_1(i) = A_1^L(i), \quad \forall i = 1, 2, \ldots, m.$$

Proof. We prove this lemma by induction. For each step within the induction, we prove by contradiction.

The first induction step is to show that $A_1(1) = A_1^L(1)$. Assume that $A_1(1) \neq A_1^L(1)$, we get $A_1(1) \subset A_1^L(1)$. There exists a_1, such that $a_1 \in A_1^L(1)$ but $a_1 \notin A_1(1)$, and a_1 represents an extreme ray in $A_1^L(1)$. The existence of such a_1 is because it is impossible for $A_1(1)$ to contain all extreme rays in $A_1^L(1)$, otherwise $A_1(1) \subset A_1^L(1)$ doesn't hold. According to the definition of $A_1^L(1)$, there exists $b_0 \in B_1$, such that

$$x = b_0 + M_{\Omega_1} a_1 \in C_1. \tag{4.16}$$

Observe that

$$A' = A_0 + A_1(1) \times A_1(2) \times \ldots \times A_1(m) \supseteq A \supseteq C_1,$$

and

$$t\mathbf{1}_{\Omega_1} \notin A_1(i), \text{ when } t > 0, \quad \forall\, i.$$

It is clear that

$$A_0 + A_1(1) \times \mathbf{0} \times \ldots \times \mathbf{0} \supseteq C_1.$$

To cover x by only using elements from A_0 and $A_1(1)$, there exist $b_0' \in A_0$ and $a_1' \in A_1(1)$ such that

$$x = b_0' + M_{\Omega_1} a_1'. \tag{4.17}$$

Note that $b_0 \in B_1$, according to the definition of B_1 and comparing equation 4.16 and equation 4.17, it is clear that there exists a nonnegative real number $\varepsilon \geq 0$ such that

$$b_0 = b_0' + \varepsilon \mathbf{1}_{\Omega_1},$$

$$a_1 = a_1' - \varepsilon \mathbf{1}_{\Omega_1}.$$

Now, we have $a_1' \in A_1(1)$ and $-\varepsilon \mathbf{1}_{\Omega_1} \in A_1(1)$. Using the additivity of a cone, we obtain that $a_1 \in A_1(1)$, which leads to a contradiction. Thus, we prove that $A_1(1) = A_1^L(1)$.

The next induction step is to show the following. Suppose that it is true that

$$A_1(i) = A_1^L(i), \quad \forall i = 1, 2, \ldots, k-1.$$

We want to show that $A_1(k) = A_1^L(k)$. This is very similar to the case $k = 1$ that has been just proved. Assume that $A_1(k) \neq A_1^L(k)$, we get $A_1(k) \subset A_1^L(k)$. There exists a_k, such that $a_k \in A_1(k)$

but $a_k \notin A_1^L(k)$. According to the definition of $A_1^L(k)$, there exists $b_{k-1} \in B_k$, such that

$$x = b_{k-1} + M_{\Omega_k} a_k \in C_k. \tag{4.18}$$

It is clear that

$$A^{(k)} = A_0 + A_1(1) \times \ldots \times A_1(k) \times \mathbf{0} \times \ldots \times \mathbf{0} \supseteq C_k.$$

Because $A_1(i) = A_1^L(i), \quad \forall i = 1, 2, \ldots, k-1$, we have

$$A^{(k)} = A_0 + A_1^L(1) \times \ldots \times A_1^L(k-1) \times A_1(k) \times \mathbf{0} \times \ldots \times \mathbf{0} \supseteq C_k.$$

To cover x by only using elements from A_0 and $\{A_1(i)\}_{i=1}^k$, there exist $b'_{k-1} \in A^{(k-1)}$ and $a'_k \in A_1(k)$ such that (note that $A^{(k-1)} = A_0 + A_1(1) \times \ldots \times A_1(k-1) \times \mathbf{0} \times \ldots \times \mathbf{0}$.)

$$x = b'_{k-1} + M_{\Omega_k} a'_k. \tag{4.19}$$

Note that $b_{k-1} \in B_{k-1}$, according to the definition of B_{k-1} and comparing equation 4.18 and equation 4.19, it is clear that there exists a nonnegative real number $\varepsilon \geq 0$ such that

$$b_{k-1} = b'_{k-1} + \varepsilon \mathbf{1}_{\Omega_{k-1}},$$

$$a_k = a'_k - \varepsilon \mathbf{1}_{\Omega_1}.$$

Now, we have $a'_k \in A_1(k)$ and $-\varepsilon \mathbf{1}_{\Omega_k} \in A_1(k)$. Using the additivity of a cone, we obtain that $a_k \in A_1(k)$, which leads to a contradiction. Thus, we prove that $A_1(k) = A_1^L(k)$.

To sum up, we have

$$A_1(i) = A_1^L(i), \quad \forall i = 1, 2, \ldots, m.$$

■

Theorem 24 (Best Lower Bound). *Suppose ρ^L is defined according to equation* (4.15). *For another group of risk measures ρ_i, we construct a factor-separable risk measure as follows.*

$$\rho'(X) = \rho_0(\sum_{i=1}^{m} \rho_i(X)\mathbf{1}_{\Omega_i}) \geq \rho(X).$$

Then it is impossible to have

$$\rho_i(x_i) \geq \rho_i^L(x_i) \quad \forall\, x_i \in \mathbb{R}^{|\Omega_i|}, \quad \forall\, i, \tag{4.20}$$

unless,

$$\rho_i = \rho_i^L,\ \forall i.$$

Proof. Considering the relationship between coherent risk measure ρ_i and its acceptance set $A_1(i)$, this theorem is immediate from lemma 4. ■

Remarks. We recall that our construction for the case of overestimation, we have demonstrated that the construction defined in equation (4.5) is absolutely optimal. For the case of underestimation, we have established a "Pareto" optimal result by the construction defined in equation (4.12).

We notice that our construction in equation (4.12) depends on the order of the iterative sequence of $\{\Omega_i\}_i$, which can be completely arbitrary. In fact, for any permutation σ of $\{1,2,\ldots,m\}$, we can define $\{A_1^L(\sigma(i))\}_i$ and $\{B_{\sigma(i)}\}_i$ according to equation (4.12). Hence we will obtain an upper bound for ρ based on the permutation σ. When we list all the possible permutations, we will have m constructions. It is possible that some of the constructions are in fact the same. Each construction provides an upper bound for ρ in a "Pareto" optimal way.

4.5.1 Expected Loss

Besides the low bound we have constructed in equation 4.12, other forms of ρ_i can also produce a lower bound for ρ. Among them, one simple construction is the expected loss function. Specifically, we can define

$$\rho_i^E(x_i) = E[x_i] = \sum_{\omega\in\Omega_i} p_\omega x_i(\omega).$$

For each i, the acceptance set of ρ_i^E is

$$A_1^E(i) = \left\{ x_i \in \mathbb{R}^{|\Omega_i|} : \sum_{\omega\in\Omega_i} p_\omega x_i(\omega) \le \mathbf{0} \right\}. \tag{4.21}$$

We can define a factor-separable risk measure based on the expected loss function.

Definition 22 (Expected Loss Risk Estimator). *Suppose risk measure ρ_i^E is defined by its acceptance set in equation 4.21 for each i. The expected loss risk estimator is the following factor-separable risk measure.*

$$\rho^E = \rho(\sum_{i=1}^m \rho_i^E(X)\mathbf{1}_{\Omega_i}). \tag{4.22}$$

Now we would establish that our construction defined in equation 4.12 provides a better lower bound than the expected loss function which is commonly used in practice for estimating risk.

Theorem 25 (Expected Loss as Lower Bound). *Suppose* ρ^E *is an expected loss risk estimator defined in equation* (4.22)*, and* ρ^E *is a factor-separable risk measure defined in equation* (4.15)*. Then for any outcome* $X \in \mathbb{R}^{|\Omega|}$*, we have*

$$\rho^E(X) \leq \rho^L(X).$$

Proof. The statement is equivalent to the following.

$$\begin{aligned} & \rho^E = \rho(\textstyle\sum_{i=1}^m \rho_i^E(X)\mathbf{1}_{\Omega_i}) \\ \leq \quad & \rho^L = \rho(\textstyle\sum_{i=1}^m \rho_i^L(X)\mathbf{1}_{\Omega_i}). \end{aligned} \tag{4.23}$$

Recall that u_i is the projection of the real probability v onto Ω_i. Now we wish to show that for each i and for any $x \in \mathbb{R}^{|\Omega_i|}$.

$$\rho_i^E(x) = \sum_{\omega \in \Omega} u_i(\omega)x(\omega) \leq \rho_i^L(X). \tag{4.24}$$

This relationship is clear when we write out the dual representation of ρ^L as follows.

$$\rho_i^L(x) = \sup_{q \in Q_i(u_i)} \sum_{\omega \in \Omega} q(\omega)x(\omega),$$

where $Q_i(u_i)$ is the set of probability measures that are absolutely continuous with respect to the real probability u_i. It is clear that $u_i \in Q_i(u_i)$. Since u_i is a feasible solution in the dual representation of ρ^L, inequality (4.24) holds. According to the monotonicity of coherent risk measures, $\rho^E(X) \leq \rho^L(X)$. ■

Corollary 4. *Suppose we are given a coherent risk measure* ρ. *We construct* ρ^E *as an expected loss risk estimator defined in equation* (4.22). *Then for any outcome* $X \in \mathbb{R}^{|\Omega|}$, *we have*

$$\rho^E(X) \leq \rho(X).$$

Proof. According to Theorem 25 and Theorem 23, it follows immediately that for any outcome $X \in \mathbb{R}^{|\Omega|}$, we have

$$\rho^E(X) \leq \rho^L(X) \leq \rho(X).$$

■

Chapter 5

Concluding Remarks

This thesis proposed a theoretical framework for systemic risk. We investigated systemic risk from different perspectives, including an axiomatic approach that accommodates the risk measurement for various forms of applications, a structural model that highlights contagion in networks, and an estimation formulation based on risk factors.

In Chapter 2, we proposed a set of axioms that capture essential properties of systemic risk measures. The proposed framework is built upon the classic work by Artzner et al. [1999]. Our theory, with a set of new axioms, focuses on the difference between a single-firm risk measure and a systemic one. We demonstrated that the special cases of our model includes many proposed systemic risk measures. The theoretical framework highlights the fact that a systemic risk measure can be decomposed to two components: a base risk measure and an aggregation function. These two components addresses different aspects of risk preferences. The dual representation derived from the axioms can be applied to attribute risk to individual components of the system, which properly accounts for the externalities imposed by each individual. Based on the attribution rule, we have further designed a decentralized taxation scheme, where the system manager can align the incentives of individuals. Finally, we showed that our methodology presented in this chapter can be extended to a general class of risk measures.

In Chapter 3, we introduced a structural model that studies contagion effects in financial networks. We investigated the interaction resulted from the asset-firm portfolio holing and re-balancing relationship. Through the analysis of endogenous asset prices that are captured by the marketing clearing condition, we characterized the risk components in the mechanism of contagion, and we identified the key sources of systemic risk. Specifically, leverage across firms is one of the critical risk element. We discovered that for two different regimes, high-leverage and low-leverage economies, the risk properties of the system appear to be completely opposite. We introduced the computable systemic leverage to define these two economic regimes. We proved that the optimal network structures that minimize systemic risk in these two cases are mutual fund network and isolated network respectively. We pointed out that diversification across assets is favored in the low-leverage economy, while the diversity of firms has more value in the high-leverage economy.

In Chapter 4, we solved theoretical problems of upper bounding and lower bounding risk measures based on risk factors. Given an original risk measure and factor information, we introduced a form of factor-separable risk measures. We connected the acceptance sets of risk measures to the factor-separable measure to provide explicit expressions for upper and lower bound of the original risk measure. We showed that a tight approximation is constructed for the upper bounding problem and a Pareto optimal construction is possible for the lower bounding problem.

However, there are many research questions that remain to be addressed. We highlight some of these below:

Statistical estimation of systemic risk. We have presented theoretical frameworks that build a solid foundation for measuring and estimating systemic risk. However, to develop statistical estimation procedures from data brings other aspects of technical challenges. In a recent paper, Rockafellar et al. [2008] show how to compute the weights for a given set of factors that best approximate the risk of a single agent. It would be interesting to extend the existing methodologies to computing the risk exposure to specific systemic factors. Another related problem is to develop procedures for identifying a sparse set of factors that best represents a given systemic risk measure. We expect that these risk factors would help in generating most risk-relevant scenarios. Even in the single-agent case, factor selection schemes to best represent coherent and related convex risk measures have not been explored in the literature.

Dynamic systemic risk control. Controlling systemic risk in a complex network involves the problem of real-time control, i.e., decision-making as events are unfolding. One example of such work is in power networks. The impact of recent blackouts inspires the development of optimization-based methods to provide control guidance in the event of a future blackout event. An interesting insight that has emerged in the literature (See Bienstock [2011]) is that taking a (greedy) control after observing very few events in a potential cascade results in very poor decisions. In a resilient network, the immediate impact may be difficult to determine and, in time, the structure recovers. Applying intelligent control after a suitable delay yields a better outcome. On the other hand, waiting too long would result in a situation where one may not have enough resources to mitigate a cascade once it has propagated. This insight is equally relevant to other systemic risk settings, including inter-bank financial networks. It would be interesting to extend this control methodology to networks where agents make self-interested decisions.

Bibliography

Beatrice Acciaio and Irina Penner. Dynamic risk measures. In *Advanced Mathematical Methods for Finance*, pages 1–34. Springer, 2011.

Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. Technical report, National Bureau of Economic Research, 2013.

V. V. Acharya, L. H. Pedersen, T. Philippon, and M. Richardson. Measuring systemic risk. Working paper, March 2010a.

V. V. Acharya, T. Yorulmazer, and J. A. C. Santos. Systemic risk and deposit insurance premiums. *Federal Reserve Bank of New York Economic Policy Review*, 16(1):89–99, August 2010b.

T. Adrian and M. Brunnermeier. CoVaR. *Federal Reserve Bank of New York Staff Reports*, 348: 1–8, 2009.

Franklin Allen and Douglas Gale. Financial contagion. *Journal of political economy*, 108(1):1–33, 2000.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, July 1999.

R. J. Aumann and L. S. Shapley. *Values of Non-Atomic Games*. Princeton University Press, 1974.

D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1999.

Daniel Bienstock. Adaptive online control of cascading blackouts. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–8. IEEE, 2011.

L. J. Billera and D. C. Heath. Allocation of shared costs: A set of axioms yielding a unique procedure. *Mathematics of Operations Research*, 7:32–39, 1982.

Markus K Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *Review of Financial studies*, 22(6):2201–2238, 2009.

A. Buch and G. Dorfleitner. Coherent risk measures, coherent capital allocations and the gradient allocation principle. *Insurance: Mathematics and Economics*, 42(1):235–242, 2008. ISSN 0167-6687.

John Y Campbell and Luis M Viceira. *Strategic asset allocation: portfolio choice for long-term investors*. Oxford University Press, 2002.

M. Castellani. A dual representation for proper positively homogeneous functions. *Journal of Global Optimization*, 16(4):393–400, 2000.

S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio. Risk measures: Rationality and diversification. *Mathematical Finance*, 2010.

A. Cherny and D. Orlov. On two approaches to coherent risk contribution. *Mathematical Finance*, 21(3):557–571, 2011.

Alexander S Cherny and Dilip B Madan. Coherent measurement of factor risks. *arXiv preprint math/0605062*, 2006.

R. Cont, A. Moussa, and E. Bastos e Santos. Network structure and systemic risk in banking systems. Working paper, December 2010.

M. Denault. Coherent allocation of risk capital. *Journal of Risk*, 4:1–34, 2001. ISSN 1465-1211.

L. Eisenberg and T. H. Noe. Systemic risk in financial systems. *Management Science*, 47(2): 236–249, February 2001.

Matthew Elliott, Benjamin Golub, and Matthew O Jackson. Financial networks and contagion. *Available at SSRN 2175056*, 2013.

T. Fischer. Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance: Mathematics and Economics*, 32:135–146, 2003.

Prasanna Gai and Sujit Kapadia. Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, page rspa20090410, 2010.

C. Gauthier, A. Lehar, and M. Souissi. Macroprudential regulation and systemic capital requirements. Technical Report 2010–4, Bank of Canada, January 2010.

K. Giesecke and B. Kim. Systemic risk: What defaults are telling us. Working paper, January 2011.

Paul Glasserman and H Peyton Young. How likely is contagion in financial networks? *Journal of Banking & Finance*, 2014.

Robin Greenwood, Augustin Landier, and David Thesmar. Vulnerable banks. Technical report, National Bureau of Economic Research, 2012.

O. Haimanko. Cost sharing: the nondifferentiable case. *Journal of Mathematical Economics*, 35: 445–462, 2001.

X. Huang, H. Zhou, and H. Zhu. A framework for assessing the systemic risk of major financial institutions. *Journal of Banking & Finance*, 33(11):2036–2049, 2009.

Dan A Iancu, Marek Petrik, and Dharmashankar Subramanian. Tight approximations of dynamic risk measures. *arXiv preprint arXiv:1106.6102*, 2011.

P. Jorion. *Value at Risk*. McGraw-Hill, New York, 2006.

S. G. Kou, X. Peng, and C. C. Heyde. External risk measures and Basel accords. Working paper, Columbia University and Hong Kong University of Science and Technology, 2009.

D. Kreps. *Notes on the Theory of Choice*. Westview Press, 1988.

A. Lehar. Measuring systemic risk: A risk management approach. *Journal of Banking & Finance*, 29(10):2577–2603, 2005.

M. Liu and J. Staum. Systemic risk components and deposit insurance premia. Technical Report 2011–03, Federal Deposit Insurance Corporation Center for Financial Research, December 2010.

J. F. Mertens. The Shapley value in the non-differentiable case. *International Journal of Game Theory*, 17:1–65, 1988.

X. Liu N. Chen and D. Yao. Interconnected balance sheets, market liquidity and the amplification effect in a financial network. Working paper, June 2014.

Frau Dipl-Math Irina Penner. *Dynamic convex risk measures: time consistency, prudence, and sustainability*. PhD thesis, Humboldt-Universität zu Berlin, 2007.

R Tyrrell Rockafellar, Stan Uryasev, and Michael Zabarankin. Risk tuning with generalized linear regression. *Mathematics of Operations Research*, 33(3):712–729, 2008.

Berend Roorda, Johannes M Schumacher, and Jacob Engwerda. Coherent acceptability measures in multiperiod models. *Mathematical Finance*, 15(4):589–612, 2005.

Dan Rosen and David Saunders. Risk factor contributions in portfolio credit risk models. *Journal of Banking & Finance*, 34(2):336–349, 2010.

A. Schied. Risk measures and robust optimization problems. *Stochastic Models*, 22(4):753, 2006.

J. C. Staum. Systemic risk components and deposit insurance premia. Working paper, January 2011a.

J. C. Staum. Counterparty contagion in context: Contributions to systemic risk. Working paper, 2011b.

N. Tarashev, C. Borio, and K. Tsatsaronis. Attributing systemic risk to individual institutions. Technical Report 308, Working paper 308, Bank of International Settlements, May 2010.

A. Tsanakas. To split or not to split: Capital allocation with convex risk measures. *Insurance: Mathematics and Economics*, 44:268–277, 2009.

Wolf Wagner. Systemic liquidation risk and the diversity–diversification trade-off. *The Journal of Finance*, 66(4):1141–1175, 2011.

Peter J. Wallison. Dodd-frank and the myth of "interconnectedness". *The Wall Street Journal*, 2012.

Appendix A

Proofs of Theorems

A.1 Proofs of Chapter 2

Theorem 3. *Suppose* $\rho = \rho_0 \circ \Lambda$ *is a systemic risk measure characterized by an aggregation function* Λ *and a base risk measure* ρ_0*. Then, for all economies* $X \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}$*,* $\rho(X)$ *can be expressed as the value of the optimization problem*

$$\begin{aligned} \rho(X) = \quad & \underset{\bar{\pi},\Xi}{\text{maximize}} && \sum_{i\in\mathcal{F}}\sum_{\omega\in\Omega} \Xi_{i,\omega} X_{i,\omega} \\ & \text{subject to} && (1,\bar{\pi}) \in \mathcal{A}^*, \\ &&& (\bar{\pi}_\omega, \Xi_\omega) \in \mathcal{B}^*, \qquad \forall\, \omega \in \Omega, \\ &&& \bar{\pi} \in \mathbb{R}^{|\Omega|},\ \Xi \in \mathbb{R}^{|\mathcal{F}|\times|\Omega|}. \end{aligned} \tag{A.1}$$

In addition, feasible points $(\bar{\pi}, \Xi)$ *for this problem must satisfy*

$$\bar{\pi} \geq \mathbf{0}_\Omega, \quad \mathbf{1}_\Omega^\top \bar{\pi} \leq 1, \quad \Xi \geq \mathbf{0}_{\mathcal{E}}, \quad \mathbf{1}_{\mathcal{F}}^\top \Xi \leq |\mathcal{F}| \bar{\pi}^\top. \tag{A.2}$$

Proof. In this proof, we use the primal representation of Theorem 2. Recall that primal representation of ρ is

$$\rho(X) = \inf_{(m,\ell) \in \mathbb{R} \times \mathbb{R}^{|\Omega|}} m + I_{\mathcal{A}}(m, \ell) + \sum_{\omega \in \Omega} I_{\mathcal{B}}(\ell_\omega, x_\omega),$$

where $I_{\mathcal{A}}$ and $I_{\mathcal{B}}$ are indicator functions of sets $\mathcal{A}$ and $\mathcal{B}$ from (2.11) respectively. Here, given a set $\mathcal{C} \in \mathbb{R}^n$, we define the indicator function $I_{\mathcal{C}} \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by

$$I_{\mathcal{C}}(x) \triangleq \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ \infty & \text{otherwise.} \end{cases} \tag{A.3}$$

For the convex set $\mathcal{A}$, define the support function by

$$S_{\mathcal{A}}(-\pi_0, \bar{\pi}) \triangleq \sup_{(m,\ell) \in \mathbb{R} \times \mathbb{R}^{|\Omega|}} -\pi_0 m + \bar{\pi}^\top \ell = \begin{cases} 0 & \text{if } (\pi_0, \bar{\pi}) \in \mathcal{A}^*, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{A}^*$ is given by (2.12). From convex duality, the indicator function of $\mathcal{A}$ can be expressed as the Fenchel-Legendre conjugate of its support function, i.e.,

$$I_{\mathcal{A}}(m, \ell) = \sup_{(\pi_0, \bar{\pi}) \in \mathbb{R} \times \mathbb{R}^{|\Omega|}} -\pi_0 m + \bar{\pi}^\top \ell - S_{\mathcal{A}}(-\pi_0, \bar{\pi}) = \sup_{(\pi_0, \bar{\pi}) \in \mathcal{A}^*} -\pi_0 m + \bar{\pi}^\top \ell.$$

Similarly, the indicator function for the set $\mathcal{B}$ can be expressed as

$$I_{\mathcal{B}}(\ell_\omega, x_\omega) = \sup_{(\xi_{0\omega}, \hat{\xi}_\omega) \in \mathcal{B}^*} -\xi_{0\omega}\ell_\omega + \hat{\xi}_\omega^\top x_\omega,$$

where $\mathcal{B}^*$ is given by (2.13).

Thus, applying convex duality to the primal representation, we obtain that

$$\begin{aligned}
\rho(X) &= \inf_{(m,\ell)\in\mathbb{R}\times\mathbb{R}^{|\Omega|}} \sup_{\substack{(\pi_0,\bar{\pi})\in\mathcal{A}^* \\ (\xi_{0\omega},\hat{\xi}_\omega)\in\mathcal{B}^*,\ \forall\omega\in\Omega}} m - \pi_0 m + \bar{\pi}^\top \ell + \sum_{\omega\in\Omega} \left(-\xi_{0\omega}\ell_\omega + \hat{\xi}_\omega^\top x_\omega\right) \\
&= \sup_{\substack{(\pi_0,\bar{\pi})\in\mathcal{A}^* \\ (\xi_{0\omega},\hat{\xi}_\omega)\in\mathcal{B}^*,\ \forall\omega\in\Omega}} \inf_{(m,\ell)\in\mathbb{R}\times\mathbb{R}^{|\Omega|}} m - \pi_0 m + \bar{\pi}^\top \ell + \sum_{\omega\in\Omega} \left(-\xi_{0\omega}\ell_\omega + \hat{\xi}_\omega^\top x_\omega\right) \\
&= \sup_{\substack{(1,\bar{\pi})\in\mathcal{A}^* \\ (\bar{\pi}_\omega,\hat{\xi}_\omega)\in\mathcal{B}^*,\ \forall\omega\in\Omega}} \sum_{\omega\in\Omega} \hat{\xi}_\omega^\top x_\omega.
\end{aligned}$$

This establishes (A.1).

Now, we show the sub-stochastic properties (A.2) for feasible dual variables $(\bar{\pi}, \Xi)$. Up to a sign change, $\mathcal{A}^*$ and $\mathcal{B}^*$ are the dual cones to $\mathcal{A}$ and $\mathcal{B}$, respectively. The monotonicity of the acceptance sets $\mathcal{A}$ and $\mathcal{B}$ implies that

$$\bar{\pi} \geq \mathbf{0}_\Omega, \quad \Xi \geq \mathbf{0}_{\mathcal{E}}.$$

From the normalization property of ρ_0, we have that $(1, \mathbf{1}_\Omega) \in \mathcal{A}$. This implies that

$$1 - \mathbf{1}_\Omega^\top \bar{\pi} \geq 0.$$

From the normalization property of Λ, we have that $(|\mathcal{F}|, \mathbf{1}_{\mathcal{F}}) \in \mathcal{B}$. This implies that

$$|\mathcal{F}|\bar{\pi}_\omega - \mathbf{1}_{\mathcal{F}}^\top \Xi_\omega \geq 0, \quad \forall\, \omega \in \Omega.$$

■

Lemma 1. *A function $g\colon \mathbb{R}^n \to \mathbb{R}$ is positively homogeneous and monotonic if and only if there exists an index set $\mathcal{S}$ where, for each $s \in \mathcal{S}$, $g^{(s)}\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a positively homogeneous, monotonic, and convex extended real-valued function, such that*

$$g(x) = \underset{s \in \mathcal{S}}{\text{minimize}}\ g^{(s)}(x),$$

for all $x \in \mathbb{R}^n$.

Proof. This proof closely follows the argument presented by Castellani [2000]. Let $\mathcal{S} \triangleq \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ denote the unit sphere in $\mathbb{R}^n$. For $s \in \mathcal{S}$, define

$$\mathcal{T}_s \triangleq \mathcal{K}_s^\circ + g(s)s,$$

where $\mathcal{K}_s \triangleq \{\lambda s : \lambda \geq 0\}$ denotes the cone generated by $s \in \mathcal{S}$, and $\mathcal{K}_s^\circ \triangleq \{y : y^\top s \leq 0\}$ denotes the polar cone of $\mathcal{K}_s$. Given $s \in \mathcal{S}$, define

$$g^{(s)}(x) \triangleq \sup_{y \in \mathcal{T}_s \cap \mathbb{R}^n_+} y^\top x. \tag{A.4}$$

Then $g^{(s)}\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is clearly monotonic, positively homogeneous and convex. Using convex duality, we can rewrite $g^{(s)}$ as

$$\begin{aligned} g^{(s)}(x) &= \sup_{y\in\mathcal{T}_s,\ y\geq 0} y^\top x \\ &= \inf_{\beta\geq 0}\ \sup_{y\in\mathcal{T}_s} y^\top(x+\beta) \\ &= \inf_{\beta\geq 0} \left\{ g(s)s^\top(x+\beta) + \inf_{z\in\mathcal{K}_s^\circ} z^\top(x+\beta) \right\} \\ &= \inf_{\beta\geq 0} g(s)s^\top(x+\beta) + I_{\mathcal{K}_s}(x+\beta), \end{aligned}$$

where $I_{\mathcal{K}_s}(z)$ denotes the indicator function (A.3) for the set $\mathcal{K}_s$. The second equality follows from strong duality since $\mathcal{T}_s$ and $\mathbb{R}^n_+$ are non-empty polyhedral sets, and the last equality follows from the fact that, for all $z\in\mathbb{R}^n$,

$$\sup_{y\in\mathcal{K}_s^\circ} y^\top z = I_{\mathcal{K}_s}(z).$$

Since $\mathcal{S}$ is a compact set, for any $x\in\mathbb{R}^n$, $\inf_{s\in\mathcal{S}} g^{(s)}(x)$ is achieved, i.e., the infimum is, in fact, a minimum. Moreover,

$$\begin{aligned} \min_{s\in\mathcal{S}} g^{(s)}(x) &= \min_{s\in\mathcal{S}}\ \inf_{\beta\geq 0} g(s)s^\top(x+\beta) + I_{\mathcal{K}_s}(x+\beta) \\ &= \inf_{\beta\geq 0}\ \min_{s\in\mathcal{S}} g(s)s^\top(x+\beta) + I_{\mathcal{K}_s}(x+\beta). \end{aligned} \tag{A.5}$$

Note that $I_{\mathcal{K}_s}(z) < \infty$ only if $z = \lambda s$, for some $\lambda \geq 0$. In that case, by positive homogeneity, $g(s)s^\top z + I_{\mathcal{K}_s}(z) = \lambda g(s) = g(\lambda s) = g(z)$. Thus, it follows that for all $z\in\mathbb{R}^n$,

$$\min_{s\in\mathcal{S}} g(s)s^\top z + I_{\mathcal{K}_s}(z) = g(z).$$

Then, from (A.5), it follows that

$$\min_{s \in \mathcal{S}} g^{(s)}(x) = \inf_{\beta \geq 0} g(x+\beta) = g(x),$$

where the second equality follows from the fact that g is monotonically increasing and $\beta \geq 0$. ■

A.2 Proofs of Chapter 3

Theorem 8. *Suppose we are given initial endowments* (Θ, θ^0) *and the risk factor x. For a vector of posted asset prices q, there exists an equilibrium with prices q if and only if q satisfies the following* ***market clearing condition****:*

$$Dq = \Pi(q, \theta^0 + \Theta^\top q, x)\left(\theta^0 + \Theta^\top q\right). \tag{A.6}$$

Proof. Suppose q satisfies the market clearing condition in equation (3.6). We take

$$\hat{\theta}_h^0 = \left(1 - \sum_{i \in \mathcal{A}} \Pi_{ih}(q, \theta^0 + \Theta^\top q, x)\right)\left(\theta_h^0 + \sum_{i \in \mathcal{A}} \Theta_{ih} q_i\right), \quad \forall\, h \in \mathcal{F},$$

and

$$\hat{\Theta}_{ih} = \frac{1}{q_i}\Pi_{ih}(q, \theta^0 + \Theta^\top q, x)\left(\theta_h^0 + \sum_{j \in \mathcal{A}} \Theta_{jh} q_j\right), \quad \forall\, i \in \mathcal{A}\, \forall\, h \in \mathcal{F}.$$

Now, for any $h \in \mathcal{F}$,

$$
\begin{aligned}
& \sum_{i\in\mathcal{A}} \left(\hat{\Theta}_{ih} - \Theta_{ih}\right) q_i + \left(\hat{\theta}_h^0 - \theta_h^0\right) \\
= \; & \sum_{i\in\mathcal{A}} \Pi_{ih}\theta_h^0 + \sum_{i\in\mathcal{A}} \Pi_{ih} \sum_{j\in\mathcal{A}} \Theta_{jh} q_j - \sum_{i\in\mathcal{A}} \Theta_{ih} q_i \\
& + \left(1 - \sum_{i\in\mathcal{A}} \Pi_{ih}\right) \theta_h^0 + \left(1 - \sum_{i\in\mathcal{A}} \Pi_{ih}\right) \sum_{j\in\mathcal{A}} \Theta_{jh} q_j - \theta_h^0 \\
= \; & 0.
\end{aligned}
$$

So the budget balance in equation (3.3) holds. According to the market clearing condition in equation (3.6), for each $i \in \mathcal{A}$ we have

$$
D_{ii} = \frac{1}{q_i} \sum_{h\in\mathcal{F}} \Pi_{ih}(q, \theta^0 + \Theta^\top q, x) \left(\theta_h^0 + \sum_{j\in\mathcal{A}} \Theta_{jh} q_j\right) = \sum_{h\in\mathcal{F}} \hat{\Theta}_{ih}.
$$

So the share balance in equation (3.4) holds. By using the proved result of budget balance in equation (3.3), for each $i \in \mathcal{A}$ and each $h \in \mathcal{F}$, we can show that

$$
\hat{\Theta}_{ih} = \frac{1}{q_i} \Pi_{ih}(q, \theta^0 + \Theta^\top q, x) \left(\theta_h^0 + \sum_{j\in\mathcal{A}} \Theta_{jh} q_j\right) = \Pi_{ih}(q, \theta^0 + \Theta^\top q, x) \left(\hat{\theta}_h^0 + \sum_{j\in\mathcal{A}} \hat{\Theta}_{jh} q_j\right).
$$

So the portfolio match condition in equation (3.5) holds. Therefore $(q, \hat{\Theta}, \hat{\theta}^0)$ established an equilibrium.

Now we suppose that there exists an equilibrium with prices q. Namely there exists a matrix

$\hat{\Theta} \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}$ and a vector $\hat{\theta}^0 \in \mathbb{R}^{|\mathcal{F}|}$ such that $(q, \hat{\Theta}, \hat{\theta}^0)$ establishes an equilibrium. For each $i \in \mathcal{A}$,

$$\begin{aligned} D_{ii}q_i &= \sum_{h\in\mathcal{F}} \hat{\Theta}_{ih} q_i \\ &= \sum_{h\in\mathcal{F}} \Pi_{ih}(q, \theta^0 + \Theta^\top q, x) \left(\hat{\theta}^0_h + \sum_{j\in\mathcal{A}} \hat{\Theta}_{jh} q_j \right) \\ &= \sum_{h\in\mathcal{F}} \Pi_{ih}(q, \theta^0 + \Theta^\top q, x) \left(\theta^0_h + \sum_{j\in\mathcal{A}} \Theta_{jh} q_j \right). \end{aligned}$$

The first equality is because of the share balance condition in equation (3.4). The second equality is due to the portfolio match condition in equation (3.5). The last equality is because of the budget balance condition in equation (3.3). Therefore, the market clearing condition is proved. ■

Theorem 9. *Suppose (q, Θ, θ^0) is a given equilibrium, and the risk factor is x_0. Assume that $\Pi(q(x), w(\theta^0(x), q(x)), x)$ is a continuously differentiable function of q, w, and x, and $\theta^0(x)$ is a continuously differentiable function of x. If the matrix $\left[I - (\Pi + \Omega)\bar{\Theta}^\top - HD^{-1}\right]$ is invertible at $x = x_0$, then there exists a neighborhood $N(x_0)$ of x_0 and a unique price function $q(x)$ on $N(x_0)$ such that for all $x \in N(x_0)$, $q(x)$ satisfies the market clearing condition,*

$$Dq(x) = \Pi(q(x), w(\theta^0(x), q(x)), x) \left(\theta^0(x) + \Theta^\top q(x) \right), \tag{A.7}$$

and we have that

$$D\frac{dq}{dx} = \left[I - (\Pi + \Omega)\bar{\Theta}^\top - HD^{-1}\right]^{-1} \left[(\Pi + \Omega) \frac{d\theta^0}{dx} + \frac{\partial \Pi}{\partial x} w \right]. \tag{A.8}$$

Proof. Equation (A.7) characterizes the relation between q and x. Since $\Pi(q(x), w(\theta^0(x), q(x)), x)$

is a continuously differentiable function of q, w, and x, and $\theta^0(x)$ is a continuously differentiable function of x, we can apply implicit function theorem to equation (A.7). By differentiating all the terms w.r.t. x, we can investigate price sensitivity with respect to shocks on x.

$$\Pi\left(\frac{d\theta^0}{dx}+\Theta^\top\frac{dq}{dx}\right)+\frac{d\Pi}{dx}w-D\frac{dq}{dx} \quad = \quad 0. \tag{A.9}$$

For each asset i, we can write out an element-wise equation for equation (A.9). We obtain that $\left[D\frac{dq}{dx}\right]_i$ equals

$$\left[\Pi\frac{d\theta^0}{dx}\right]_i+\left[\Pi\Theta^\top\frac{dq}{dx}\right]_i+\sum_{h\in\mathcal{F}}w_h\left(\frac{\partial\Pi_{ih}}{\partial x}+\sum_{j\in\mathcal{A}}\frac{\partial\Pi_{ih}}{\partial q_j}\frac{dq_j}{dx}+\frac{\partial\Pi_{ih}}{\partial w_h}\left(\Theta^\top\frac{dq}{dx}+\frac{d\theta^0}{dx}\right)\right), \tag{A.10}$$

where $\frac{\partial\Pi}{\partial x}\in\mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}$ is the portfolio sensitivity w.r.t. to x when asset prices q and w are fixed. $\frac{\partial\Pi_{ih}}{\partial q_j}\in\mathbb{R}$ characterizes portfolio sensitivity w.r.t. prices q, and $\frac{\partial\Pi_{ih}}{\partial w_h}\in\mathbb{R}$ characterizes portfolio sensitivity w.r.t. wealth w.

By definition, the elements of H and Ω we identify that

$$H_{ij}=[(G^iw)^\top]_j=\sum_{h\in\mathcal{F}}\frac{\partial\Pi_{ih}}{\partial q_j}w_h, \quad \forall\, i,j\in\mathcal{A}. \tag{A.11}$$

$$\Omega_{ih}=\frac{\partial\Pi_{ih}}{\partial w_h}w_h, \quad \forall\, i\in\mathcal{A}\;\forall\, h\in\mathcal{F}. \tag{A.12}$$

Writing equation (A.10) in a matrix form and re-arranging terms, we obtain the following equations.

$$\begin{aligned}
\Pi\frac{d\theta^0}{dx}+\Pi\Theta^\top\frac{dq}{dx}+\frac{\partial\Pi}{\partial x}w+H\frac{dq}{dx}+\Omega\Theta^\top\frac{dq}{dx}+\Omega\frac{d\theta^0}{dx} &= D\frac{dq}{dx} \\
\left[D-(\Pi+\Omega)\Theta^\top-H\right]\frac{dq}{dx} &= (\Pi+\Omega)\frac{d\theta^0}{dx}+\frac{\partial\Pi}{\partial x}w \\
\left[I-(\Pi+\Omega)\bar{\Theta}^\top-HD^{-1}\right]D\frac{dq}{dx} &= (\Pi+\Omega)\frac{d\theta^0}{dx}+\frac{\partial\Pi}{\partial x}w.
\end{aligned} \tag{A.13}$$

For the last equality, we let $\bar{\Theta}=D^{-1}\Theta\in\mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ denote the normalized initial portfolios. When $\left[I-(\Pi+\Omega)\bar{\Theta}^\top-HD^{-1}\right]$ is invertible, we can solve for $\frac{dq}{dx}$ in equation (A.13), the result of the theorem follows. ■

Theorem 10. *Suppose (q,Θ,θ^0) is a given equilibrium, with the risk factor $x=x_0$. Assume that the conditions of Theorem 9 hold, and $H=\mathbf{0}$ and $\Omega=\mathbf{0}$. If the initial endowment is in equilibrium and satisfies equation* (3.20)*, then we have the following results.*

1. $$\Pi\bar{\Theta}^\top=Y(X),$$

   where $Y(X)\triangleq X\,\mathrm{diag}(b)\,\mathrm{diag}(w)\,\mathrm{diag}(b)X^\top M^{-1}$, where $M=D\,\mathrm{diag}(q)$ is a diagonal matrix, where the diagonal entries are the total market chaptalization of each asset.

2. *The eigenvalues of $Y(X)$ are real and non-negative.*

3. *The maximal network amplifier is given by*

   $$\mathit{MNA}=\max_{1\le i\le|\mathcal{A}|}|1-\lambda_i(Y(X))|^{-1}, \tag{A.14}$$

   where each λ_i is the ith eigenvalue of the matrix.

Proof. Each element of $\Pi\bar{\Theta}^\top$ is

$$(\Pi\bar{\Theta}^\top)_{ij} = \sum_{h=1}^{|\mathcal{F}|} \Pi_{ih} \frac{w_h \Pi_{jh}}{q_j D_{jj}} = \frac{1}{q_j D_{jj}} \sum_{h=1}^{|\mathcal{F}|} \Pi_{ih} \Pi_{jh} w_h. \quad \forall\, i\, \forall\, h$$

Now we can re-write $\Pi\bar{\Theta}^\top$ as

$$\Pi\bar{\Theta}^\top = \Pi\,\mathrm{diag}(w)\Pi^\top M^{-1} = X\,\mathrm{diag}(b)\,\mathrm{diag}(w)\,\mathrm{diag}(b)X^\top M^{-1} = Y(X).,\tag{A.15}$$

Since $N = X\,\mathrm{diag}(b)\,\mathrm{diag}(w)\,\mathrm{diag}(b)X^\top$ is a positive semidefinite matrix, so it has real non-negative eigenvalues. We note that $M^{-1} = D^{-1}\,\mathrm{diag}(q)^{-1}$ is a positive diagonal matrix, $L = M^{-\frac{1}{2}} N M^{-\frac{1}{2}}$ is also a positive semidefinite matrix, so it also has real non-negative eigenvalues. Since $\Pi\bar{\Theta}^\top = M^{\frac{1}{2}} L M^{-\frac{1}{2}}$ is similar to L, $\Pi\bar{\Theta}^\top = Y(X)$ has real non-negative eigenvalues.

Therefore, the maximal network amplifier is simplified to equation (3.21). ■

Theorem 11.

$$\overline{\lambda_{min}} \leq \underline{\lambda_{max}}.\tag{A.16}$$

Proof. We define

$$\gamma = \frac{\sum_{h\in\mathcal{F}} w_h b_h^2}{\sum_{h\in\mathcal{F}} w_h b_h}.$$

To prove this theorem, we need to establish several facts.

$$\underline{\lambda_{max}} = \gamma,$$

and

$$\overline{\lambda_{min}} \leq \gamma.$$

First, for $\underline{\lambda_{max}}$, we can show that the optimization program in equation (3.23) has an optimal solution $X^* = \frac{1}{v} Dq \mathbf{1}_{\mathcal{F}}^\top$ for equation (3.23) such that

$$\underline{\lambda_{max}} = \lambda_{max}(Y(X^*)) = \gamma$$

and

$$Y(X^*)Dq = \gamma Dq.$$

Namely, the corresponding eigenvector is Dq. To obtain this result, we re-write the original program (3.23) as

$$\begin{array}{llll} \underset{X \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{F}|}}{\text{minimize}} & & t & \\ \text{subject to} & tI_{\mathcal{A}} - X \operatorname{diag}(b) \operatorname{diag}(w) \operatorname{diag}(b) X^\top M^{-1} & \succeq & \mathbf{0}, \\ & Dq & = & X \operatorname{diag}(b) w, \\ & \mathbf{1}_{\mathcal{A}}^\top X & = & \mathbf{1}_{\mathcal{F}}^\top, \\ & X & \geq & \mathbf{0}. \end{array} \tag{A.17}$$

This program is equivalent to the following semidefinite program by using Schur complement.

$$\begin{array}{lrcl}\underset{X\in\mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}}{\text{minimize}} & & t & \\ \text{subject to} & \begin{pmatrix}\operatorname{diag}(w)^{-1} & \operatorname{diag}(b)X^{\top}M^{-\frac{1}{2}} \\ M^{-\frac{1}{2}}X\operatorname{diag}(b) & tI_{\mathcal{A}}\end{pmatrix} & \succeq & \mathbf{0}, \\ & Dq & = & X\operatorname{diag}(b)w, \\ & \mathbf{1}_{\mathcal{A}}^{\top}X & = & \mathbf{1}_{\mathcal{F}}^{\top}, \\ & X & \geq & \mathbf{0}.\end{array} \tag{A.18}$$

Take $X = \frac{1}{v}Dq\mathbf{1}_{\mathcal{F}}^{\top}$ and $t = \gamma$. It is not hard to show that (X,t) is feasible in program (A.18). Now, we want to prove that $(X,t) = (\frac{1}{v}Dq,\gamma)$ is indeed the optimal solution for program (A.18). To prove it, we notice that the constraint $X \geq 0$ is not binding at $(X,t) = (\frac{1}{v}Dq,\gamma)$. We construct the following program that modifies program (A.18) by removing the constraint $X \geq 0$.

$$\begin{array}{lrcl}\underset{X\in\mathbb{R}^{|\mathcal{A}|\times|\mathcal{F}|}}{\text{minimize}} & & t & \\ \text{subject to} & \begin{pmatrix}\operatorname{diag}(w)^{-1} & \operatorname{diag}(b)X^{\top}M^{-\frac{1}{2}} \\ M^{-\frac{1}{2}}X\operatorname{diag}(b) & tI_{\mathcal{A}}\end{pmatrix} & \succeq & \mathbf{0}, \\ & Dq & = & X\operatorname{diag}(b)w, \\ & \mathbf{1}_{\mathcal{A}}^{\top}X & = & \mathbf{1}_{\mathcal{F}}^{\top}.\end{array} \tag{A.19}$$

The optimal value of the new program (A.19) is no greater than optimal value of the new program (A.18). We want to prove that $(X,t) = (\frac{1}{v}Dq,\gamma)$ is the optimal solution for program (A.19).

The dual program for the primal program (A.19) is the following.

$$
\begin{aligned}
\underset{Z,\,\mu_A,\,\mu_F}{\text{maximize}} \quad & -\mu_A^\top Dq - \mu_F^\top b - tr(W^{-1}Z_{FF}) \\
\text{subject to} \quad & Z = \begin{pmatrix} Z_{FF} & Z_{FA} \\ Z_{FA}^\top & Z_{AA} \end{pmatrix} \succeq 0, \\
& tr(Z_{AA}) = 1, \\
& 2Z_{FA}M^{-\frac{1}{2}} = w\mu_A^\top + \mu_F \mathbf{1}_A^\top, \\
& Z \in \mathbb{R}^{(|\mathcal{F}|+|\mathcal{A}|)\times(|\mathcal{F}|+|\mathcal{A}|)}, \\
& \mu_A \in \mathbb{R}^{|\mathcal{A}|},\ \mu_F \in \mathbb{R}^{|\mathcal{F}|}.
\end{aligned}
\tag{A.20}
$$

Take $\mu_A = 0$, $\mu_F = -\frac{2}{w^\top b} Wb$ and $Z = uu^\top$, where $u^\top = (u_F^\top, u_A^\top)$,

$$u_F = -\frac{1}{\sqrt{w^\top b}} Wb,$$

and

$$u_A = \frac{1}{\sqrt{w^\top b}} \operatorname{diag}(M^{\frac{1}{2}}).$$

Notice that Z is a symmetric rank-one matrix. So Z is positive-semidefinite. Also,

$$Z_{FF} = u_F u_F^\top, \quad Z_{FA} = u_F u_A^\top, \quad Z_{AA} = u_A u_A^\top.$$

We can compute that

$$tr(Z_{AA}) = \frac{1}{w^\top b} d^\top q = 1,$$

and

$$2Z_{FA}M^{-\frac{1}{2}} = 2u_F u_A^\top M^{-\frac{1}{2}} = -\frac{2}{w^\top b} W b \mathbf{1}_A^\top = w\mu_A^\top + \mu_F \mathbf{1}_A^\top.$$

So (μ_A, μ_F, Z) is a feasible solution for the dual program (A.20). The objective value is

$$-\mu_A^\top Dq - \mu_F^\top b - tr(W^{-1}Z_{FF}) = b^\top \frac{2}{w^\top b} W b - b^\top \frac{1}{w^\top b} W b = \gamma.$$

So far, we have demonstrated that there exists a dual feasible solution for the dual program (A.20), such that the dual objective value is also γ. $(X,t) = (\frac{1}{v}Dq, \gamma)$ is indeed the optimal solution for program (A.19). Therefore, it is also the optimal solution for program (A.18).

Now we prove the other fact, $\overline{\lambda_{min}} \leq \gamma$.

Let $R = \operatorname{diag}(b)\operatorname{diag}(w)\operatorname{diag}(b)$. Take

$$T = M^{-1/2} X R X^\top M^{-1/2}.$$

Since T is similar to $XRX^\top M^{-1}$, they have the same eigenvalues. We re-write the original program (3.24) as

$$\begin{array}{llcl} \underset{X \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{F}|}}{\text{maximize}} & \lambda_{min}(T) & & \\ \text{subject to} & Dq & = & X \operatorname{diag}(b) w, \\ & \mathbf{1}_{\mathcal{A}}^\top X & = & \mathbf{1}_{\mathcal{F}}^\top, \\ & X & \geq & \mathbf{0}. \end{array} \tag{A.21}$$

Recall $v = \mathbf{1}^\top Dq = w^\top b$. Notice that

$$T = \left(\frac{1}{v}M\right)^{-1/2} X \left(\frac{1}{v}R\right) X^\top \left(\frac{1}{v}M\right)^{-1/2}.$$

Let $e = \mathrm{diag}(\left(\frac{1}{v}M\right)^{1/2})$. According to the constraints on X in program (A.21), we can show that $T \succeq \mathbf{0}$, $T \geq \mathbf{0}$, and

$$e^\top T e = \mathbf{1}_{\mathcal{A}}^\top X \left(\frac{1}{v}R\right) X^\top \mathbf{1}_{\mathcal{A}} = \mathbf{1}_{\mathcal{F}}^\top \left(\frac{1}{v}R\right) \mathbf{1}_{\mathcal{F}} = \gamma.$$

So we can construct the following program as a relaxed version of program (A.21) .

$$\begin{aligned} \underset{X \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{F}|}}{\text{maximize}} \quad & \lambda_{min}(T) \\ \text{subject to} \quad & e^\top T e = \gamma, \\ & T \succeq \mathbf{0}, \\ & T \geq \mathbf{0}. \end{aligned} \tag{A.22}$$

Program (A.22) is equivalent to the following.

$$\begin{aligned} \underset{X \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{F}|}}{\text{maximize}} \quad & t \\ \text{subject to} \quad & T - tI \succeq \mathbf{0}, \\ & e^\top T e = \gamma, \\ & T \succeq \mathbf{0}, \\ & T \geq \mathbf{0}. \end{aligned} \tag{A.23}$$

The dual of the program (A.23) is the following.

$$\begin{array}{rrcl} \underset{\beta\in\mathbb{R};S_1,S_2\in\mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}}{\text{minimize}} & \beta\gamma & & \\ \text{subject to} & Tr(S_1) & = & 1, \\ & S_1+S_2 & \leq & \beta e^\top e, \\ & S_1,\, S_2 & \succeq & \mathbf{0}. \end{array} \tag{A.24}$$

One observation is that $\beta = 1$, $S_1 = ee^\top$, and $S_2 = 0$ is dual feasible for program (A.24), since $Tr(ee^\top) = \sum_{i\in\mathcal{A}} \frac{D_{ii}q_i}{v} = 1$. Then we know that the optimal value of program (A.24) is no greater than γ. Since program (A.24) is a relaxation of program (A.21), $\overline{\lambda_{min}} \leq \gamma$. Now the result of this theorem follows immediately.

■

Theorem 12. *In any economy,*

$$\underline{\lambda_{max}} = \gamma. \tag{A.25}$$

Proof. This result follows immediately from the proof of Theorem 11. ■

Theorem 13. *Suppose we are given a low-leverage economy. The minimum value of of the maximal network amplifier can be achieved by a mutual-fund economy. Namely,*

$$X^* \triangleq x^*\mathbf{1}_{\mathcal{F}}^\top \in \underset{X\in X}{\operatorname{argmin}}\, MNA(X),$$

where

$$x^* = \frac{Dq}{\mathbf{1}_{\mathcal{A}}^\top Dq}.$$

For the holding network X^*, every firm invests in the same market portfolio, where assets are held in proportion to the market capitalization of assets.

Proof. This result follows immediately from the proof for Theorem 11. ■

Theorem 14. *For a low-leverage economy, suppose every firm has a constant leverage level $0 < \bar{b} < 1$, and $b = \bar{b}\mathbf{1}_{\mathcal{F}}$. Suppose the holding network $X \in \mathcal{X}$ is a non-negative matrix. Then the maximal network amplifier does not depend on the configuration of the holding network X.*

Proof. When every firm has a constant leverage level $0 < \bar{b} < 1, \gamma = \bar{b} < 1$. So this is a low-leverage case. We observe that for any feasible holding network $X \in \mathcal{X}$,

$$\begin{aligned}
\Pi\bar{\Theta}^\top Dq &= X \operatorname{diag}(b) \operatorname{diag}(w) \operatorname{diag}(b) X^\top D^{-1} \operatorname{diag}(q)^{-1} Dq \\
&= X \operatorname{diag}(b) \operatorname{diag}(w) \operatorname{diag}(b) X^\top \mathbf{1}_{\mathcal{A}} \\
&= X \operatorname{diag}(b) \operatorname{diag}(w) \operatorname{diag}(b) \mathbf{1}_{\mathcal{F}} \\
&= \bar{b} X \operatorname{diag}(b) \operatorname{diag}(w) \mathbf{1}_{\mathcal{F}} \\
&= \bar{b} X \operatorname{diag}(b) w \\
&= \bar{b} Dq.
\end{aligned}$$

This proves that for any feasible holding network $X \in \mathcal{X}$, $\bar{b}$ is always an eigenvalue of the matrix $Y(X) = X \operatorname{diag}(b) \operatorname{diag}(w) \operatorname{diag}(b) X^\top M^{-1}$. The corresponding eigenvector is Dq.

Since X is a matrix with non-negative components, $Y(X)$ is a matrix with non-negative components. Also, Dq is an eigenvector with strictly positive components. By Corollary 8.1.30, in Horn and Johnson, “Matrix Analysis”.**TODO:** add reference, the eigenvalue associated with Dq must be the spectral radius of $Y(X)$. So we can show that for any feasible holding network $X \in \mathcal{X}$, $\bar{b}$ is always the largest eigenvalue of $Y(X)$. Therefore the maximal network amplifier is always

$$\mathsf{MNA}(X) = \frac{1}{1-\bar{b}}.$$

■

Theorem 15. *Suppose we are given a high-leverage economy. The minimum value of the maximal network amplifier is bounded below by* $\frac{1}{\gamma-1}$.

$$\min_{X\in\mathcal{X}} \mathsf{MNA}(X) = \frac{1}{\overline{\lambda_{min}}-1} \geq \frac{1}{\gamma-1}.$$

Proof. Recall that in the proof of Theorem 11, we have established that $\overline{\lambda_{min}} \leq \gamma$. Since the economy is a high-leverage one, $\overline{\lambda_{min}} > 1$. According to the proof of Theorem 11, we immediately obtain that $1 < \overline{\lambda_{min}} \leq \gamma$ for a high-leverage economy. Therefore the maximal network amplifier

$$\min_{X\in\mathcal{X}} \mathsf{MNA}(X) = \frac{1}{\overline{\lambda_{min}}-1} \geq \frac{1}{\gamma-1}.$$

■

Theorem 16. *Suppose the leverage is* $b = \bar{b}\mathbf{1}_{\mathcal{F}}$ *where* $\bar{b} > 1$*, wealth is* $w = \bar{w}\mathbf{1}_{\mathcal{F}}$*, and the market capitalization is* $m = \bar{m}\mathbf{1}_{\mathcal{A}}$*. We also assume that* $|\mathcal{F}| = n|\mathcal{A}|$*, where* $n \in N$*. The following network*

X^* minimizes the maximal network amplifier.

$$X^* = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \tag{A.26}$$

Specifically,

$$X^*_{ih} = \begin{cases} 1 \quad , & \mathit{if}\ (i-1)n+1 \leq h \leq in, \\ 0 \quad , & \mathit{otherwise}. \end{cases} \tag{A.27}$$

Proof. First, X^* is a feasible solution for program (3.24). According to the definition of X^*, we obtain that

$$\begin{aligned} Y(X^*) &= X^* \operatorname{diag}(b) \operatorname{diag}(w) \operatorname{diag}(b) X^{*\top} M^{-1} \\ &= \bar{b} \frac{\bar{w}\bar{b}}{\bar{m}} \frac{|\mathcal{F}|}{|\mathcal{A}|} I_{\mathcal{A}} \\ &= \bar{b} I_{\mathcal{A}}. \end{aligned}$$

The last equality is because of the fact that the total market capitalization $\bar{m}|\mathcal{A}|$ equals the total wealth invested in risky assets $\bar{w}\bar{b}|\mathcal{F}|$. The minimum eigenvalue is

$$\lambda_{min}(Y(X^*)) = \bar{b}.$$

In this case, $MNA(X^*) = \frac{1}{\bar{b}-1}$.

We observe that $\overline{\lambda_{min}} \geq \lambda_{min}(Y(X^*)) = \bar{b} > 1$, so this is a high-leverage economy. We notice that $\gamma = \bar{b}$ for this symmetric case. Since the economy is a high-leverage one, $\overline{\lambda_{min}} > 1$. According to the proof of Theorem 11, we immediately obtain that $1 < \overline{\lambda_{min}} \leq \gamma = \bar{b}$. Therefore,

$$\min_{X \in X} \mathsf{MNA}(X) = \frac{1}{\overline{\lambda_{min}} - 1} \geq \frac{1}{\bar{b} - 1}.$$

We obtain that $\frac{1}{\bar{b}-1}$ is the minimum achievable level for the maximal network amplifier. So this feasible X^* is also the optimal holding network for the given condition. ■

Theorem 17. *Suppose A is fixed and we take $F \to +\infty$. For any given F, we generate a set of firms $\mathcal{F}$ where $|\mathcal{F}| = F$. The leverage and wealth of $\mathcal{F}$ satisfy the setting in this section. Indexed by F, we can construct a sequence of sets of feasible holding networks $\mathcal{X}^{(F)}$. Then, indexed by F, there exists a sequence of partitions $\mathcal{P}^{(F)}$ of $\mathcal{F}$ that defines a sequence of holding networks $X(\mathcal{P}^F)$ according to equation* (3.31) *such that the sequence of $X(\mathcal{P}^{(F)})$ is asymptotically the optimal solution to the problem of minimizing the maximal network amplifier. Namely, as $F \to +\infty$, $X(\mathcal{P}^F)$ is asymptotically a feasible solution. Specifically,*

$$X(\mathcal{P}^F) \in \mathcal{X}^{(F)}, \quad \textit{for all but finitely many } F\textit{, almost surely.}$$

Also, $X(\mathcal{P}^F)$ provides the optimality asymptotically.

$$MNA(X(\mathcal{P}^F)) - \min_{X \in \mathcal{X}^{(F)}} \mathsf{MNA}(X) \to_P 0.$$

Proof. To prove the theorem, we will construct the sequence of partitions $\mathcal{P}^F$ of $\mathcal{F}$. This theorem has two parts to prove. The first part is to prove that the sequence of $X(\mathcal{P}^F)$ asymptotically belongs to feasible holding networks $\mathcal{X}^{(F)}$. The second part is to prove that the sequence of $X(\mathcal{P}^F)$ asymptotically produces the optimal solution.

For a given F, we construct a partition $\mathcal{P}^F$ of $\mathcal{F}$. The partition consists of non-intersecting subsets of $\mathcal{F}$ specified by $\mathcal{P}^F = \{\mathcal{F}_i\}_{i\in\mathcal{A}} + \mathcal{R}$. Here $\mathcal{F}_i \subseteq \mathcal{F}$ denotes firms which invest in each asset i, and $\mathcal{R} \subseteq \mathcal{F}$ denotes the remaining firms. In this construction, only the cardinality $|\mathcal{F}_i|$ of $\mathcal{F}_i$ matters for this proof. So we first decide the cardinality of $\mathcal{F}_i$ for each $i \in \mathcal{A}$.

$$|\mathcal{F}_i| = \max\{\lceil \bar{m}_i F \rceil - 1 - \lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor, 0\}. \tag{A.28}$$

where $\varepsilon \in (0, \frac{1}{2})$ is a small number that is arbitrarily chosen and fixed. It is clear that $|\mathcal{F}_i| < \bar{m}_i F$, so $\sum_{i\in\mathcal{A}} |\mathcal{F}_i| \leq F$. For each $i \in \mathcal{A}$, we define

$$\mathcal{F}_i = \{h \in \mathcal{F} : \sum_{j=1}^{i-1} |\mathcal{F}_j| + 1 \leq h \leq \sum_{j=1}^{i} |\mathcal{F}_j|\}.$$

$$\mathcal{R} = \mathcal{F} \setminus \bigcup_{i\in\mathcal{A}} \mathcal{F}_i.$$

Now based on the partition $\mathcal{P}^F$, we can construct $X(\mathcal{P}^F)$ according to equation (3.31). To prove the asymptotic feasibility of $X(\mathcal{P}^F)$, we introduce some notations. For a given F, let E_F denotes the event that $X(\mathcal{P}^F) \geq 0$.

$$Pr(E_F) = 1 - Pr\left(\cup_{i\in\mathcal{A}}\{m_i' < 0\}\right) \geq 1 - \sum_{i\in\mathcal{A}} Pr\left(m_i' < 0\right).$$

In the above equation, the first equality is because the only possibility of $X(\mathcal{P}^F)$ has negative elements is from events $\{m_i' < 0\}$. The inequality is because of the relaxation of the union of events. We can bound $Pr(m_i' < 0)$. For each $i \in \mathcal{A}$, define

$$Z_i = \sum_{h \in \mathcal{F}_i} b_h w_h - \bar{m}_i \sum_{h \in \mathcal{F}} b_h w_h.$$

The mean of Z_i is

$$\mu_{Z_i} = (|\mathcal{F}_i| - \bar{m}_i F)\,\bar{b}\bar{w}.$$

Since $|\mathcal{F}_i| < \bar{m}_i F$, $\mu_{Z_i} < 0$. The variance of Z_i is

$$\sigma_{Z_i}^2 = \left(|\mathcal{F}_i| + \bar{m}_i^2 F\right) \sigma_{b_h w_h}^2.$$

For $Pr(m_i' < 0)$ we have

$$Pr\left(m_i' < 0\right) = Pr(Z_i > 0) = Pr(Z_i - \mu_{Z_i} > -\mu_{Z_i}) \leq Pr\left(|Z_i - \mu_{Z_i}| > \frac{|\mu_{Z_i}|}{\sigma_{Z_i}} \sigma_{Z_i}\right) \leq \frac{\sigma_{Z_i}^2}{\mu_{Z_i}^2}.$$

The first inequality is because $\mu_{Z_i} < 0$ and we relax the one-side condition to a two-side condition. The second inequality is due to Chebyshev's inequality. As $F \to +\infty$,

$$\frac{\sigma_{Z_i}^2}{\mu_{Z_i}^2} = \frac{\left(|\mathcal{F}_i| + \bar{m}_i^2 F\right)}{(|\mathcal{F}_i| - \bar{m}_i F)^2} \frac{\sigma_{b_h w_h}^2}{\bar{b}^2 \bar{w}^2} = \frac{\frac{\left(|\mathcal{F}_i| + \bar{m}_i^2 F\right)}{F}}{\left(\frac{|\mathcal{F}_i| - \bar{m}_i F}{F^{1/2}}\right)^2} \frac{\sigma_{b_h w_h}^2}{\bar{b}^2 \bar{w}^2} \to 0.$$

The above convergence is due to the following two observations.

$$\frac{\left(|\mathcal{F}_i| + \bar{m}_i^2 F\right)}{F} \to \bar{m}_i + \bar{m}_i^2. \tag{A.29}$$

$$\frac{|\mathcal{F}_i| - \bar{m}_i F}{F^{\frac{1}{2}}} \rightarrow (\bar{m}_i)^{\frac{1}{2}+\varepsilon} (F)^{\varepsilon} \rightarrow +\infty. \tag{A.30}$$

To show the above results, we first notice that in equation (A.28), the definition of $|\mathcal{F}_i|$ provides the following facts.

$$\frac{\lceil \bar{m}_i F \rceil - 1}{\bar{m}_i F} \in [1 - \frac{1}{\bar{m}_i F}, 1) \rightarrow 1. \tag{A.31}$$

$$\frac{\lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor}{F} = \frac{\lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor}{(\bar{m}_i F)^{\frac{1}{2}+\varepsilon}} \frac{\bar{m}_i}{(\bar{m}_i F)^{\frac{1}{2}-\varepsilon}} \rightarrow 0. \tag{A.32}$$

$$\frac{\lceil \bar{m}_i F \rceil - 1 - \bar{m}_i F}{F^{\frac{1}{2}}} \in [-\frac{1}{F^{\frac{1}{2}}}, 0) \rightarrow 0. \tag{A.33}$$

$$\frac{\lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor}{F^{\frac{1}{2}}} = \frac{\lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor}{(\bar{m}_i F)^{\frac{1}{2}+\varepsilon}} (\bar{m}_i)^{\frac{1}{2}+\varepsilon} (F)^{\varepsilon} \rightarrow (\bar{m}_i)^{\frac{1}{2}+\varepsilon} (F)^{\varepsilon} \rightarrow +\infty. \tag{A.34}$$

By equation (A.31) and equation (A.32), we obtain that

$$\frac{\lceil \bar{m}_i F \rceil - 1 - \lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor}{F} \rightarrow \bar{m}_i.$$

Then

$$\frac{|\mathcal{F}_i|}{F} \rightarrow \bar{m}_i, \tag{A.35}$$

and equation (A.29) immediately follows. By equation (A.33) and equation (A.34), we obtain that

$$\frac{\lceil \bar{m}_i F \rceil - 1 - \bar{m}_i F - \lfloor (\bar{m}_i F)^{\frac{1}{2}+\varepsilon} \rfloor}{F^{\frac{1}{2}}} \rightarrow (\bar{m}_i)^{\frac{1}{2}+\varepsilon} (F)^{\varepsilon} \rightarrow +\infty.$$

Then equation (A.30) follows. Now we have proved equation (A.29) and equation (A.30).

Then $Pr(m_i' < 0) \rightarrow 0$, and hence $Pr(E_F) \rightarrow 1$. By using the second Borel-Cantelli lemma, we obtain that

$$Pr\left(\limsup_{F\to\infty} E_F\right) = 1.$$

Namely, $X(\mathcal{P}^F) \geq 0$ for all but finitely many F, as $F \to +\infty$. Additionally, equation (3.31) provides that $X(\mathcal{P}^F)$ always satisfy all other constraints apart from $X(\mathcal{P}^F) \geq 0$ in the definition (17) of the feasible holding networks $\mathcal{X}^{(F)}$. Therefore $X(\mathcal{P}^F)$ asymptotically belongs to feasible holding networks $\mathcal{X}^{(F)}$ almost surely.

Now we prove that the sequence of $X(\mathcal{P}^F)$ asymptotically produces the optimal solution almost surely. According to equation (3.31), we can calculate $Y(X(\mathcal{P}^F))$.

$$Y(X(\mathcal{P}^F)) = X(\mathcal{P}^F)\,\mathrm{diag}(b)\,\mathrm{diag}(w)\,\mathrm{diag}(b)X(\mathcal{P}^F)^\top M^{-1} = A(X(\mathcal{P}^F)) + B(X(\mathcal{P}^F)),$$

where

$$A(X(\mathcal{P}^F)) = \begin{pmatrix} \frac{1}{m_1}\sum_{h\in\mathcal{F}_1} w_h b_h^2 & 0 & \dots & 0 \\ 0 & \frac{1}{m_2}\sum_{h\in\mathcal{F}_2} w_h b_h^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{m_{|\mathcal{A}|}}\sum_{h\in\mathcal{F}_2} w_h b_h^2 \end{pmatrix},$$

$$B(X(\mathcal{P}^F)) = \begin{pmatrix} \frac{\alpha^2}{m_1}|\mathcal{R}|m_1' m_1' & \frac{\alpha^2}{m_2}|\mathcal{R}|m_1' m_2' & \dots & \frac{\alpha^2}{m_2}|\mathcal{R}|m_1' m_{|\mathcal{A}|}' \\ \frac{\alpha^2}{m_1}|\mathcal{R}|m_2' m_1' & \frac{\alpha^2}{m_2}|\mathcal{R}|m_2' m_2' & \dots & \frac{\alpha^2}{m_2}|\mathcal{R}|m_2' m_{|\mathcal{A}|}' \\ \dots & \dots & \dots & \dots \\ \frac{\alpha^2}{m_1}|\mathcal{R}|m_{|\mathcal{A}|}' m_1' & \frac{\alpha^2}{m_2}|\mathcal{R}|m_{|\mathcal{A}|}' m_2' & \dots & \frac{\alpha^2}{m_2}|\mathcal{R}|m_{|\mathcal{A}|}' m_{|\mathcal{A}|}' \end{pmatrix} = \alpha^2|\mathcal{R}|m' m'^\top M^{-1}.$$

In the above formulation, matrix $A(X(\mathcal{P}^F))$ characterizes the asset-to-asset impact generated by $\cup_{i\in\mathcal{A}}\mathcal{F}_i$. Meanwhile, matrix $B(X(\mathcal{P}^F))$ characterizes the asset-to-asset impact generated by the remaining firms $\mathcal{R}$. For each asset $i \in \mathcal{A}$,

$$A_{ii}(X(\mathcal{P}^F)) = \frac{\sum_{h\in\mathcal{F}_i} w_h b_h^2}{m_i} = \frac{\frac{\sum_{h\in\mathcal{F}_i} w_h b_h^2}{|\mathcal{F}_i|}}{\frac{\sum_{h\in\mathcal{F}} w_h b_h}{F}} \frac{|\mathcal{F}_i|}{\bar{m}_i F} \tag{A.36}$$

As $F \to +\infty$, $|\mathcal{F}_i| \to +\infty$. Recall equation (A.35),

$$\frac{|\mathcal{F}_i|}{\bar{m}_i F} \to 1.$$

In our construction, each b_h and w_h are in fact indexed by F implicitly. For a given F, each b_h and w_h are drawn from the given distribution for $h = 1 \dots F$. For different F, b_h and w_h are a different set of realizations of the same i.i.d. random variables. So we identify a triangular array of i.i.d. random quantities. We can apply the weak law of large numbers for triangular arrays. So the other components of equation (A.36) converge.

$$\frac{\sum_{h\in\mathcal{F}_i} w_h b_h^2}{|\mathcal{F}_i|} \to E[w_h b_h^2], \quad \text{in probability.}$$

$$\frac{\sum_{h\in\mathcal{F}} w_h b_h}{F} \to E[w_h b_h], \quad \text{in probability.}$$

Therefore, as $F \to +\infty$,

$$A_{ii}(X(\mathcal{P}^F)) \to \bar{\gamma}, \quad \text{in probability.} \tag{A.37}$$

For every asset i and j,

$$B_{ij}(X(\mathcal{P}^F)) = \frac{\alpha^2}{m_j}|\mathcal{R}|m_i^{'}m_j^{'} \leq \frac{|\mathcal{R}|}{m_j} \to 0, \quad \text{in probability.} \tag{A.38}$$

In the above expression, the inequality is due to the fact that as $F \to +\infty$,

$$\alpha^2 m_i^{'}m_j^{'} = \frac{m_i^{'}}{\mathbf{1}_{|\mathcal{A}|}^\top m'}\frac{m_j^{'}}{\mathbf{1}_{|\mathcal{A}|}^\top m'} \leq 1, \quad \text{almost surely,}$$

since $m_i^{'} \geq 0$ almost surely for any i. In addition, the statement $\frac{|\mathcal{R}|}{m_j} \to_P 0$ in equation (A.38) can be proved by the weak law of large numbers for triangular arrays.

$$\frac{|\mathcal{R}|}{m_j} = \frac{F - \sum_{k\in\mathcal{A}}|\mathcal{F}_k|}{F}\frac{1}{\frac{\bar{m}_j\sum_{h\in\mathcal{F}}w_h b_h}{F}} \to \left(1 - \sum_{k\in\mathcal{A}}\bar{m}_k\right)\frac{1}{\bar{m}_j E[w_h b_h]} \to 0, \quad \text{in probability.}$$

Therefore

$$B(X(\mathcal{P}^F)) \to \mathbf{0}, \quad \text{in probability.}$$

We have obtained that

$$Y(X(\mathcal{P}^F)) = A(X(\mathcal{P}^F)) + B(X(\mathcal{P}^F)) \to \bar{\gamma}I, \quad \text{in probability.}$$

We conclude that asymptotically the dominant term is $A(X(\mathcal{P}^F))$, i.e., the asset-to-asset impact generated by $\cup_{i\in\mathcal{A}}\mathcal{F}_i$. The other contributing term $B(X(\mathcal{P}^F))$ that characterizes the asset-to-asset impact generated by the remaining firms $\mathcal{R}$ vanishes as $B(X(\mathcal{P}^F)) \to \mathbf{0}$.

Since eigenvalues are continuous functions of matrix elements, according to the continuous mapping theorem, it follows that

$$\lambda_i(Y(X(\mathcal{P}^F))) \to \bar{\gamma}, \quad \text{in probability}, \quad \forall\, i.$$

We immediately know that

$$\lambda_{min}(Y(X(\mathcal{P}^F))) \to \bar{\gamma}, \quad \text{in probability}. \tag{A.39}$$

Also, we notice when $F \to +\infty$,

$$\gamma^F = \frac{\sum_{h\in\mathcal{F}} w_h b_h^2}{\sum_{h\in\mathcal{F}} w_h b_h} = \frac{\frac{\sum_{h\in\mathcal{F}} w_h b_h^2}{F}}{\frac{\sum_{h\in\mathcal{F}} w_h b_h}{F}} \to \bar{\gamma}, \quad \text{in probability}. \tag{A.40}$$

We recall from the proof of Theorem 11 that for any given F and generated wealth w and leverage b of firms, the following inequalities always hold.

$$\lambda_{min}(X(\mathcal{P}^F))\mathbf{1}_{\{X(\mathcal{P}^F)\in\mathcal{X}^{(F)}\}} \le \max_{X\in\mathcal{X}^{(F)}} \lambda_{min}(Y(X)) \le \gamma^F. \tag{A.41}$$

The first inequality is because when $X(\mathcal{P}^F) \in \mathcal{X}^{(F)}$, the feasible solution doesn't exceed the optimal solution, and the optimal solution is non-negative. The second inequality is from the proof of Theorem 11.

According to equation (A.40), the right hand side of inequality (A.41) converges to $\bar{\gamma}$ in probability. For the first part of this theorem, we have proved that the sequence of $X(\mathcal{P}^F)$ that we construct satisfies that

$$X(\mathcal{P}^F) \in \mathcal{X}^{(F)}, \quad \text{for all but finitely many } F, \text{ almost surely.}$$

This implies that $\mathbf{1}_{\{X(\mathcal{P}^F)\in\mathcal{X}^{(F)}\}}$ converges to 1 almost surely. With the result in equation (A.39), we apply Slutsky's theorem to obtain that $\lambda_{min}(X(\mathcal{P}^F))\mathbf{1}_{\{X(\mathcal{P}^F)\in\mathcal{X}^{(F)}\}}$ converges to $\bar{\gamma}$ in distribution. Since $\bar{\gamma}$ is a constant, the left hand side of inequality (A.40) also converges to $\bar{\gamma}$ in probability.

According to the sandwich theorem for convergence, we can conclude that

$$\max_{X\in\mathcal{X}^{(F)}} \lambda_{min}(Y(X)) \to \bar{\gamma}, \quad \text{in probability.} \tag{A.42}$$

Recall that $MNA(X(\mathcal{P}^F)) = \frac{1}{\lambda_{min}(Y(X(\mathcal{P}^F)))-1}$ when $\lambda_{min}(Y(X(\mathcal{P}^F))) > 1$. Let $E_S^{(F)}$ denote the event when $MNA(X(\mathcal{P}^F)) = \frac{1}{\lambda_{min}(Y(X(\mathcal{P}^F)))-1}$. Let $\bar{E}_S^{(F)}$ denote the complementary event of $E_S^{(F)}$.

According to equation (A.39) and the fact that $\bar{\gamma} > 1$, we can show that for any given $\delta > 0$,

$$\begin{aligned}
& Pr(|MNA(X(\mathcal{P}^F)) - \tfrac{1}{\bar{\lambda}-1}| > \delta) \\
= \;& Pr(|MNA(X(\mathcal{P}^F)) - \tfrac{1}{\bar{\lambda}-1}| > \delta,\ E_S^{(F)}) \\
& + Pr(|MNA(X(\mathcal{P}^F)) - \tfrac{1}{\bar{\lambda}-1}| > \delta \mid \bar{E}_S^{(F)}) Pr(\bar{E}_S^{(F)}) \\
\to \;& Pr(|MNA(X(\mathcal{P}^F)) - \tfrac{1}{\bar{\lambda}-1}| > \delta,\ E_S^{(F)}) \\
= \;& Pr(|\tfrac{1}{\lambda_{min}(Y(X(\mathcal{P}^F)))-1} - \tfrac{1}{\bar{\lambda}-1}| > \delta,\ E_S^{(F)}) \\
\leq \;& Pr(|\tfrac{1}{\lambda_{min}(Y(X(\mathcal{P}^F)))-1} - \tfrac{1}{\bar{\lambda}-1}| > \delta) \\
\to \;& 0.
\end{aligned}$$

The first convergence is because as F goes to infinity, the probability of $E_S^{(F)}$ goes to 1, and the probability of $\bar{E}_S^{(F)}$ goes to 0. The last convergence is because of the continuous mapping theorem and equation (A.39). Now we conclude that the level of the maximal network amplifier converges.

$$MNA(X(\mathcal{P}^F)) \to \frac{1}{\bar{\gamma}-1}, \quad \text{in probability.} \tag{A.43}$$

Recall that $\min_{X \in \mathcal{X}^{(F)}} \mathsf{MNA}(X) = \frac{1}{\max_{X \in \mathcal{X}^{(F)}} \lambda_{min}(Y(X)) - 1}$ when $\max_{X \in \mathcal{X}^{(F)}} \lambda_{min}(Y(X)) > 1$. Let $E_{HL}^{(F)}$ denote the following event

$$E_{HL}^{(F)} = \left\{ \min_{X \in \mathcal{X}^{(F)}} \mathsf{MNA}(X) = \frac{1}{\max_{X \in \mathcal{X}^{(F)}} \lambda_{min}(Y(X)) - 1} \right\}.$$

Let $\bar{E}_{HL}^{(F)}$ denote the complementary event of $E_{HL}^{(F)}$. Since $\bar{\gamma} > 1$, asymptotically the economy is a high-leveraged economy. Namely, as F goes to infinity, the probability of the economy being a high-leveraged economy converges to 1. The probability of $E_{HL}^{(F)}$ goes to 1, and the probability of $\bar{E}_{HL}^{(F)}$ goes to 0.

According to equation (A.42) and the fact that $\bar{\gamma} > 1$, we can show that for any given $\delta > 0$,

$$
\begin{aligned}
& Pr(|\min_{X\in\mathcal{X}^{(F)}} \mathsf{MNA}(X) - \frac{1}{\bar{\lambda}-1}| > \delta) \\
= \;& Pr(|\min_{X\in\mathcal{X}^{(F)}} \mathsf{MNA}(X) - \frac{1}{\bar{\lambda}-1}| > \delta,\ E_{HL}^{(F)}) \\
& + Pr(|\min_{X\in\mathcal{X}^{(F)}} \mathsf{MNA}(X) - \frac{1}{\bar{\lambda}-1}| > \delta \mid \bar{E}_{HL}^{(F)}) Pr(\bar{E}_{HL}^{(F)}) \\
\to \;& Pr(|\min_{X\in\mathcal{X}^{(F)}} \mathsf{MNA}(X) - \frac{1}{\bar{\lambda}-1}| > \delta,\ E_{HL}^{(F)}) \\
= \;& Pr(|\frac{1}{\max_{X\in\mathcal{X}^{(F)}} \lambda_{min}(Y(X)) - 1} - \frac{1}{\bar{\lambda}-1}| > \delta,\ E_{S}^{(F)}) \\
\leq \;& Pr(|\frac{1}{\max_{X\in\mathcal{X}^{(F)}} \lambda_{min}(Y(X)) - 1} - \frac{1}{\bar{\lambda}-1}| > \delta) \\
\to \;& 0.
\end{aligned}
$$

The last convergence is because of continuous mapping theorem and equation (A.42).

Therefore we conclude that

$$
\min_{X\in\mathcal{X}^{(F)}} \mathsf{MNA}(X) - \frac{1}{\bar{\gamma}-1} \to 0, \quad \text{in probability.} \tag{A.44}
$$

By equation (A.43) and equation (A.44), we obtain that

$$
MNA(X(\mathcal{P}^F)) - \min_{X\in\mathcal{X}^{(F)}} \mathsf{MNA}(X) \to_P 0.
$$

Now we have proved the second part of this theorem. The construction of the asymptotically isolated holding network provides the optimal solution. ■