Three essays on the optimal allocation of risk with illiquidity, intergenerational sharing and systemic institutions

Dimitrov, D.K.

Publication date
2022

Document Version
Final published version

tinbergen institute

This dissertation considers three problems of risk allocation and applies approaches from theoretical finance and risk management to address several policy debates from a macro-finance point of view. The first essay included in the thesis addresses a classical finance problem of allocating risks efficiently in an investment portfolio. Non-triviality arises when there is uncertainty in the immediate availability of a market for the savings assets. The second essay puts the risk allocation problem into a policy-relevant setting by considering how illiquidity, in the form of trading costs in the retirement savings portfolio of the elderly, affects the ability of different generations to share financial risks with each other. The third essay shows how monitoring and evaluating systemic risk in a financial network can be done through a risk management lens, using a credit portfolio approach based on implied market views from the Credit Default Swap market.

Daniel Dimitrov holds a B.Sc. in Economics from the University of Magdeburg and a M.Sc. in Econometrics and Mathematical Economics from Tilburg University. He is a Ph.D. candidate at the University of Amsterdam under the supervision of Prof. Roel Beetsma and Prof. Sweder van Wijnbergen. His research areas lie in the field of macro-finance and asset pricing.

Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions   Daniel Dimitrov

803

Daniel Dimitrov

Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions

Universiteit van Amsterdam

Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions

Daniel K. Dimitrov

ISBN: TBD

Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J Maex

ten overstaan van een door het College voor Promoties ingestelde

commissie, in het openbaar te verdedigen in de Agnietenkapel

op maandag 29 november 2022, te 12.00 uur

door Daniel K. Dimitrov

geboren te Sofia, Bulgarije

Promotiecommissie:

Promotores	Prof. Dr. S.J.G. van Wijnbergen	*University of Amsterdam*
	Prof. dr. R.M.W.J. Beetsma	*University of Amsterdam*
Overige leden:	Prof. Dr. C. van Ewijk	*University of Amsterdam*
	Dr. W.E. Romp	*University of Amsterdam*
	Prof. dr. J.M. Schumacher	*University of Amsterdam*
	Prof. Dr. B.J.M. Werker	*Tilburg University*
	Prof. Dr. D.W.G.A. Broeders	*Maastricht University*

Faculteit: Economie en Bedrijfskunde

You shall possess the good of the earth and sun, (there are millions of suns left,)
You shall no longer take things at second or third hand, nor look through the eyes of the dead, nor feed on the spectres in books,
You shall not look through my eyes either, nor take things from me,
You shall listen to all sides and filter them from your self.

– Walt Whitman, *Song to Myself*

So the guy says, "What are you doing? You come to fix the radio, but you're only walking back and forth!" I say, "I'm thinking!" Then I said to myself, "All right, take the tubes out, and reverse the order completely in the set." [...] "He fixes radios by thinking!" The whole idea of thinking, to fix a radio - a little boy stops and thinks, and figures out how to do it - he never thought that was possible.

– Richard Feynman, *Surely You're Joking, Mr. Feynman!*

Acknowledgments

I would like to thank many people for helping me along the way in making this thesis possible.

First of all, I am grateful to my supervisors, Roel Beetsma and Sweder van Wijnbergen. Roel in the very beginning saw potential in my Ph.D. application, and even though I was from the industry and disconnected from academia at the time, gave me a chance. His honest, and timely feedback on my drafts taught me the patience and diligence needed to make sound research. Sweder's scientific curiosity and "pleasure of finding things out" is what I will always aspire to have in research. I am grateful that he saw the potential synergies between my industry experience, policy work, and the current academic debate, and engaged me in a very interesting projects on systemic risk which gave birth to the third chapter of this thesis. I am thankful to all the visiting researchers, and all seminar participants for the helpful discussions and comments. I am grateful to the teachers at Tinbergen Institute. They showed me that science is an open door.

I am much grateful to my colleagues at the UvA. It has been a privilege to spend time among all the young and smart people that I met at the UvA. Especially: Stan, who was also a guide to everything Dutch and happily, or at least without complaints, enjoyed our Balkan company, which dominated the office room; Ioana, who was our experienced guide to the realities and merits of Ph.D. life; Merve that I regularly discussed Balkan culture and politics with; our UvA MInt colleagues Christian, Ward, Elisabeth, Franc, Kees, Dirk - I am thankful for the all the lively discussions over lunch; the TI colleagues Stefan, Konstantin, Daniel, Lisa, Xu Lin who joined our ranks; Massimo Giuliodori for the great teaching experience I had in tutoring his course in Fiscal and Monetary Policy; Marcelo and Andras for being great tutor buddies.

I would also like to acknowledge the financial support I have received from MN services, who partially funded this Ph.D., as well as De Nederlandsche Bank who partially funded my stay within the Financial Stability division, where I worked on the last chapter of the thesis. I am grateful to the Financial Stability team and the Research team for the support and the fruitful discussions.

I have to mention the great opportunity to take part in the Open Source Economics Bootcamp 2019 at the University of Chicago where I learned Python and its application to quantitative Economics. Those six weeks of great lecturers, and hard work paid off eventually and will continue to do so over the coming years. I am grateful for this to Rick Evans and to all participants and tutors who made this a great experience of learning and collaboration.

Finally, I am grateful to my family – my sisters Anna and Asya; who encouraged and supported me unconditionally, while they on their own followed a similar path; my father, who was always available to talk to; my mom, who I am sure would have been very proud. I am grateful to Hristina, who agreed to follow me on this journey.

List of authors

Strategic Asset Allocation with Private Assets: Untangling Illiquidity

This paper is single-authored and based on my master thesis at Tilburg University in 2017 under the name of *Portfolio Choice with Liquidity Frictions* where it was supervised by Prof. Bas Werker. The research question under investigation and the numerical methods used were improved significantly and the paper was re-written.

Intergenerational Risk Sharing with Market Liquidity Risk

This paper is single-authored based on an idea by Prof. Roel Beetsma, who also closely supervised and reviewed the project.

Quantifying Systemic Risk in the Presence of Unlisted Banks: Application to the Dutch Financial Sector

This paper is co-authored with Prof. Sweder van Wijnbergen. I developed the framework, implemented the numerical methods and wrote the paper with Sweder. I partially worked for this thesis during my research visit at De Nederlandsche Bank (DNB). I gratefully acknowledge the financial support and the data access granted by DNB. However, the views expressed in this paper are those of the authors, and not necessarily those of the DNB or its committees.

Contents

1

Introduction

In this thesis, we consider three non-trivial problems of risk allocation and apply approaches from theoretical finance and risk management to address several policy debates from a macro-finance point of view. The three topics considered are diverse, but there is a common theme that runs through each chapter. In each, we focus on the resolution of barriers to a first-best risk allocation rule. The barriers can be in the form of market illiquidity, current generations not being able to participate in the shocks affecting future generations, or risk spillovers between systemic institutions.

The first essay in Chapter 2 addresses a classical finance problem of allocating risks efficiently in an investment portfolio, in the spirit of Merton (1969). In our set-up, illiquidity in one of the assets exists as uncertainty in the immediate availability of a market where price risk can be traded, as in Ang et al. (2014). As a result, illiquidity acts as an additional non-hedgeable risk component, affecting the portfolio choice decision. We illustrate that the temporal dimension of liquidity naturally pushes the problem in the area of dynamic portfolio choice, leading us away from the standard static portfolio optimization (Markowitz, 1952) that is ill-equipped to tackle illiquidity risk. From that point of view, we show how dynamic programming techniques can have practical applications in asset allocation. We tailor the problem to an issue that long-term investors (pension funds, endowments, etc.) with private asset classes in their investment mix typically face. We determine the optimal strategic and the tactical portfolio weights for investors in hedge funds, private equity, real estate, and infrastructure.

Furthermore, we show how illiquidity endogenizes the risk aversion of investors, making essentially the portfolio choice and consumption decisions functions of the share of illiquid wealth. Quantitatively, for a reasonable parametrization, we find that if the investor cannot trade the asset over five years on average, the welfare loss, quantified as Certainty Equivalent Consumption loss, is about 2% relative to an equivalent completely liquid mix. Still, adding to the investment strategic mix private asset classes, for which secondary markets do not exist or are too frail, increases significantly the investor's welfare relative to a traditional equity-bonds allocation.

The second essay in Chapter 3 puts the allocation problem into a policy-relevant setting by considering how illiquidity in the form of uncertain trading costs affects the ability of different generations to share financial risk with each other. In that setting, the optimal allocation is considered across generations with partially illiquid savings.

We use a stylized two-period overlapping generations framework, where each generation makes a portfolio allocation decision for retirement. In this context, designing optimal social security institutions can also be seen as a risk management problem in the spirit of Shiller (1999). A policymaker balances, on one hand, the benefits of a wider risk-bearing pool by integrating the young into the financial shock affecting the currently old, and on the other, the costs of potentially destabilizing the young's retirement savings over time with risk being imported into the youth's starting wealth.

In this essay, we show that illiquidity reduces the range of transferable shocks between generations and thus lowers the potential benefits from risk-sharing. We still find that a contingent transfers policy based on a reasonably parametrized savings portfolio with liquid and illiquid assets increases aggregate welfare. However, higher illiquidity justifies higher levels of risk sharing to compensate for the trading friction.

The third essay in Chapter 4 looks at a different market failure problem. The default of a systemically important institution (SIFI) presents an externality when the failure has the potential to generate fire sales and risk spillovers to other institutions. Regulators have resorted to charging SIFIs with systemic capital buffers as a way of enhancing the loss-absorbing capacity of the system. In this chapter, we show how the macro-prudential problem of monitoring and evaluating institutions in the context of the risks of other institutions in the system can be seen through a risk management lens as well. The supervisor takes ownership of the potential losses in case of system-wide distress, so implicitly it needs to manage a portfolio of defaultable loans, and its problem can be seen as one of minimizing the tail risk of the portfolio.

We, thus, propose a credit portfolio approach for evaluating systemic risk and attributing it across institutions by constructing a model that can be estimated from high-frequency CDS data. This captures risks from privately held institutions and cooperative banks, extending approaches that rely on information from the public equity market (Adrian and Brunnermeier, 2016; Acharya et al., 2017). We account for correlated losses between the institutions, overcoming a modeling weakness in earlier studies. A latent risk factor with heterogeneous exposures fitted on the default probabilities implied from single-name CDSs quantifies the potential for joint distress and losses. We apply the model to a universe of Dutch banks and insurers and show how the framework can be used to complement the current regulatory framework used for setting optimal systemic buffers (EBA, 2020).

2

Strategic Asset Allocation with Private Assets: Untangling Illiquidity

2.1 Introduction

This paper examines the problem of optimal strategic asset allocation (SAA) for long-term investors with illiquid asset classes in the investment universe. The liquidity friction that we consider appears in the form of trading uncertainty in a dynamic portfolio choice setting. We address several imminent problems that follow from the lack of predictable liquid trading opportunities. We solve for the optimal portfolio allocations and optimal consumption of a utility maximizing agent and determine the size of the consumption cost and the size of the liquidity premium that an investor would require for holding an illiquid asset, otherwise comparable to public equity. For a reasonable calibration of the model, we find that adding a private asset class to the strategic allocation mix, such as hedge funds, private equity, direct real estate, or infrastructure, increases significantly the certainty equivalent consumption (CEC) of a long-term investor, despite the lack of reliable secondary markets for such assets.

[0]Acknowledgements: I am grateful to my supervisor Roel Beetsma for his suggestions and support in polishing up this paper; Bas Werker who commented on an earlier draft as a master thesis supervisor at Tilburg University; Albert J. Menkveld, Thierry Foucault and participants in the Market Microstructure Summer School for the valuable feedback and suggestions.

Private (or also alternative) asset classes have become a widely accessed investment option for long-term institutional investors, including pension funds, sovereign wealth funds, and endowments (Andonov et al., 2015, 2021). The common wisdom is that these investors are well-poised to exploit any long-term premia and diversification opportunities that such asset classes may offer. At the same time, they may face spending targets or the need to accommodate liabilities arising from participants that enter the retirement phase. This raises the need for a clear understanding of how to handle their specific liquidity properties in the strategic allocation process, where a secondary market usually does not exist or is very frail. As relocation cannot be achieved at demand, this will inevitably have an *ex-ante* effect on the allocation decision, where in anticipation of the lack of liquidity, the agent adjusts in advance the strategic asset weights compared to a fully liquid market. At the same time, in periods without liquidity, the agent adjusts tactically the allocation to the liquid classes.

We base our analysis on a model by Ang et al. (2014) that introduces uncertainty in the frequency with which an asset can be traded. This is an extension to the standard model of intertemporal portfolio optimization (Merton, 1969, 1971, 1973). We further extend the framework by adding multiple liquid assets to the investment universe thus bringing the model closer to the strategic allocation process used in practice. In the process, we establish an efficient numerical algorithm to solve the resulting portfolio choice problem.

We approximate the decision process of long-term institutional investors by an infinite horizon dynamic allocation problem with intermediate consumption. The nature of liabilities for many institutional investors, such as spending rules or continuous payment of retirement benefits, justify the modeling choice of agents with intermediate consumption. The dynamic nature of the problem then arises by considering the temporal dimension of liquidity. In this interpretation, liquidity represents the opportunity to trade, occurring randomly with intensity determined by the period over which the asset cannot be traded on average. We find that the effect of illiquidity on the portfolio allocations is economically significant when calibrated to major asset classes.

This paper then looks at the asset allocation problem from several different angles. First, we review the baseline dynamic model of portfolio choice with continuous trading (Merton, 1971) as a benchmark model whose implications are relevant for the investment in traditional asset classes, such as money market funds, bonds, and equity. Then we reformulate the problem by introducing illiquidity and investigate the properties of the resulting optimal solution. Finally, we apply the models to the SAA problem of long-term investors with access to illiquid private asset classes along with traditional liquid assets.

The framework that we employ allows us to make a distinction between SAA, as the asset class portfolio weights that an investor will come back to at every opportunity when liquidity is available, and tactical asset allocation (TAA), as the sub-optimal achievable weights that an investor can resort to, when liquidity is available only for a subset of the assets in the portfolio. This provides a way to factor in *ex-ante* the costs associated with illiquidity and its effect on the optimal consumption and allocation across assets.

The illiquid asset cannot be traded for periods of random length. This exposes the investor to unhedgeable liquidity risks, as the portfolio allocations cannot be adjusted when price changes lead to sub-optimal allocation. Also, the investor can consume out of liquid wealth and is allowed to transfer between the liquid and illiquid wealth only with uncertain timing. This gives rise to the risk of not being able to meet consumption needs when wealth is locked up in illiquid holdings. Liquid and illiquid wealth are thus imperfect substitutes and the investor will try to minimize the disutility coming from states of the world where liquid wealth cannot finance consumption.

Numerically, we confirm the intuitive premise that with higher illiquidity, the investor will lower *ex-ante* the strategic allocation to the illiquid asset. The liquidity friction, however, begins to have economic significance when an average waiting time between trading events is above two years. At two years, the strategic allocation to the illiquid asset gets significantly reduced below the continuous trading case and the certainty equivalent cost of illiquidity starts increasing notably[1].

As far as the TAA is concerned, we find that at high shares of illiquid holdings and without foreseeable ability to trade, the investor will tend to reduce sharply the allocation to the liquid risky asset in favor of cash in order to reduce liquid wealth volatility. The share of illiquid risky holdings thus acts as a state variable that also determines the optimal investment in the liquid asset (Figure 2.2b).

[1] See Table 2.2 for details.

A simulation shows that for levels of the friction compatible with the most common alternative investments, the agent can rebalance often enough, so that disproportionately large shares of illiquid wealth are avoided and the tactical weights remain in proximity of the SAA. Overall, the liquidity friction tends to reduce only marginally the average tactical liquid asset holdings. Still, it creates a fat left tail in the distribution of the liquid risky asset weights, as the tactical weights of the liquid assets get adjusted to uncontrolled movements in the illiquid wealth share.[2]

The lack of a full control on the illiquid asset share fluctuations also leads to a lower average consumption rate when a strong liquidity friction is present. The investor does not have a way of offsetting drops in the value of the liquid asset during periods in which the illiquid cannot be traded. Thus, even though consumption is determined directly only by the level of liquid wealth, fluctuations in illiquid wealth result in fluctuations in the consumption share and thus endogenize the investor's risk aversion. This results in a fat left tail in the distribution of the consumption rate over time (Figure 2.3e).

To finance consumption over time, the agent needs to keep a higher cash cushion as liquidity friction increases (Figure 2.1a). As the friction increases, the overall allocation to risky assets is decreased. As a result, over the long run, the investor needs to stay more conservative and cannot fully exploit the market risk premium. As a result, the investor's wealth and consequently, consumption level is lower over the long run compared to the completely liquid case (Figure 2.3).

Evaluating the welfare implications of illiquidity, we find that with a five-year friction the CEC is reduced by 5% compared to an identical completely liquid investment (Table 2.2). Intuitively, this reduction in risk-free equivalent consumption can be seen as triggered through two channels. First, the reduced strategic holdings in the illiquid asset lowers the diversification potential of the private asset in the portfolio. Second, the fact that consumption can be financed out of liquid wealth only, implies a higher risk that liquidity could be exhausted before a trading opportunity arrives.

Similarly, we find a clear positive dependency between the severity of the illiquidity, the consumption cost, and the liquidity premium. An asset that cannot be traded for ten

[2]See histogram plots in Figure 2.3f for details.

years on average would require a premium of at least 1.26% above the expected return of an otherwise identical liquid asset (Table 2.2).

In addition, the diversification benefits from the low correlation between the liquid and illiquid assets are muted relative to the model with no frictions. The CE cost is much stronger for assets with high negative correlation, which could otherwise diversify the portfolio. The cost approaches zero for assets that have almost no diversification potential (Table 2.4a).

The paper continues as follows: Section 2.2 provides an overview of the related literature; Section 2.3 outlines the benchmark continuous trading model and provides the corresponding analytical solutions; Section 2.4 provides the basic quantitative structure of the market with liquidity friction; Section 2.5 investigates the properties of the illiquid model and looks into the potential for liquidity premia and illiquidity costs of a synthetic asset comparable to public equities; and finally, Section 2.6 investigates to strategic allocations for several private asset classes that typically long-term investors can typically access.

2.2 Related Literature

First, we relate to the broader literature on dynamic portfolio choice with trading frictions. The frictions can broadly be classified along three main dimensions: transaction costs, delayed execution, and inability to access a market. The aspect of market illiquidity that each direction considers has specific implications on the optimal strategies recommended for investors.

The first dimension, where asset liquidity can be accessed at any point in time by paying a transaction cost, has been widely explored (Zabel, 1973; Magill and Constantinides, 1976; Gennotte and Jung, 1994; Boyle and Lin, 1997). In essence, this approach examines portfolio choice with a riskless asset and a set of risky assets subject to proportional trading costs. Most studies find that with the transaction costs it becomes optimal for agents to trade only if the illiquid asset allocation moves outside of a no-trade region. In general equilibrium, Constantinides (1986) determines that, as a result, overall trading is significantly reduced even for modest transaction costs. By trading smaller amounts and less frequently, agents accommodate the transaction costs without sacrificing consumption. Consequently, they do not need a significant liquidity premium to compensate for the friction. This argument is debated for example by Dai et al. (2011) who find that with the interaction between asset trading costs and position constraints within investors' portfolios large liquidity premia may arise.

Later studies introduce a liquid risky asset along with the risky illiquid one and emphasize that having access also to liquid risky investments, creates hedging and diversification motives, thus reducing the utility cost of the transaction friction, and alleviating the impact that illiquidity on one asset has on the investment in other assets and the pressure that this creates on prices (Buss et al., 2015; Bichuch and Guasoni, 2018). To capture these effects fully, we consider liquid and illiquid risky assets. The curse of dimensionality makes such problems harder to solve. We apply a numerical algorithm in the spirit of Cai et al. (2013), who offer a recipe relying on quadrature approximations and value function iteration for solving non-trivial portfolio choice problems with multiple assets subject to transaction costs. We adapt the algorithm to be compatible with our problem.

Longstaff (2001) explores the second dimension, where illiquidity appears as a friction on the quantity that can be traded per period. In his setup, a constraint exists on the speed with which trades can be executed, so larger trades are executed slowly at a known rate. By waiting for a deterministic period, the agent is then subject to price risk without complete control over the allocations in the portfolio and has to hedge against both expected and unexpected portfolio weight changes. In contrast to the previous set of studies, they find that the agents optimally trade whenever needed and as much as needed.

Across the third dimension, Miklós and Ádám (2002) approximate the inability to trade an asset immediately by introducing a deterministic lag between the time an order is placed on the market and the time a trade actually takes place. Ang et al. (2014), on which our study is closely based, further extend this idea by adding uncertainty in the waiting time. The idea of uncertain waiting time is closely related to the OTC markets where search frictions usually imply that the time between transactions is stochastic (Diamond, 1982; Duffie et al., 2004) as counterparties need to find each other to trade. This is especially relevant for the private asset classes that we consider later on, such as private equity and real estate, for which an exchange does not exist and a counterparty to the trade is usually not immediately available. A similar approach, with uncertain trading time under a limited horizon set up, is applied by Jansen and Werker (2020) to estimate the potential cost of illiquidity for the same asset classes.

In general, the literature has identified several market imperfections which can exacerbate market liquidity (Vayanos and Wang, 2012). We do not focus on the reasons for which illiquidity may be present, but it is worthwhile pointing out that each of these are present in the market for private asset classes that we consider in (2.6), thus justifying the modelling approaches that we consider. First, clientele effects exist, which allow only investors with sufficient capital or expertise to trade, especially when participation costs and transaction costs are present and agents face charges for trading and monitoring market movements. Second, imperfect competition may allow large players to exert market power and may push out smaller investors causing a strain on liquidity. Third, the presence of asymmetric information may motivate buyers to leave the market. In addition, funding constraints, such as the low access to credit for investors may exacerbate illiquidity further (Brunnermeier and Pedersen, 2009). Also, search frictions, especially in over-the-counter markets, typically create a decentralized network structure of the market where investors need to spend time looking for an appropriate trading counterparty to the trade as in Duffie et al. (2005).

We also relate to the literature on liquidity premia. In Section 2.5 we evaluate the return surcharge that a utility-maximizing investor is willing to accept (in utility-equivalent terms) in order to convert the illiquid asset to a liquid one with otherwise comparable risk-return properties. Even though we do not consider the risk premia in a general equilibrium framework, the model still provides an indication of the extra return that investors would require in order to accept holding an illiquid asset. Jansen and Werker (2020) call this *the shadow cost of illiquidity.*

Empirically, there is an academic debate on how large liquidity premia are and if they even exist given biases endemic to illiquid asset classes (Ang, 2011) and the difficulty of establishing a reliable observational measure of illiquidity even for publicly traded assets

(Goyenkoa and Trzcinkab, 2009). Our approach provides a theoretically justified link between the severity of the liquidity friction and the size of the corresponding premium.[3]

There are very few studies that examine the effect of illiquidity on risk premia and asset allocations for long-term alternative asset classes, such as hedge funds, private equity, real estate, or infrastructure investments, so we try to fill this gap. Korteweg and Westerfield (2022) provide an overview of the main challenges in terms of data transparency, performance measurement and investing in private asset classes. In terms of liquidity, Franzoni et al. (2012) find evidence that private equity shares the same liquidity factor as public equity, reducing the funds alpha and the diversification potential typically expected by investors. Jansen and Werker (2020) applies a portfolio choice model with liquidity costs for investors with a limited horizon, calibrating the model to regulatory data of alternative investment holdings. We follow a similar approach but overcome the difficulty to estimate the actual return and risk characteristics on illiquid asset classes by directly employing forward-looking expected return and risk data from JP Morgan (2022).

Finally, we extend the literature on SAA for institutional investors[4] by explicitly incorporating a liquidity friction as a latent Poisson factor in the spirit of Ang et al. (2014). The approach can be related to studies of dynamic portfolio choice with jump risk (Wu, 2003; Liu et al., 2003). The main difference with the current approach is that at the moment the Poisson jump occurs, we allow for the decision-maker to access liquidity on the private asset and to reset the portfolio allocations to optimality.

2.3 Baseline Model of Portfolio Choice

We proceed by describing the dynamic investment problem with full market liquidity.

[3]Overall, the liquidity premia literature tends to focus on market microstructure, for example within equity (Amihud and Mendelson, 1986; Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005), corporate bonds (Bongaerts et al., 2012), government bonds (Yakov Amihud and Pedersen, 2005), CDS derivatives Qiu and Yu (2012) and crypto-currency (Brauneis et al., 2021) markets among other classes. Generally, the effect of illiquidity on risk premia in these studies is assumed to work through channels that can be summarized as level, when asset-specific characteristic leading to an additional market premium, and risk, when the asset's returns are sensitive to market-wide liquidity shocks, where liquidity as risk may also affect liquid assets as well Amihud and Mendelson (2015). For example, Lou and Sadka (2005) argues that in a market crisis when liquidity dries out, portfolio managers are likely to de-lever by selling their risky liquid assets first.

[4]See for example Campbell et al. (2004); Cochrane (2022).

2.3.1 Liquid Markets

In the liquid case, we assume that the market is complete[5] and there are no arbitrage opportunities. Denoting the instantaneous risk-free rate as r, the price of a risk-free asset follows from the process:

$$dB_t/B_t = rdt \tag{2.1}$$

The n-dimensional vector containing all risky asset returns $\boldsymbol{dS}/\boldsymbol{S}$ in is defined as:

$$\begin{aligned} \frac{\boldsymbol{dS}_t}{\boldsymbol{S}_t} &= \boldsymbol{\mu} dt + \boldsymbol{\sigma} \boldsymbol{dZ}_t \\ &= (r\mathbb{1} + \boldsymbol{\sigma\lambda})dt + \boldsymbol{\sigma} \boldsymbol{dZ}_t \end{aligned} \tag{2.2}$$

where $\mathbb{1}$ stands for a column vector of ones, $\boldsymbol{dZ}_t$ is a vector of n independent Brownian motions supported by probability space $(\Omega, \Im, \mathbb{P})$; $\boldsymbol{\mu}$ is a $n \times 1$ vector of expected returns; $\boldsymbol{\sigma}$ is $n \times n$ matrix holding the sensitivity of the risky asset returns to the Brownian uncertainties; $\boldsymbol{\lambda} = \boldsymbol{\sigma}^{-1}(\boldsymbol{\mu} - r\mathbb{1})$ is the price of risk. In a market with two risky assets we have:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1, \mu_2 \end{bmatrix}^\intercal, \boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{bmatrix}, \boldsymbol{\Sigma} = \boldsymbol{\sigma\sigma}' = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{2.3}$$

2.3.2 Preferences and Wealth Dynamics

The market is populated by homogeneous infinitely-lived agents with the same investment portfolio. They do not receive any income from non-financial sources and consume out of the portfolio continuously. Their wealth dynamics are thus determined by the return of the portfolio they hold as a weighted average of the returns on all investment assets minus the consumption rate c_t:

$$\frac{dW_t}{W_t} = (r + \boldsymbol{\pi}_t'(\boldsymbol{\mu} - r\mathbb{1}) - c_t)dt + \boldsymbol{\pi}_t' \boldsymbol{\sigma} \boldsymbol{dZ}_t \tag{2.4}$$

[5] Completeness implies that any future uncertain pay-off is attainable by combining in a portfolio the assets which are available on the market. It is achievable when the number of marketable assets is the same as the number of uncertainty sources.

[5] A self-financing trading strategy starts with zero investment cost, generates a non-negative payoff with probability one, and positive payoff with positive probability.

Over time, agents control two major aspects of the wealth's evolution - the consumption (withdrawal) rate and the portfolio composition determined by the vector asset allocation weights $\boldsymbol{\pi}_t$.

Agents derive utility from consumption. It is time-separable and is assumed to be of the CRRA form:

$$u(C_s) = \frac{C_s^{1-\gamma}}{1-\gamma}$$

with $\gamma > 0, \gamma \neq 1$ as the coefficient of relative risk aversion, where higher γ implies a higher risk aversion. Over time, agents aim to maximize their expected lifetime utility by setting the optimal level of consumption and the risky asset weights in the investment portfolio:

$$\sup_{(\pi_s, C_s)} E_t \int_t^{\infty} \mathrm{e}^{-\beta(s-t)} u(C_s) ds \tag{2.5}$$

By making an optimizing decision agents consider all relevant market factors, such as the risk-free rate, the expected return of the marketable assets, their price variance, and correlations, and in the process are subject to the dynamic budget constraint defined by the wealth process (2.4).

2.3.3 The Martingale Solution

First, we use the martingale approach to determine the current value of the uncertain future consumption streams generated by the stochastic wealth processes. For the purpose, we define the Stochastic Discount Factor (SDF) as a process M_t, with the purpose of pricing future random payoffs. In particular, for a security with price P_t and instantaneous payoff X_t, the SDF allows us to write the relationship

$$P_t = E_t \left[\frac{M_s}{M_t} X_s \right] \quad s \geq t \tag{2.6}$$

In the market defined so far, the SDF can be shown[6] to evolve as follows:

$$dM_t = -M_t(rdt + \boldsymbol{\lambda}' dZ_t), \text{ with } M_0 = 1 \tag{2.7}$$

[6] See for example Campbell and Viceira (2002); Munk (2013); Duffie (2001).

The SDF uniquely defines prices if and only if the market is complete and there are no opportunities for arbitrage in the economy. In that case, a unique trading strategy can replicate future payoffs in any state of the world. As a result, Cox and Huang (1989) show that agents can transform the dynamic optimization problem of (2.5) into an equivalent static problem:

$$\begin{aligned} &\sup_{C_s} E \int_0^\infty \mathrm{e}^{-\beta s} u(C_s) ds \\ &\text{s.t. } E \int_0^\infty M_s C_s ds = W_0 M_0 \end{aligned} \tag{2.8}$$

The constraint becomes intuitive if we view the wealth process as the total return of an asset with price W_s, which continuously pays out a dividend of C_s, rewriting (2.4) as:

$$\frac{dW_s + C_s ds}{W_s} = (r + \boldsymbol{\pi}_s' \boldsymbol{\sigma} \boldsymbol{\lambda}) ds + \boldsymbol{\pi}_s' \boldsymbol{\sigma} \boldsymbol{dZ}_s$$

Then, the current market value at time $t < s$ of this asset can be found by pricing the infinite stream of discounted future cash flows spun by the consumption process, such that:

$$W_t = E_t \int_t^\infty \frac{M_s}{M_t} C_s ds$$

If wealth is optimally invested and optimally consumed over time, then at time $t = 0$ its market value spun by the SDF will equal to initial value W_0 available to the agent. In other words, in optimality, the starting wealth needs to be sufficient to finance expected future consumption.

The static problem can then be solved by the Lagrange multiplier method:

$$\sup_{(C_s)} \mathcal{L} = \sup_{(C_s)} \left\{ E \int_0^\infty \mathrm{e}^{-\beta s} u(C_s) ds + \phi \left(W_0 - E \int_0^\infty M_s C_s ds \right) \right\}$$

where ϕ stands for the Lagrangian constant. This can be solved by first finding the optimal consumption C_s^* from the first-order condition w.r.t. consumption applied on the Lagrangian function. Then the multiplier ϕ^* can be found by substituting consumption in the constraint. This will provide the optimal wealth dynamics W_t^* from which optimal portfolio asset allocations can be implied. Following this procedure, the optimal consumption and investment strategies can be derived (see Appendix 2.A.1) as

$$c_t = c^* = \frac{\beta + r(\gamma - 1)}{\gamma} + \frac{1}{2}\frac{\gamma - 1}{\gamma^2}\|\boldsymbol{\lambda}\|^2 \tag{2.9}$$

$$\boldsymbol{\pi}_t = \boldsymbol{\pi}^* = \frac{1}{\gamma}(\boldsymbol{\sigma}')^{-1}\boldsymbol{\lambda} = \frac{1}{\gamma}(\boldsymbol{\sigma}\boldsymbol{\sigma}')^{-1}(\boldsymbol{\mu} - r\mathbb{1}) \tag{2.10}$$

The landmark result observed seen here is that in the Merton world defined so far, both the investment weights and the consumption share are fixed and independent of the current wealth.

The optimal wealth process allows us to define the value function $V(W_t^*)$ as the cumulative expected utility when the optimal wealth trajectory is followed. In Appendix (2.A.1) we show that this simplifies to

$$V(W_t^*) = H \cdot (W_t^*)^{1-\gamma} \tag{2.11}$$

where $H = \frac{1}{1-\gamma}\left(\frac{1}{c^*}\right)^{\gamma}$.

2.3.4 The Dynamic Programming Solution

An alternative is to consider the problem within the context of dynamic optimization. This approach will prove itself useful when we relax the liquidity assumption in the next section, so we explore it here.

The dynamic problem that agents need to solve at any time instant is

$$\begin{aligned} &\sup_{(\boldsymbol{\pi}_s, C_s)} E_t \int_t^{\infty} \mathrm{e}^{-\beta(s-t)} u(C_s) ds \\ &\text{s.t. } dW_s = W_s(r + \boldsymbol{\pi}_s'\boldsymbol{\sigma}\boldsymbol{\lambda})ds - C_s ds + W_s\boldsymbol{\pi}_s'\boldsymbol{\sigma}\boldsymbol{dZ}_s \end{aligned} \tag{2.12}$$

Reexamining the value function over a short time period Δt we can follow the Bellman principle of optimality to arrive at a more tractable optimizing equation. In particular, we split today's value function into an optimizing decision over the upcoming time span (from t until $t + \Delta t$) and an optimal strategy afterward:

$$V(W_t) = \sup_{(\boldsymbol{\pi}_s, C_s)} \int_t^{t+\Delta t} \mathrm{e}^{-\beta(s-t)} u(C_s) ds + \mathrm{e}^{-\beta \Delta t} E[V(t + \Delta t, W_{t+\Delta t})] \tag{2.13}$$

Note that the time subscript from the expectation has been dropped, as the wealth dynamics are driven by independent random shocks, implying that knowledge of the present does not help in forecasting the expected value function, so the unconditional expectation can be used instead of the conditional. In the limit for small Δt this converges to the continuous time Bellman equation, as shown in Appendix (2.A.2):

$$\beta V(W_t) = \sup_{(\boldsymbol{\pi}_t, C_t)} u(C_t) + E\left[dV(W_t)\right] \tag{2.14}$$

where $E\left[dV(W_t)\right]$ stands for the drift term of the value process $dV(W_t)$.

An appealing intuition can be derived by rewriting the Bellman equation (2.14) in the following form[7] indicating that over any short interval of time the optimal allocation and consumption strategies have to ensure that the discounted utility exactly offsets any expected decline in the discounted value function:

$$\sup_{(\boldsymbol{\pi}_t, C_t)} \mathrm{e}^{-\beta t} u(C_t) + E\left[d(\mathrm{e}^{-\beta t} V(W_t))\right] dt = 0 \tag{2.15}$$

Taking advantage of Ito's calculus, we show in Appendix 2.A.2 that the Bellman principle gives rise to the following partial differential equation that the value function, and the optimal strategies for consumption C_t and portfolio weight allocations $\boldsymbol{\pi}_t$ have to satisfy (known as the Hamilton-Jacobi-Bellman (HJB) equation):

$$\begin{aligned} &\mathcal{L}^C + \mathcal{L}^\pi - \beta V = 0 \\ &\quad \mathcal{L}^C = \sup_{C_t} \left\{ u(C_t) - C_t V_W \right\} \\ &\quad \mathcal{L}^\pi = \sup_{\boldsymbol{\pi}_t} \left\{ (r + \boldsymbol{\pi}_t'(\mu - r\mathbb{1}))\, V_W W_t + \frac{1}{2} V_{WW} W_t^2 \boldsymbol{\pi}_t' \boldsymbol{\Sigma} \boldsymbol{\pi}_t \right\} \end{aligned} \tag{2.16}$$

[7] Follows by reversing the terms in $d(\mathrm{e}^{-\beta t} V(W_t)) = -\beta \mathrm{e}^{-\beta t} V(W_t) + \mathrm{e}^{-\beta t} dV(W_t)$ and substituting in Equation (2.14)

and where we define $V \equiv V(W_t)$, $V_W \equiv \frac{\partial V(W_t)}{\partial W_t}$ and $V_{WW} \equiv \frac{\partial^2 V(W_t)}{\partial W_t^2}$. This allows us to derive the Envelope Theorem, implying that at optimality, the marginal utility from consuming a little more needs to be equal to the marginal value of investing a little more:

$$u'(C_t) = V_W \tag{2.17}$$

As a result, optimal consumption can be determined

$$C_t^* = I_{u'}(V_W) \tag{2.18}$$

where $I_{u'}(.)$ denotes the inverse of the marginal utility function. For CRRA utility we have $C_t^* = (V_W)^{-\frac{1}{\gamma}}$.

The first order condition with respect to the optimal assets allocation then implies:

$$\boldsymbol{\pi}^* = -\frac{V_W}{W_t V_{WW}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbb{1}) = -\frac{V_W}{W_t V_{WW}} (\boldsymbol{\sigma}')^{-1} \boldsymbol{\lambda} \tag{2.19}$$

In total the agent will invest the fraction $-\frac{V_W}{W V_{WW}} \mathbb{1}'(\boldsymbol{\sigma}')^{-1}\boldsymbol{\lambda}$ in risky assets and the rest in the risk-free asset. The fraction $-\frac{V_W}{W V_{WW}}$ defines the relative risk tolerance of individuals given their indirect utility $V(W)$ and is the inverse of the Relative Risk Aversion measure constructed as $-\frac{u''(C)C}{u'C}$.

The optimal risky asset investment is thus determined by two components: (1) the preference-specific relative risk tolerance V_W/WV_{WW} and (2) a vector, which is the same for everyone on the market and which is composed of the inverse of the assets' variance-covariance matrix and the excess expected returns: $(\boldsymbol{\sigma}')^{-1}\boldsymbol{\lambda}$.

Note that the second-order conditions for maxima are satisfied both for (2.17) and (2.19) as long as $V(W)$ is concave in W and $u(C)$ is concave in C. Substituting the maximizing allocation and consumption terms in the HJB equation of (2.16) and applying explicitly the assumed CRRA utility form simplifies further to the equation:

$$\beta V = \frac{\gamma}{1-\gamma} V_W^{1-1/\gamma} + r W_t V_W - \frac{1}{2} \frac{V_W^2}{V_{WW}} \|\boldsymbol{\lambda}\|^2 \tag{2.20}$$

This second order Partial Differential Equation (PDE) can be solved by making a guess for the value function of the form $V(W_t, t) = g(t)^{\gamma} \frac{W_t^{1-\gamma}}{1-\gamma}$. Substituting in (2.20) verifies that the guess solves the PDE for $g(t) = \frac{1}{A}$, where A is the same constant we found in (2.42). The optimal consumption share is again $c^* = \frac{C_t}{W_t} = A$. Note that $g(t)$ is a constant function in the case of an infinite horizon problem.

2.4 Baseline Model with an Illiquid Asset

When the assumption of a complete market does not hold, we need to resort to the DP approaches outlined earlier to solve the agents' optimization problem.[8]The presence of illiquidity violates the market completeness property, since in periods of illiquidity certain market outcomes will not be reachable. In a way, illiquidity risk presents a new source of uncertainty that cannot be managed and hedged within the portfolio.

Asset prices are again defined through relation (2.2). The additional component now is that the investor cannot be certain that a market for n-th asset with price S_n will exist over time. Following Ang et al. (2014), we formalize this by assuming that a trading opportunity for the asset arrives with the intensity η of a Poisson process N_t, which is independent of the Brownian motions.[9]A jump in N_t indicates that a liquidity opportunity for the asset has materialized, the investor can trade without further costs or frictions and can freely rebalance the portfolio to her strategic objectives. We assume that the illiquid asset cannot be used as collateral to issue riskless debt.[10]

This changes the structure of the problem and the agents' wealth dynamics. Investors can only consume directly out of liquid wealth W_t and can transfer funds dI_t from illiquid

[8]An alternative is to adjust the martingale method. For more details see for example Karatzas and Shreve (1998). We do not explore this option here.

[10]The assumption ensures that the illiquidity constraint cannot be circumvented by pledging the illiquid asset and issuing debt. Still, the assumption is empirically valid, as the private assets we consider later on are rarely accepted as collateral.

[10]The counting process $\{N_t, t \in [0, \infty)\}$ is a Poisson process with rate $\eta > 0$ if: (1) The starting value of the process is zero: $N_0 = 0$ (2) Increments of the process are stationary and independent (3) The number of events that occur during any time increment of length Δt is Poisson distributed with mean $\eta \Delta t$. Formally, for any $t \geq 0, \Delta t > 0$:

$$P(N_{t+\Delta t} - N_t = n) = \mathrm{e}^{-\eta \Delta t} \frac{(\eta \Delta t)^n}{n!}, \quad n = 0, 1, ...$$

See Appendix (2.B.2) for more details.

wealth X_t only when an appropriate opportunity arises. Assume that $\boldsymbol{\theta}_t$ is an $n-1$ vector holding the fractions of liquid wealth that are invested into the liquid risky assets, and c_t is the fraction of liquid wealth consumed, then the wealth dynamics will evolve as follows:

$$\begin{aligned} dW_t/W_t &= (r + \boldsymbol{\theta}_t'(\mu_{1:n-1} - r\mathbb{1}_{1:n-1}) - c_t)dt + \boldsymbol{\theta}_t'\boldsymbol{\sigma}_{n-1}\boldsymbol{dZ}_t - dI_t/W_t \\ dX_t/X_t &= \mu_n dt + \boldsymbol{\sigma}_n \boldsymbol{dZ}_t + dI_t/X_t \end{aligned} \tag{2.21}$$

where the subscripts $1:n-1$ and n indicate the first $n-1$ rows, or respectively the n-th row of the corresponding matrix or vector, specified in (2.2).

The value function for this problem can be written as

$$V(W_t, X_t) = \sup_{\theta_s, dI_s, c_s} E_t \int_t^\infty \mathrm{e}^{-\beta(s-t)} u(C_s) ds \tag{2.22}$$

We denote $Q_t = W_t + X_t$ as total liquid and illiquid wealth, and ξ_t as the proportion of illiquid asset holdings to total wealth.[11] For a CRRA utility, we can define the reduced-form value function $H(\xi) \equiv V((1-\xi), \xi)$. From the homogeneity properties of the value function, it follows that:

$$V(W_t, X_t) = (Q_t)^{1-\gamma} H(\xi_t) \tag{2.23}$$

As a result, we can transform the optimization problem of (2.22) from one of finding the value function $V(W_t, X_t)$ itself to solving for the the concave function $H(\xi_t)$.

Whenever liquidity is available, when the Poisson shock hits, agents are free to move to the top of the $H(\xi_t)$ curve.[12] They do so by rebalancing between liquid and illiquid wealth, such that a fixed portion ξ^* is held in the illiquid asset. As a result, they set $\xi^* = arg \max_\xi H(\xi)$ by choosing the appropriate amount of $dI_\tau = (\xi^* - \xi_{\tau^-})W_{\tau^-}$ where ξ_{τ^-} is the initial weight before rebalancing occurs and ξ^* is the target weight that can be reached immediately after rebalancing. Whenever the asset stays illiquid, the agent is stuck with sub-optimal ξ_t and is below the optimum of the $H(\xi_t)$. Once we find the

[11] We diverge here from the notation of the previous section. To reconcile the two, note (suppressing the time notation) that $\boldsymbol{\pi}_{1:n-1} = \boldsymbol{\theta}(1-\xi)$ is the allocation to the first (liquid) risky asset as a proportion of total wealth, $\boldsymbol{\pi}_n = \xi$ is the allocation to the illiquid asset, and $c(1-\xi)$ is now the consumption rate out of total rate, where in the illiquid case, c is consumption out of liquid wealth.

[12] Ang et al. (2014) show analytically that $H(\xi)$ is finite, continuous, concave and is maximized at ξ^* for $\xi \in [0, 1)$.

function $H(\xi)$ then, we can determine both the optimal mix between liquid and illiquid assets ξ^*, and the value function $V(W_t, X_t)$ itself.

Going forward, we will interpret ξ^* and $\boldsymbol{\theta}(\xi^*)$ as the strategic allocation mix. The reaction function $\boldsymbol{\theta}(\xi)$ when trading opportunity in the illiquid asset is not available, will form the tactical allocation as a reaction to the illiquid wealth share in the portfolio.

2.4.1 The HJB Equation with Three Assets

Assuming that there are three assets: liquid risk-free, liquid risky and illiquid risky, and following the form established in (2.3), the wealth dynamics will evolve as

$$\begin{aligned} dW_t/W_t &= (r + (\mu_1 - r)\theta_t - c_t)dt + \theta_t\sigma_1 dZ_{1t} - dI_t/W_t \\ dX_t/X_t &= \mu_2 dt + \sigma_2\rho dZ_{1t} + \sigma_2\sqrt{1-\rho^2}dZ_{2t} + dI_t/X_t \end{aligned}$$

We can then derive the HJB equation governing optimal consumption and allocation as functions of the illiquid wealth share[13]:

$$\begin{aligned} &\mathcal{L}^C + \mathcal{L}^\theta + \mathcal{L} = \beta V \\ &\qquad \mathcal{L}^C = \sup_{C_t}\left\{u(C_t) - C_t V_W\right\} \\ &\qquad \mathcal{L}^\theta = \sup_{\theta_t}\left\{(r + \theta_t(\mu_1 - r\mathbb{1}))\, V_W W_t + \frac{1}{2}V_{WW}W_t^2\theta_t^2\sigma_1^2 + V_{WX}W_t X_t\theta\sigma_2\sigma_1\rho\right\} \\ &\qquad \mathcal{L} = V_X\mu_2 + \frac{1}{2}V_{XX}X^2 + \eta(V^* - V) \end{aligned} \tag{2.24}$$

with V^* the point at which the value function jumps once the Poisson shock hits and the agent changes the illiquid-to-total wealth ratio back to the strategic ξ^*.[14]

By construction, consumption in the periods between the Poisson liquidity arrivals can be funded from liquid wealth only, so by maximizing $\mathcal{L}^C$ we arrive at a modified version of

[13] See Appendix (2.A.4) for details.

[14] We can also establish bounds for the $H(\xi)$ function. Agents will invest in the illiquid asset only if that leads to higher cumulative utility than the one-asset Merton case. At the same time, with a liquidity friction agents will not be able to reach the two-asset Merton case, where liquidity is fully available. This leads to the value function of (2.22) being bounded from above and below:

$$\begin{aligned} H_{M1}W^{1-\gamma} \le V(W, X) \le H_{M2}(W + X)^{1-\gamma} \le 0 \\ \implies H_{M1}\,(1-\xi)^{1-\gamma} \le H(\xi) \le H_{M2} \le 0 \end{aligned}$$

where the constants H_{M1} and H_{M2} are the one and two asset value function constants from (2.11).

the Envelope Theorem (2.17). At optimum, the marginal utility of liquid wealth is equal to the value of a marginal change in liquid wealth, such that

$$\begin{aligned} & u'(C_t) = V_W(W_t, X_t) \\ \Longleftrightarrow \; & C_t^* = I_{u'}\left(V_W(W_t, X_t)\right) \equiv \left(V_W(W_t, X_t)\ \right)^{-\frac{1}{\gamma}} \end{aligned}$$

In the CRRA case in particular, $V_w(W_t, X_t) = W_t^{-\gamma}\left((1-\gamma)H(\xi_t) - \xi_t H'(\xi_t)\right)$ and $u'(c_t W_t) = u'((1-\xi_t)W_t) = (1-\gamma)(cW)^{-\gamma}$. As a result total wealth can be canceled out and the optimal consumption rate (out of total wealth) c_t can be expressed as

$$c_t = \Big((1-\gamma)H(\xi_t) - H'(\xi_t)\xi_t\Big)^{-\frac{1}{\gamma}}(1-\xi_t)^{-1} \tag{2.25}$$

In contrast to the fixed Merton optimal consumption rate in (2.9), the consumption rate between trading events now is time-varying with ξ_t. As the share of illiquid wealth floats randomly in periods of illiquidity, the optimal consumption rate will change as well.

The optimal investment in the risky liquid asset can be derived correspondingly:

$$\theta_t = \frac{\mu_1 - r}{\sigma_1^2}\underbrace{\left(-\frac{V_W}{V_{WW}W_t}\right)}_{\Phi(\xi_t)} + \frac{\sigma_2\rho}{\sigma_1}\underbrace{\left(-\frac{V_{WX}X_t}{V_{WW}W_t}\right)}_{\Psi(\xi_t)} \tag{2.26}$$

Note that we can split the optimal investment in the liquid risky asset into two parts. First, the *investment demand* in the liquid risky asset is driven by

$$\frac{\mu_1 - r}{\sigma_1^2}\Phi(\xi_t)$$

where it can be seen that the function $\Phi(\xi_t) = -\frac{V_W}{V_{WW}W_t}\frac{1}{1-\xi_t}Q_t$ is decreasing in ξ_t as $-\frac{V_W}{V_{WW}} \geq 0$. This implies that, even with zero correlation between the two risky assets, the optimal investment allocation in the risky liquid asset will still be state-dependent, and decreasing in the share of current illiquid asset holdings.

Second, the *hedging demand* term is represented by

$$\frac{\sigma_2\rho}{\sigma_1}\Psi(\xi_t)$$

where $\Psi(\xi_t) = -\frac{V_{WX}}{V_{WW}}\left(\frac{\xi_t}{1-\xi_t}\right)$ again can be shown to be decreasing in $\xi_t \in (0,1)$, as due to the convexity of the value function we have $-\frac{V_{WX}}{V_{WW}} \leq 0$. If the correlation between the two assets then is negative, increases in the share of the illiquid asset will increase the hedging demand for the liquid asset, counterbalancing the investment demand effect. For positive correlation, however, the hedging and the investment demand will be going in the same direction, both decreasing the share of liquid risky investments when the share of illiquid wealth increases.

2.4.2 The Certainty Equivalent and Utility Loss

Apart from determining the optimal investment and consumption strategies, we want to quantify the utility-equivalent cost of illiquidity. For this purpose, we define the CEC, as the guaranteed continuous consumption stream which makes agents indifferent between risk-free and uncertain consumption. It is determined by equating the cumulative utility associated with risk-free consumption to the indirect utility associated with a particular risky consumption stream C_s:

$$\int_t^{\infty} \mathrm{e}^{-\beta s} u(CES_t) ds = E_t \int_t^{\infty} \mathrm{e}^{-\beta s} u(C_s) ds \tag{2.27}$$

The right-hand side of the equation corresponds to the value function derived from the optimization problem the agent faces. The left-hand-side of this equation can be simplified to

$$u(CEC_t) \int_t^{\infty} \mathrm{e}^{-\beta s} ds = \frac{1}{\beta} u(CEC_t)$$

which implies that the CEC is

$$CEC_t = I_u \left(\beta E_t \int_t^{\infty} \mathrm{e}^{-\beta s} u(C_s) ds \right) \tag{2.28}$$

Expressing the CEC as a fraction of total wealth, we can relate it to the reduced cumulative utility functions ((2.11) in the case of full liquidity, and (2.23) in the case of illiquidity). Thus, dividing both sides of (2.28) by Q_t, we get

$$\begin{aligned}\frac{CEC_t^M}{Q_t} &= (\beta(1-\gamma)H_M)^{\frac{1}{1-\gamma}} \\ &= \beta^{\frac{\gamma}{1-\gamma}} \left(\frac{1}{c_M}\right)^{\frac{1}{1-\gamma}}\end{aligned}$$

In the illiquid case, it will be dependent on the current share of illiquid wealth

$$\frac{CEC_t^{il}(\xi)}{Q_t} = (\beta(1-\gamma)H(\xi_t))^{\frac{1}{1-\gamma}} \tag{2.29}$$

As a result, we define the cost of liquidity $L(\xi)$ as the CEC loss associated with a lack of liquidity. It is measured as the percentage of CE consumption the agent would be willing to give up in order to make the second risky asset completely liquid. In particular we have

$$L(\xi_t) \equiv 1 - \frac{CEC_t^{il}(\xi)}{CEC_t^{M2}} = 1 - \left(\frac{H(\xi_t)}{H_{M2}}\right)^{\frac{1}{1-\gamma}} \tag{2.30}$$

where CEC_t^{M2} is determined for the Merton two-asset case. Note that the utility loss is again state-dependent through ξ_t.

It can be shown that $L(\xi_t)$ is also the percentage loss on current total wealth that individuals are willing to take in order to make their wealth fully liquid. For the individual to be indifferent between holding fully liquid or holding illiquid wealth, we need the value function in the two-asset Merton case as a function of total wealth minus the acceptable loss to be equal to the value function in the illiquid case:

$$\begin{aligned}&V^{M2}(Q_t(1-L(\xi_t))) = V(W_t, X_t) \\ &H_{M2}Q_t^{1-\gamma}(1-L(\xi_t))^{1-\gamma} = H(\xi_t)Q^{1-\gamma} \\ &L(\xi_t) = 1 - \left(\frac{H(\xi_t)}{H_{M2}}\right)^{1-\gamma}\end{aligned}$$

The last term is precisely the expression in (2.30).

Table 2.1: Parameters

Parameter	Definition	Default Value
r	risk-free rate	0.02
β	personal discount rate	0.03
μ_1	first (liquid) risky asset expected return	0.055
σ_1	first (liquid) risky asset volatility	0.14
μ_2	second (illiquid) risky asset expected return	0.055
σ_2	second (illiquid) risky asset volatility	0.14
ρ	correlation	0
$1/\eta$	average time (in years) between a trading opportunity arises	10
γ	risk aversion parameter in the CRRA utility function	6

Note. The table shows a short summary of the parameter values that are used for the presented model. Unless stated otherwise, these parameters are used for the numerical work. The return and risk of the assets are based on annual projections for U.S. public equity and cash in JP Morgan (2022)

2.5 Numerical Evaluation of Illiquidity

In this section we investigate the properties of the illiquid asset and the effect of illiquidity of premia and allocations where the continous trading case is used as a benchmark. We assume the same means and variances for the liquid and illiquid risky assets. To calibrate the model, we look at investment data projections for a number of private asset classes with substantial illiquidity.[15]

We use a risk-aversion coefficient $\gamma = 6$, as in the Merton two-asset case, under the given initial parametrization, this produces roughly 60% investment in risky assets and 40% in risk-free bonds which is a standard long-term investment strategy for moderately risk-averse agents.

The input variables used in the numerical procedures and their default values are given in Table 2.1. Unless stated otherwise, those are the base case parameter values used for

[15]Note that we make sure that a number of parameter restrictions are satisfied in setting up the parametrization of the models, as posed in Ang et al. (2014). First, the illiquid asset has to have at least as large a Sharpe ratio as that of the liquid asset:

$$\frac{\mu_2 - r}{\sigma_2} \geq \frac{\mu_1 - r}{\sigma_1}$$

Second, a standard discount rate restriction (see (Back, 2010)) holds:

$$\beta > (1-\rho)\left(r + \frac{k^2}{2\rho}\right)$$

where $k = \sqrt{\boldsymbol{\lambda}'\boldsymbol{\lambda}} = (\boldsymbol{\mu} - r\mathbb{1})\Sigma^{-1}(\boldsymbol{\mu} - r\mathbb{1})$. The first assumption is needed to discourage shorting the illiquid asset, and the second - to ensure existence of the HJB equation as stated in (2.16).

the current section. We use five-year ahead capital market projections developed by JP Morgan (2022).

We evaluate the chance at least once over a year to meet a counterparty willing to trade in the illiquid asset. This underlies the calibration of trading friction, which for a Poisson process through the relationship is (Appendix (2.B.2)):

$$p = 1 - \mathrm{e}^{-\eta \Delta t} \tag{2.31}$$

The model is solved numerically through value function iteration on the discretized Bellman equation corresponding to the continuous-time problem as outlined in Appendix (2.B.1).

Table 2.2: Sensitivity of Cost and Decision Variables to the Average Trading Interval

		Cost		Consumption		Allocation				Rabalancing	
η	p	$rp(\xi^*)$	$LC(\xi^*)$	$CEC(\xi^*)$	$E[c(\xi)(1-\xi)]$	$\theta(\xi^*)(1-\xi^*)$	$E[\theta(\xi)]$	ξ^*	$E[\xi]$	$E\left[dI^+/Q\right]$	$E\left[dI^-/Q\right]$
M1	-			2.53	2.60	29.76		29.76			
M2	-			3.04	3.03	29.76		29.76			
1/12	100.00	0.01	0.54	3.03	3.03	29.35	29.35	29.35	29.40	0.83	0.96
1/4	98.17	0.01	0.54	3.03	3.03	29.35	29.34	29.33	29.58	1.09	1.41
1/2	86.47	0.05	0.57	3.02	3.03	29.34	29.33	29.18	29.66	1.39	2.18
1	63.21	0.06	0.66	3.02	3.03	29.33	29.29	28.72	29.89	1.66	2.86
2	39.35	0.08	0.88	3.01	3.02	29.29	29.21	27.71	29.29	2.02	4.27
5	18.13	0.28	2.33	2.97	2.91	29.19	26.98	22.58	30.41	2.38	10.62
10	9.52	1.21	6.43	2.85	2.79	29.20	26.95	14.06	21.60	2.11	11.61

Note. This table shows the cost of liquidity and the sensitivity of optimal consumption and allocation to the strength of the liquidity uncertainty. The liquidity friction is quantified by the first two columns - trading interval and probability of being able to trade over the coming year. The first two lines stand for the continuous-trading Merton case with one and two assets (M1 and M2 respectively). All numbers are in percentage points.

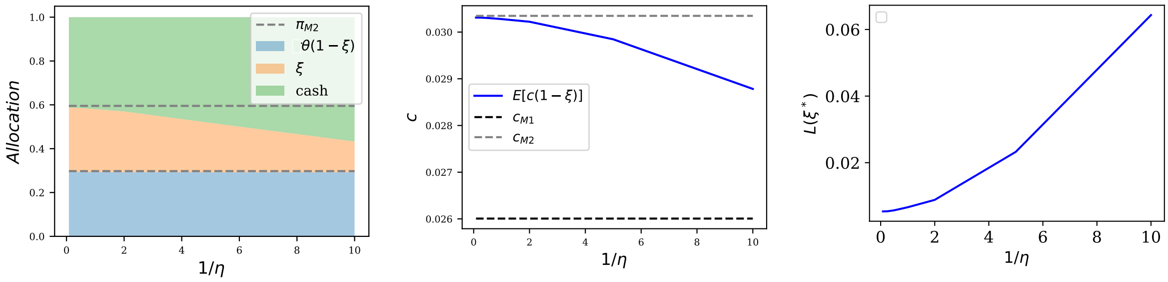
2.5.1 Optimal Allocations

First, we examine how the optimal portfolio allocation is affected by the severity of the liquidity friction. Table 2.2 presents the effects of different average waiting times to trade ($1/\eta$) on the SAA. The connection between the average waiting time and the trading probability over a year occurs through the Poisson process in equation (2.31).

As indicated in Table 2.2, at the full rebalancing point the allocation in the second asset when trading can occur up to a year on average point is still close to the continuous

Figure 2.1: Holdings and Cost Sensitivity to the Trading Friction

(a) (b) (c)



Note. This figure shows the effect of the trading friction (a) on the asset allocation as a function of the average waiting time to trade, (b) on the illiquid asset holdings as a function of the probability of being able to trade over the coming year, the dashed line indicating the Merton two-asset solution. Part (c) shows certainty equivalent consumption loss as percent to total wealth. The grey dashed lines stands for the two-asset continuous trading case, the black dashed line for the one-asset continuous trading case.

trading case, around 29.76%. As ξ^* is significantly reduced, however, the trading friction intensifies – up to 14.06% allocation for a 10-year average waiting time. Figure 2.1a also illustrates the point – the optimal base allocation to the illiquid asset starts decreasing with a friction above one year and the cash cushion starts increasing.

How fast rebalancing to base allocations can be expected to happen will depend on the average frequency with which one can trade in the illiquid asset. The propensity of the assets to drift away from the optimal allocations will then guide the size of the allocation correction agents will resort to when the opportunity arises. As a result, there is a discrepancy between the base allocations ξ^*, and the average realized weights $E[\xi]$ that is larger, the more severe the liquidity friction. Occasionally, when the Poisson liquidity shock hits, agents will be able to rebalance back to the base weights (ξ^* and $\theta(\xi^*)$) and avoid getting stuck with extreme levels of illiquid wealth. The rest of the time, they will be forced to hold sub-optimal allocations.

From that point of view, the average realized allocation moves down less sharply, and the gap between base and average realized allocation becomes larger as the friction increases. This can be interpreted as agents consistently targeting lower illiquid wealth shares. As the premium on illiquid wealth is positive, the share in illiquid wealth then tends to grow on average over time while agents are not able to withdraw from it. It is not surprising then that the average withdrawal rates from illiquid wealth $E\left[dI^-/Q\right]$ become larger, the longer is the expected waiting time to trade, while the average investment rate $E\left[dI^-/Q\right]$ is not significantly affected.

During periods of illiquidity, the agents can only adjust consumption and investment in the first (liquid) asset as a response to the stochastic share of illiquid wealth. This motivates Figures (2.2) where the (reduced) value function, consumption, and the liquid risky asset allocation are represented as response functions to the current share of illiquid wealth.

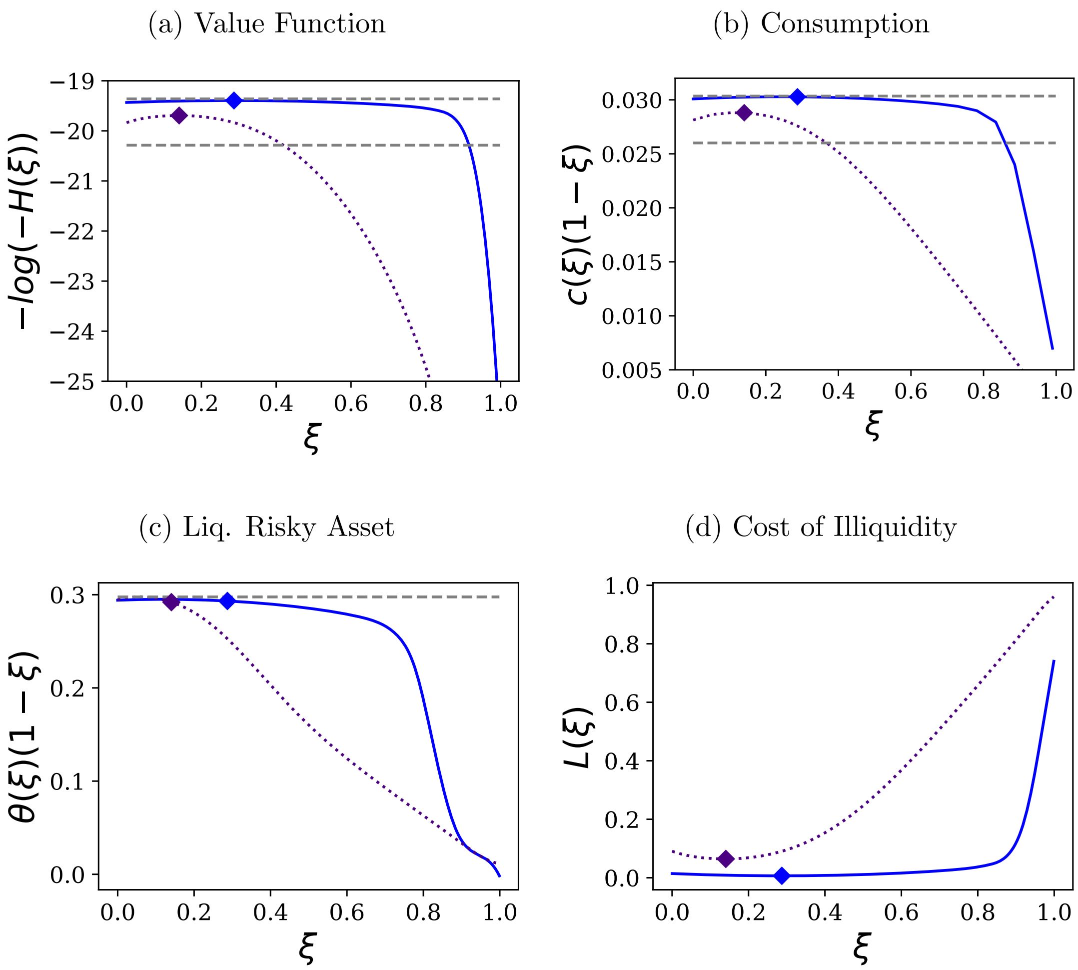
Figure 2.2c illustrates the allocation curves to the liquid asset for a one-year and ten-year friction. The Merton liquid case is presented for comparison by the two flat dashed lines (one- and two-asset market solutions respectively). In the Merton case agents can continuously rebalance to optimality, and allocations in one asset are insensitive to the current endowment in other assets.

In the one-year illiquid case, the base illiquid wealth allocation ξ^* stands close to the two-asset case liquid solution of 29.8% indicated by the diamonds on the graphs. The value function also stays close to that of the two-asset cases whenever illiquid wealth is below 80% of total wealth. Below that threshold, all one-year curves are flat, and the agent can finance more or less the same consumption rate as in the continuous trading case at any level of ξ. In each chart, a gap between the Merton flat lines and the illiquid solutions forms sharply as illiquidity grows beyond that point and as the chance grows for the agent to end up in scenarios where liquid wealth is too low to finance consumption. In particular, agents need to prevent scenarios in which liquid wealth is exhausted before the next trading opportunity arises, so consumption and liquid risky investments are getting reduced preventively. Liquid asset investments are reduced sharply at the upper edge of the curve, which curtails the volatility of liquid wealth. In the case of a ten-year friction the same effect occurs much sooner on the ξ dimension. At around $\xi = 20\%$ the reaction curves of consumption and liquid investment start decreasing linearly.

2.5.2 Consumption and Welfare Losses

Table 2.2 shows the CEC for agents as a function of the liquidity friction. This serves as an indication of the risk-free consumption equivalent to the consumption which can

Figure 2.2: Consumption and Allocation Responses to Illiquid Asset Holdings



Note. This set of figures displays consumption and liquid asset holdings as a reaction to the illiquid asset allocation level. In each chart, we presents a one-year (solid curve), ten-year friction (dotted curve) and the continuous trading cases denoted by the flat dashed lines. Panel 2.2a shows the reduced form of the value function $H(\xi)$, compared to a Merton liquid market with one asset (bottom dashed line) and with two assets (top dashed line). Panel 2.2b shows optimal consumption. Panel 2.2c displays optimal liquid asset holdings. Panel 2.2d illustrates the certainty equivalent cost associated with a one-year and a ten-year friction. If a trading opportunity arises, the agent returns to the point indicated by a blue diamond in each of the charts. In the Merton cases, rebalancing to optimality is possible all the time, so the decision curves are insensitive to the illiquid asset holdings - the agent always can rebalance to achieve optimality.

be financed by the risky portfolio. The reduction in CEC when the liquidity friction is strengthened is also an indication of the welfare loss due to illiquidity relative to the completely liquid case.

The CEC is reduced from 3.04 in the liquid two-asset case to 2.85 (per 100 total wealth) in the illiquid case with a 10-year friction. Still, it stays above the certainty equivalent of 2.53 for the one-asset case. Correspondingly, $L(\xi^*)$ indicates that the agents will be willing to give up 6.43% of their current wealth (Figure 2.1c) in order to make fully liquid an asset with a ten-year waiting time.

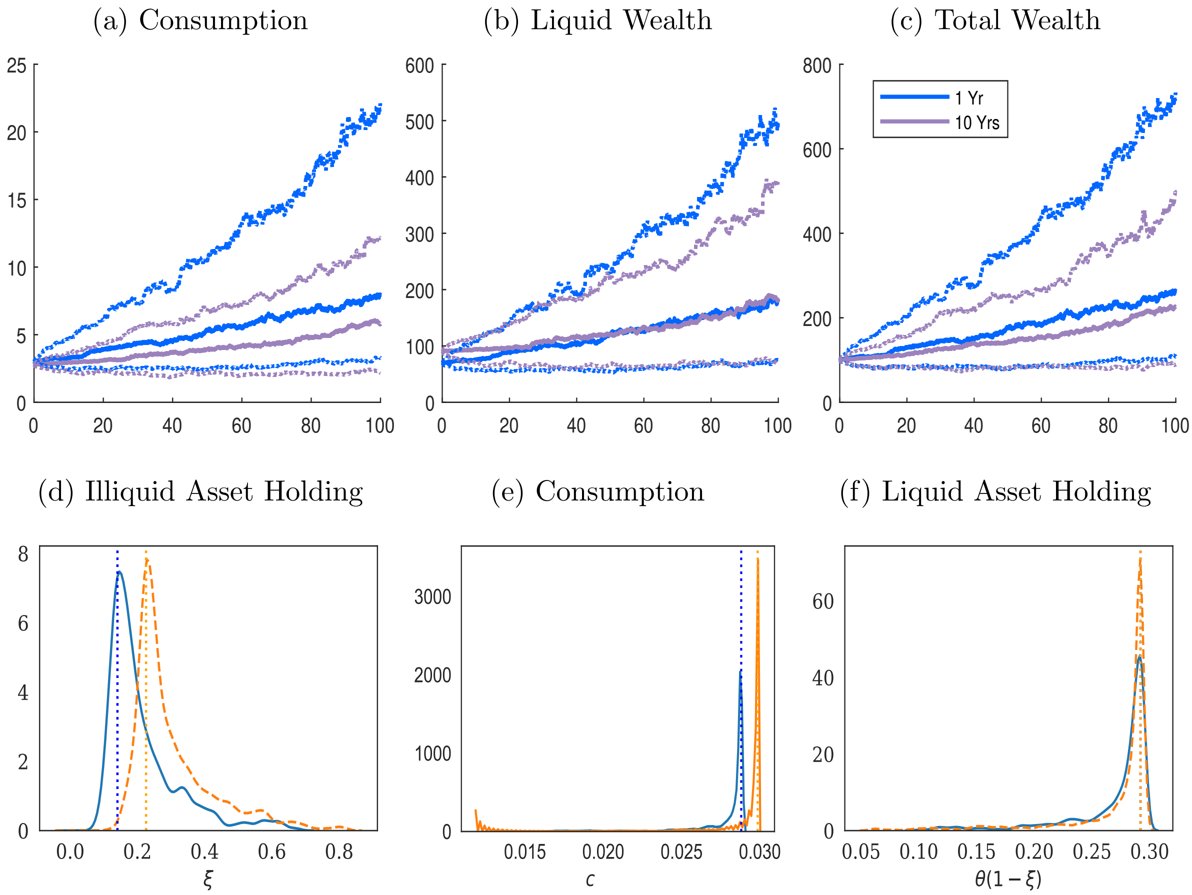
Note that here we evaluate the CEC loss only at the optimal wealth rebalancing point for $\xi_t = \xi^*$. Figure 2.2d illustrates that this is the lower limit on the losses that can be expected. For a one-year waiting time (the solid line), the loss curve is very flat for most of the range but rises sharply beyond a certain threshold, after which the agents find it difficult to finance consumption. For the ten-year waiting time (dashed curve) the loss rises notably even for moderate deviations of ξ_t from ξ^*.

Overall, the welfare loss from the illiquid and the continuous-trading case can be attributed to two factors. First, higher liquidity friction implies a higher constraint on the ability of agents to finance their consumption, when liquid holdings decrease too much. Second, the lack of trading opportunity leads to lower diversification, as *ex-ante* the holdings in the illiquid asset are reduced. The longer the expected waiting time until trading can happen, the higher each of these two components to the welfare cost will be.

The illiquidity cost can also be translated into a liquidity premium – the spread over the illiquid asset's expected return in order to make it just as attractive as a liquid asset with otherwise the same risk characteristics. The column α_{lp} in Table 2.2 makes this point. For an asset that can be traded semi-annually on average, the liquidity premium is about $5bps$. The premium grows to above 1.21% for assets which can be expected to be traded ten years from now. This premium is again computed at the high point of the value curve and presents a lower bound on the possible observed losses.

The cost that illiquidity posts on long-term investors is also shown in Figure 2.3 where we simulate for the two assets a number of time paths of length 100 years and allow the investor to follow the optimal strategies discussed so far. It can be seen that consumption and wealth grow at a lower rate when the friction is higher. More severe illiquidity induces

Figure 2.3: Simulation Paths with Illiquidity



This set of charts shows a model simulation. Charts (a)-(c) show the simulated paths of consumption, liquid, and total wealth over 100 years with monthly discretization of the stochastic processes. We show the median levels at every point in time as well as a 90% confidence intervals. The blue lines are based on a one-year friction; the light purple line is based on a ten-year friction. The model is parameterized for a 10 years trading friction. Charts (d)-(f) show the corresponding density of the decision variables in the model. All fractions are expressed as a percentage of total wealth. Consumption and liquid wealth allocation vary with the level of illiquid asset holdings ξ. The dotted vertical line in each case shows the optimal values at the base allocation point when liquidity is available. The orange line here is for the one-year friction, the blue line is for the ten-year one.

more conservative investment, which can also be traced back in the optimal composition of the portfolio in Figure 2.1a, stating that as the friction intensifies, the investor holds a higher proportion of the portfolio into cash as a cushion on the wealth volatility and illiquidity risk. A lower share of risky assets, in turn, hampers the investor's ability to build up wealth over time and consumption. We can see that in the case of the ten-year friction (purple), consumption grows at a slower pace compared to the case with a more liquid asset (blue).

Figure 2.1a also provides the simulated densities of the decision variables of the model. In (2.3d), the illiquid asset allocation exhibits a fat right tail and a tendency on average to float above the target level ξ^* indicated by the vertical dashed line. The expected return of the two risky assets is the same, still, the continuous consumption withdrawal from liquid wealth reduces its expected return below that of the expected return of illiquid wealth, and as a result, the share ξ tends to grow on average. The consumption share and the illiquid asset share both exhibit fat left tails as a reaction to the fact that ξ tends to drift above optimal levels, which often requires consumption and risky allocations to be reduced.

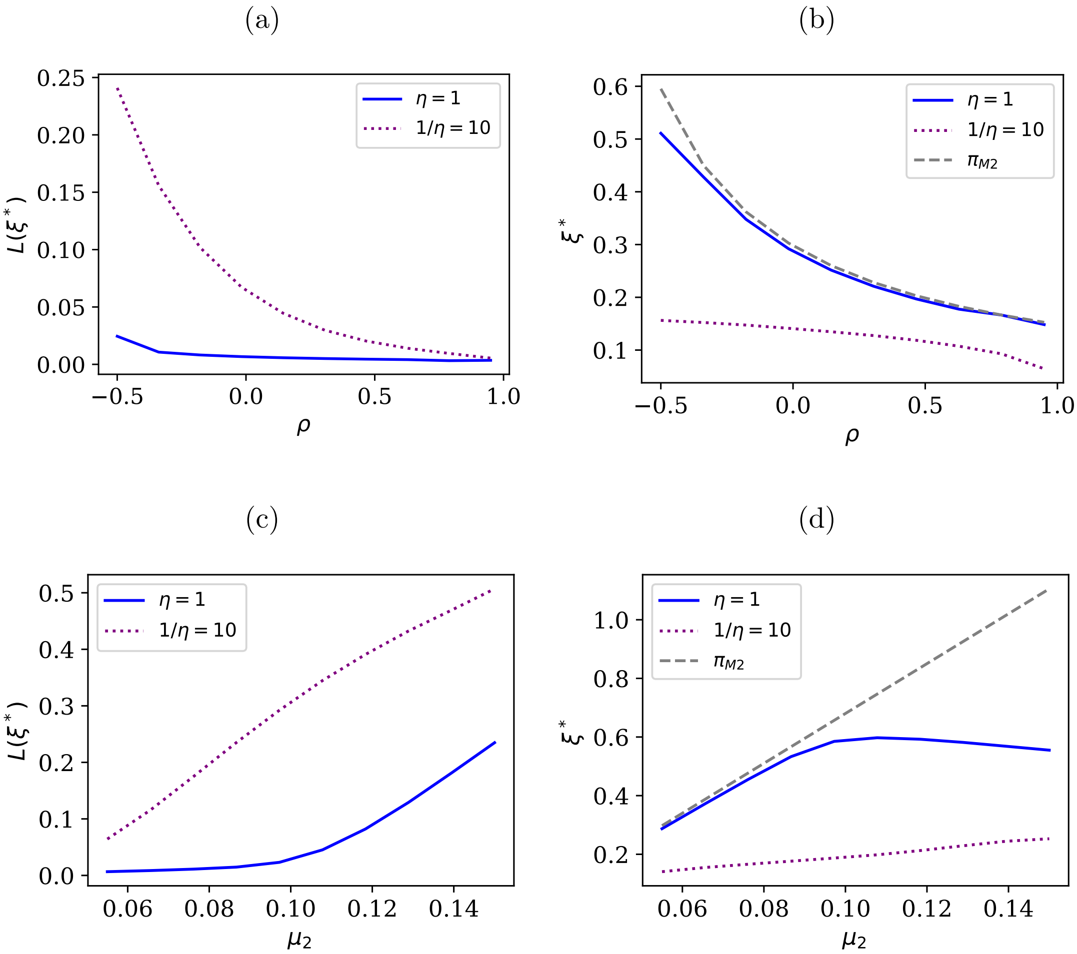
2.5.3 Parameter Sensitivity

Figure 2.4 explores how the liquidity cost and base allocations to the illiquid risky assets change as the correlation and respectively the expected return in the illiquid asset increases.

We can see in chart (2.4a) that for high expected waiting time between trades (the dotted line represents a ten-year friction) the liquidity cost is larger for negative and low correlation levels between the two risky assets. If the two assets were fully liquid, with low correlation, agents could continuously adjust allocations to hedge risk in one of the assets with risk in the other asset. The higher the waiting time between the assets, however, the lower the ability of agents to do so.

The higher the expected return of the second risky asset, the wider the gap between its allocation in the completely liquid case and in the illiquid case (Figure 2.4d). With higher expected return on the illiquid asset, illiquid wealth tends to grow faster than liquid

Figure 2.4: Holdings and Cost Sensitivity to Return and Correlation



This figure shows the effect varying correlation and the expected return on the second asset on the liquidity cost and optimal allocation of the portfolio.

wealth, so the agent will prefer to transfer larger amounts of wealth when the opportunity arises, and will set the base illiquid allocation to a lower rate.

2.6 Strategic Allocation with Private Asset Classes

In this section, we relate the two dynamic portfolio choice models presented so far to the SAA process for long-term investors. The optimal portfolio weights at the point where rebalancing is possible will present the base weights to which an investor will be aiming to return whenever the trading opportunity arises. Conceptually this corresponds to the SAA of a long-term investor.

First, we evaluate the optimal allocation when only traditional asset classes are part of the potential investment mix. For these, well-developed markets exist and liquidity is not a concern from the long-term investment perspective that we consider here. Then, we introduce in the allocation decision also several alternative classes widely known to have limited liquidity and immature secondary markets. In the process, we extend the model by Ang et al. (2014) to incorporate multiple risky liquid assets, thus bringing it closer to practice.

We compare several investment universes. The long-term investor will allocate wealth between a money market (MM) account that pays the risk-free rate and (1) long-term government bonds (GB) subject to interest rate risk[16], (2) public equity (PuE), (3) both GB and PuE, or in a combination with an alternative illiquid investment option such as (4) GB and PuE with hedge funds (HF), with (5) GB and PuE with private equity (PrE), with (6) GB and PuE with direct real estate holdings (RE), or with (7) infrastructure (Inf) investments. Cases (1), (2), and (3) can be solved with the continuous rebalancing model of Section 2.3.1. Each of the cases involving an alternative asset class is solved through the model incorporating a liquidity friction from Section 2.4.

2.6.1 Calibration

To calibrate the models, we use data from the capital market assumptions report by JP Morgan (2022). The JPM report provides volatility, expected return, and correlation projections, and is widely used by pension funds and investors to gauge their SAA. The trading probabilities for the illiquid alternative asset classes, on the other hand, are calibrated based on holding period estimates from Ang et al. (2014) and references therein. Tables (2.3) and (2.4) summarize the overall data.

Table 2.3 shows the average time between transactions for several private asset classes according to Ang et al. (2014). We will use this as a crude measure of the severity of the asset class illiquidity. Some asset classes, such as direct infrastructure investments, lie on the very extreme of the illiquidity spectrum. For others, such as private equity and hedge funds, secondary markets have only recently come into existence, and even

[16] We do not model interest rate dynamics explicitly.

then trading is typically very thin, and funds are priced at large discounts (Kleymenova et al., 2012; Ramadorai, 2012). It is still the norm that private equity and venture capital are exited by finding an appropriate counterparty to buy into the investments once the firms in the fund have matured. From that point of view, the average holding time is an appropriate measure of how easy it is for an investor to rebalance back to the SAAs. In hedge funds, investments are usually redeemed directly from the fund manager once contractual lock-ups and notice periods are satisfied. The contractual clauses for different funds can vary, but we are focused on the SAA, which involves a decision on aggregate investment classes before particular funds are selected to invest in. As a result, the average liquidity properties of hedge funds are appropriate.

We use market projections, rather than actual asset return data for several reasons. First, the SAA decision is by definition forward-looking, and JPM's projections represent a reliable source for long-term investors' expectations about market developments. Second, by using projections rather than directly historical data, we avoid engaging in data biases endemic to the historical returns of many alternative assets. Since markets for these assets are frail, data is often based on appraisals rather than market trading. Often, alternative fund managers have the discretion when and how to report historical returns, with the clear incentive to report returns when they are good. Even when trading occurs, selection bias is not excluded, as the tendency exists for the market to generate observable returns when asset prices are high and sellers are willing to enter the market. Furthermore, empirically, it has been observed that infrequent trading tends to over-smooth historical return data (Ang, 2014), biasing the asset variance downward.

2.6.2 Optimal Allocation and Welfare Gains from Alternative Investment

First, we examine if the introduction of one of the alternative asset class leads to a significant improvement in the CEC of agents following the optimal trading strategy. We evaluate the starting period CEC at the rebalancing point ξ^*, assuming that strategic allocations are set at the moment when agents can transact in the illiquid asset class and will set the illiquid asset allocation at the optimal target. Table 2.5 shows the resulting

Table 2.3: Asset Classes

	JPM Class	$1/\eta$	p	μ	σ
MM	U.S. Cash	0.0	1.00	1.30	0.00
GB	U.S. Long Treasuries	0.0	1.00	2.44	14.00
PuE	AC World Equity	0.0	1.00	6.17	12.75
HF	Diversified Hedge Funds	1.0	0.63	3.82	6.84
Pr	Private Equity	4.0	0.22	9.66	18.68
RE	U.S. Core Real Estate	9.0	0.11	3.82	6.84
Inf	Global Core Infrastructure	55.0	0.02	6.64	10.74

Note. This table shows the expected return and volatility asset class data used to calibrate each of the consequent strategic allocation cases.

Table 2.4: Asset Classes, Correlations

	GB	PuE	HF	PrE	RE	Inf
GB	1.0	-0.3	-0.33	-0.57	-0.3	-0.34
PuE	-0.3	1.0	0.76	0.84	0.4	0.6
HF	-0.33	0.76	1.0	0.80	0.37	0.46
PrE	-0.57	0.84	0.80	1.0	0.43	0.65
RE	-0.3	0.4	0.37	0.43	1.0	0.38
Inf	-0.34	0.6	0.46	0..65	0.38	1.0

Note. This table shows the assumed expected correlations between the risky asset classes.

percentage improvement in CEC relative to the benchmark continuous rebalancing case (GB+PuE). Figure 2.6 shows the allocation to all assets in the portfolio for each of the considered cases.

We can see that each alternative asset class provides an economically significant improvement in the risk-free equivalent consumption. Most favorable, with 35% improvement, is the combination of traditional asset classes and direct real estate with 16.4% portfolio allocation. Next in the ranking is the combination with infrastructure investments, despite its relatively low allocation of 6.3% and despite the long assumed average waiting time to transact of 55 years.

Furthermore, we can observe that the illiquid model generates diversified portfolios, thus beating down a major concern that theoretical allocation models may generate overly concentrated portfolios. Between 6% and 27% are allocated to the private asset classes when they are available, and in each case, it is optimal to invest in every available asset class.

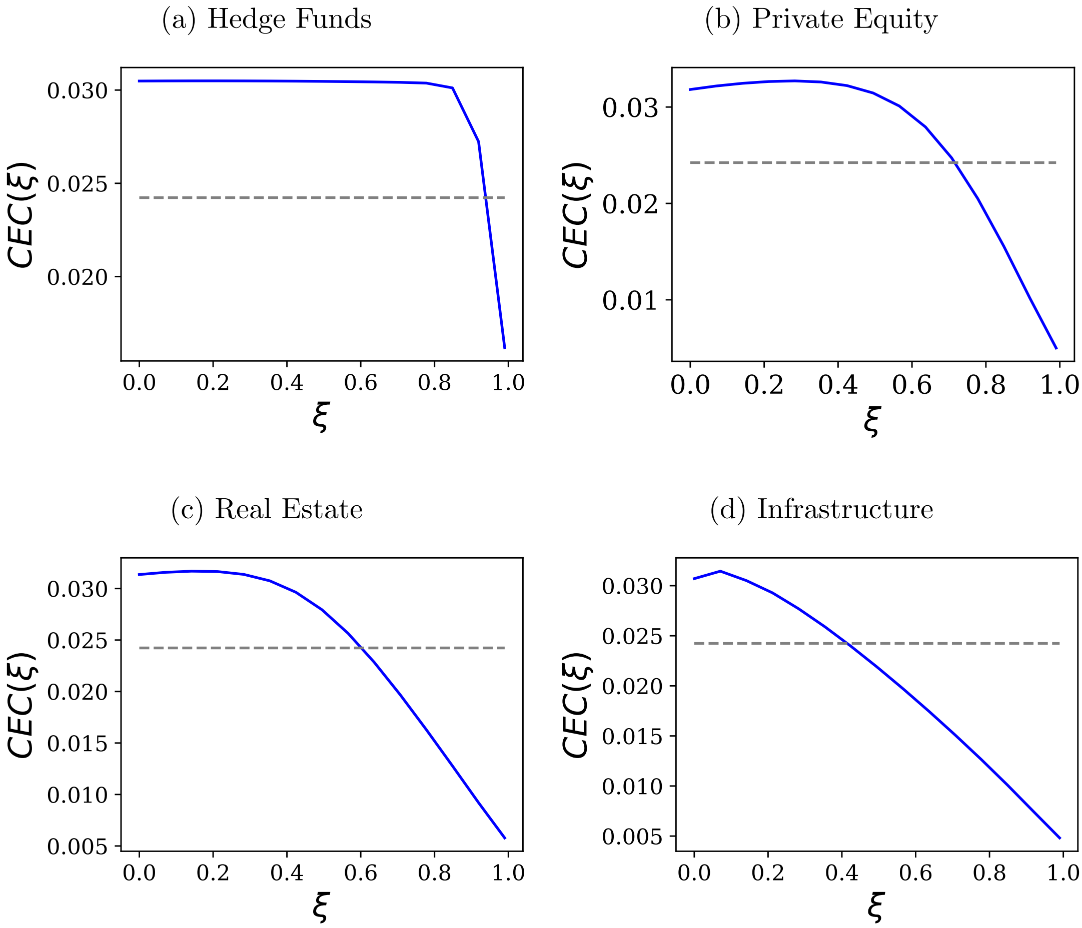
Table 2.5: Asset Classes

	$CEC(\xi^*)$	Improvement
GB	1.47	-40%
PuE	2.11	-13%
GB+PuE	2.42	0%
GB+Pu+HF	3.15	30%
GB+Pu+PrE	3.05	26%
GB+Pu+RE	3.27	35%
GB+Pu+Inf	3.17	31%

Note. This Table shows CEC at the rebalancing point. The last column shows the improvement in welfare, measured by the percentage change in CEC relative to the case where investors have access to government bonds and public equity (GB+Pu).

Figure 2.4 examines the shape of the certainty equivalent as a function of the illiquid asset holdings. The more concave the curve around the rebalancing point at ξ^*, the faster the CEC will fall once the illiquid share of wealth floats away from the optimal strategic allocation. The dashed line shows the threshold below which the illiquid opportunity becomes less favorable in CEC terms compared to the benchmark case of investing in bonds and public equity (the *GB+PuE* case). In all cases, we can see that even though the CEC is falling once the illiquid asset share moves away from the optimal ξ^* at the

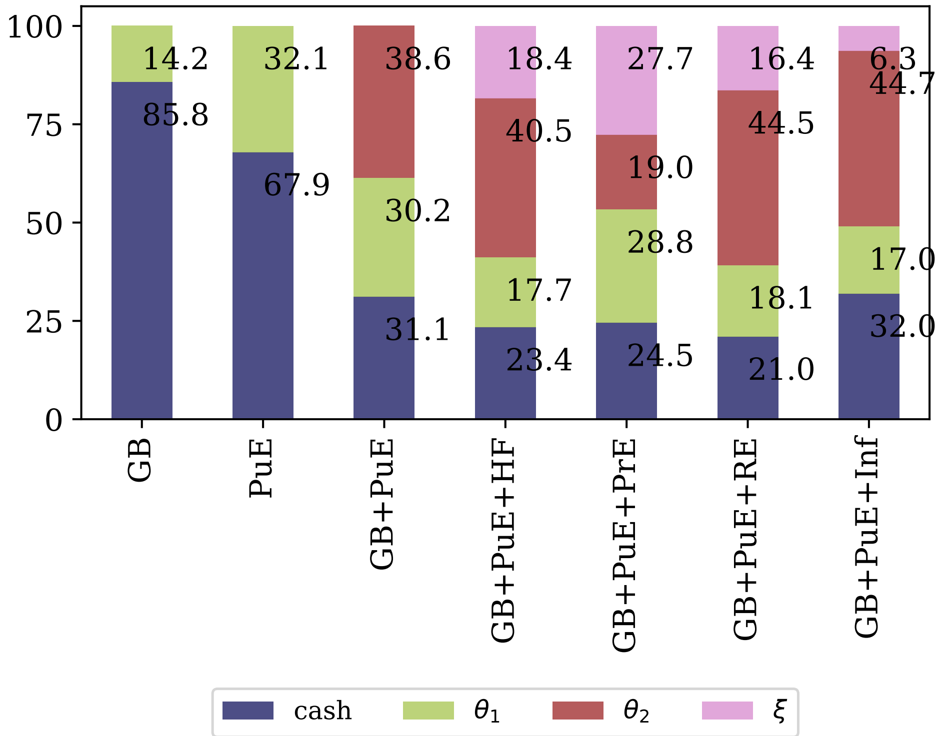
Figure 2.5: CEC Profile per Asset Class



Note. This figure shows the curvature of the CEC profile as a function of the illiquid asset share for each alternative class scenario. The gray dashed lines represent the CEC of the liquid investment scenario of holding MM, GB and PuE.

top of the curve, there is still a sizable buffer until reaching the threshold benchmark line. For example, for the most severe illiquid case of investing in infrastructure (Figure 2.5d), the optimal share of infrastructure is 6.3%. Still, if the share of the illiquid asset stays below around 40% in periods when liquidity is not available, the CEC of this portfolio is still better than the purely liquid portfolio benchmark.

Figure 2.6: Strategic Asset Allocation



Note. This figure shows the allocations to alternative asset classes when the ivestor has access to different allocation universes.

2.7 Conclusion and Further Research

We looked at the presence of stochastic liquidity friction in the asset market and its effect on optimal portfolio choice for an infinitely lived investor. We found economically significant effects from the presence of illiquidity most notably on the allocation to the illiquid asset. The effects correlate with the degree of the friction and manifest in lower consumption rate, reduced holding of the illiquid asset, high utility costs, and sub-optimal holding of the liquid asset compared to a market with full liquidity.

We looked at a specific channel through which illiquidity affects wealth, allocation, and consumption decisions. The investor can consume directly only out of the liquid asset, and transfers between liquid and illiquid wealth are allowed with uncertain timing. This gives rise to risks for the investor of not being able to meet consumption objectives as part of the total investment wealth may stay locked up in illiquid holdings.

As the illiquid asset cannot be traded for periods of random length, it exposes the investor to risks that are not hedgeable. The investor cannot always adjust allocations to account for optimality. Thus when market conditions change, the allocation between liquid and illiquid wealth is driven away from optimality.

Calibrating the model to market projections used by pension funds and long-term investors, we find that despite their illiquid nature, private asset classes have the potential to significantly increase the welfare of investors.

Throughout the paper, we have compared our findings to predictions from the classical baseline Merton model. We illustrated the solution methods for each of the models and drew parallels between them. We illustrated for reasonable data used in the SAA process that the model produces diversified portfolio with reasonable allocations.

Overall, the model presents an intuitive framework for the discussion of liquidity. The definition of illiquidity is crucial and we discussed a number of alternatives considered in the literature such as proportional transaction costs and limitations in the quantity of shares the market can absorb at a given instant. Still, the trading uncertainty and the split between liquid and illiquid risky wealth considered in this study can capture a very wide area of private asset classes and can accommodate the long-term SAA decision.

The portfolio choice model presented here provides a way for practitioners, especially long-term investors in the area of SAA, to embed illiquid investments in their optimization decision. Yet, further extensions are possible and may be necessary to capture the nuances of each specific asset class. Fixed costs of investment, for example, are a main concern for smaller investors, and may be a reason why smaller pension funds tend to shy away from private assets. Often private assets require particular expertise and organization of the due diligence process that make such investments unpopular choice. Furthermore, more research can be put into modeling the common factors that affect returns across asset classes. Extensions of the liquidity friction could, for example, be relaxed by adding the possibility of accessing secondary markets, which are still frail for private assets but have started developing in recent years. Yet, the model presented here provides a sound basis for refinements.

2.A Proofs and Derivations

2.A.1 Liquid Market: Martingale Solution

Restating the static problem specified with the Martingale approach in (2.3.3):

$$\begin{aligned} &\sup_{C_s} E \int_0^\infty \mathrm{e}^{-\beta s} u(C_s) ds \\ &\text{s.t. } E \int_0^\infty M_s C_s ds = W_0 M_0 \end{aligned} \tag{2.32}$$

The problem implies the Lagrange form:

$$\sup_{(C_s)} \mathcal{L} = \sup_{(C_s)} \left\{ E \int_0^\infty \mathrm{e}^{-\beta s} u(C_s) ds + \phi \left(W_0 - E \int_0^\infty M_s C_s ds \right) \right\} \tag{2.33}$$

Apply the first order condition w.r.t. C_s on the Lagrange function and substitute the CRRA utility.

$$\begin{aligned} &\mathrm{e}^{-\beta s} u'(C_s^*) = \phi M_s \\ \iff\ & C_s^* = I_{u'} \left(l M_s \mathrm{e}^{\beta s} \right) \end{aligned}$$

where $I_f(x)$ stands for the inverse function of $f(x)$. In particular, for the CRRA utility function, $u'(C_t) = C_t^{-\gamma}$ implies

$$C_s^* = \left(\phi M_s \mathrm{e}^{\beta s} \right)^{-\frac{1}{\gamma}} \tag{2.34}$$

Substitute the optimal consumption in the budget constraint and solve for the Lagrange constant ϕ at the optimum:

$$W_0 M_0 = W_0 = E \int_0^\infty M_s C_s^* ds = E \int_0^\infty M_s I_{u'} \left(\phi M_s \mathrm{e}^{\beta s} \right) ds$$

In particular for a CRRA utility:

$$W_0 = E \int_0^\infty M_s \left(\phi M_s \mathrm{e}^{\beta s} \right)^{-\frac{1}{\gamma}} ds = (\phi)^{-\frac{1}{\gamma}} E \int_0^\infty \mathrm{e}^{-\frac{\beta}{\gamma} s} M_s^{1-\frac{1}{\gamma}} ds$$

It is useful to define an auxiliary function g_t for $t < s$ such that

$$g_t \equiv E_t \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} \left(\frac{M_s}{M_t}\right)^{1-1/\gamma} ds \tag{2.35}$$

and we get

$$W_0 = (\phi)^{-\frac{1}{\gamma}} g_t \tag{2.36}$$

In that case, note that we also have $g_0 = E \int_0^\infty \mathrm{e}^{-\frac{\beta}{\gamma}s} M_s^{1-1/\gamma} ds$ and we get $W_0 = (\phi)^{-\frac{1}{\gamma}} g_0$. So, we can solve for the optimal Lagrange constant as:

$$\phi^* = \left(\frac{W_0}{g_0}\right)^{-\gamma} = W_0^{-\gamma} g_0^\gamma \tag{2.37}$$

and substitute back in the optimal consumption expression

$$\begin{aligned} C_s^* &= \left(\phi^* M_s \mathrm{e}^{\beta s}\right)^{-\frac{1}{\gamma}} \\ &= \left(W_0^{-\gamma} g_0^\gamma M_s \mathrm{e}^{\beta s}\right)^{-\frac{1}{\gamma}} \\ &= \frac{W_0}{g_0} \mathrm{e}^{-\frac{\beta}{\gamma}s} M_s^{-\frac{1}{\gamma}} \end{aligned} \tag{2.38}$$

Then the optimal wealth trajectory W_t^* can be determined by following the optimal consumption policy, so we substitute C_s^* in the budget constraint:

$$\begin{aligned} W_t^* &= \frac{1}{M_t} E_t \int_t^\infty M_s C_s^* ds \\ &= \frac{1}{M_t} E_t \int_t^\infty M_s \frac{W_0}{g_0} \mathrm{e}^{-\frac{\beta}{\gamma}s} M_s^{-\frac{1}{\gamma}} ds \\ &= \frac{W_0}{g_0} \frac{1}{M_t} E_t \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}s} M_s^{1-1/\gamma} ds \\ &= \frac{W_0}{g_0} M_t^{-\frac{1}{\gamma}} \mathrm{e}^{-\frac{\beta}{\gamma}t} E_t \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} \left(\frac{M_s}{M_t}\right)^{1-1/\gamma} ds \\ &= \frac{W_0}{g_0} M_t^{-\frac{1}{\gamma}} \mathrm{e}^{-\frac{\beta}{\gamma}t} g_t \end{aligned} \tag{2.39}$$

We can also conveniently write the last equation as $\frac{W_t^*}{g_t} = \frac{W_0}{g_0} M_t^{-\frac{1}{\gamma}} \mathrm{e}^{-\frac{\beta}{\gamma}t}$. Note that the right hand side of this is then equal to the optimal consumption from (2.38). This allows us to find current consumption as a function of current wealth, rather than initial wealth. So, reconciling the two we get that g_t is proportional to the optimal wealth-to-consumption ratio:

$$C_t^* = \frac{W_t^*}{g_t} \tag{2.40}$$

As indicated above we can also find C_s^* as a function of the information (wealth) available in time t. It can be derived by restating (2.38), expanding for M_t and combining the result with (2.39):

$$C_s^* = \frac{W_0}{g_0} \mathrm{e}^{-\frac{\beta}{\gamma}s} M_s^{-\frac{1}{\gamma}} = \frac{W_0}{g_0} \mathrm{e}^{-\frac{\beta}{\gamma}t} M_t^{-\frac{1}{\gamma}} \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} \left(\frac{M_s}{M_t}\right)^{-\frac{1}{\gamma}} = \frac{W_t^*}{g_t} \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} \left(\frac{M_s}{M_t}\right)^{-\frac{1}{\gamma}} \tag{2.41}$$

Finally, we can solve 2.35 for an explicit form of the g_t function.

$$\begin{aligned} g_t &= E_t \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} \left(\frac{M_s}{M_t}\right)^{1-1/\gamma} ds \\ &= \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} E_t \left[\left(\frac{M_s}{M_t}\right)^{1-1/\gamma}\right] ds \end{aligned}$$

Using the explicit formula for the SDF process $M_t = \exp\left\{-rt - \boldsymbol{\lambda}'\boldsymbol{Z}_t - \frac{1}{2}\|\boldsymbol{\lambda}\|^2 t\right\}$ we can simplify g_t:

$$\begin{aligned} g_t &= \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} E_t \left[\left(\mathrm{e}^{-r(s-t) - \boldsymbol{\lambda}'(Z_s - Z_t) - \frac{1}{2}\|\boldsymbol{\lambda}\|^2(s-t)}\right)^{1-1/\gamma}\right] ds \\ &= \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t) - \left(1-\frac{1}{\gamma}\right) r(s-t) - \frac{1}{2}\left(1-\frac{1}{\gamma}\right)\|\boldsymbol{\lambda}\|^2(s-t)} E_t \left(\mathrm{e}^{-\left(1-\frac{1}{\gamma}\right)\boldsymbol{\lambda}'(Z_s - Z_t)}\right) ds \end{aligned}$$

We can use the formula for the expectation of the exponent of a normal random variable, where for $X \sim N(\mu, \sigma^2)$ we have $E\left(\mathrm{e}^{\alpha X}\right) = \mathrm{e}^{-\alpha\mu_1 + \frac{1}{2}\alpha^2\sigma^2}$. In our case, $\alpha = -\left(1 - \frac{1}{\gamma}\right)$ and $x = \boldsymbol{\lambda}'(\boldsymbol{Z}_s - \boldsymbol{Z}_t)$, where $\boldsymbol{Z}_s - \boldsymbol{Z}_t$ is a Brownian Motion increment so it is normally distributed with mean zero and variance $\|\boldsymbol{\lambda}\|^2(s-t)$. As a result:

$$g_t = \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t) - \left(1-\frac{1}{\gamma}\right) r(s-t) - \frac{1}{2}\left(1-\frac{1}{\gamma}\right)\|\boldsymbol{\lambda}\|^2(s-t)} \mathrm{e}^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)^2\|\lambda\|^2(s-t)} ds = \int_t^\infty \mathrm{e}^{-A(s-t)} ds = \frac{1}{A}$$

where A is a constant, which can safely be assumed positive for plausible parametrization:

$$\begin{aligned} A &\equiv \frac{\beta + r(\gamma-1)}{\gamma} + \frac{1}{2}\frac{\gamma-1}{\gamma^2}\|\boldsymbol{\lambda}\|^2 \\ &= \frac{\beta + r(\gamma-1)}{\gamma} + \frac{1}{2}\frac{\gamma-1}{\gamma^2}(\boldsymbol{\mu} - r\mathbf{1})'(\boldsymbol{\sigma\sigma'})^{-1}(\boldsymbol{\mu} - r\mathbf{1}) \end{aligned} \tag{2.42}$$

The end result shows that in the infinite-horizon case with intermediate consumption the ratio of optimal consumption to wealth is a constant. From (2.40) we have that $C_t^* = AW_t^*$. As a result:

$$c_t^* = \frac{C_t^*}{W_t^*} = \frac{1}{g(t)} = A = \frac{\beta + r(\gamma-1)}{\gamma} + \frac{1}{2}\frac{\gamma-1}{\gamma^2}\|\boldsymbol{\lambda}\|^2 \tag{2.43}$$

The final step is to solve for $\boldsymbol{\pi}^*$, the optimal investment in the risky asset. For this purpose, we employ (2.39). As $g_t = g_0 = 1/A$ we can cancel the terms out and get:

$$dW^*(M_t, t) = W_t^* = \frac{W_0}{g_0} M_t^{-\frac{1}{\gamma}} \mathrm{e}^{-\frac{\beta}{\gamma}t} g_t = W_0 M_t^{-\frac{1}{\gamma}} \mathrm{e}^{-\frac{\beta}{\gamma}t}$$

Applying Ito's rule then and focusing on the relevant terms we get:

$$\begin{aligned} dW^*(M_t, t) &= \frac{\partial W^*}{\partial t}dt + \frac{\partial W^*}{\partial M_t}dM_t + \frac{1}{2}\frac{\partial^2 W^*}{\partial^2 M_t}[dM_t]^2 \\ &= (...)dt + \frac{1}{\gamma}W_0 \mathrm{e}^{-\frac{\beta}{\gamma}t} M_t^{-1-\frac{1}{\gamma}} M_t \boldsymbol{\lambda' dZ}_t \\ &= (...)dt + \frac{1}{\gamma}W_t^* \boldsymbol{\lambda' dZ}_t \end{aligned} \tag{2.44}$$

From the wealth process we also have

$$dW_t = (...)dt + W_t \boldsymbol{\pi}_t' \boldsymbol{\sigma dZ}_t$$

The underlying assumption in the Martingale approach (2.8) is that wealth is optimally invested. Thus, reconciling the two wealth processes one can get the optimal allocation:

$$\begin{aligned} &\frac{1}{\gamma}\boldsymbol{\lambda'} = \boldsymbol{\pi}_t'\boldsymbol{\sigma} \\ \iff &\boldsymbol{\pi}_t^* = \frac{1}{\gamma}(\boldsymbol{\sigma'})^{-1}\boldsymbol{\lambda} = \frac{1}{\gamma}(\boldsymbol{\sigma\sigma'})^{-1}(\boldsymbol{\mu} - r\mathbb{1}) \end{aligned} \tag{2.45}$$

Using the optimizing consumption and allocation strategies, one can now solve for the value function as well.

$$\begin{aligned}
V(W_t^*) &= E_t \int_t^\infty \mathrm{e}^{-\beta(s-t)} u(C_s^*) ds \\
&= \frac{1}{1-\gamma} E_t \int_t^\infty \mathrm{e}^{-\beta(s-t)} (C_s^*)^{1-\gamma} ds \\
&= \frac{1}{1-\gamma} \left(\frac{W_t^*}{g_t}\right)^{1-\gamma} E_t \int_t^\infty \mathrm{e}^{-\frac{\beta}{\gamma}(s-t)} \left(\frac{M_s}{M_t}\right)^{1-1/\gamma} ds \\
&= \frac{1}{1-\gamma} {g_t}^{\gamma} (W_t^*)^{1-\gamma} \\
&= \frac{1}{1-\gamma} \left(\frac{1}{A}\right)^{\gamma} (W_t^*)^{1-\gamma}
\end{aligned} \tag{2.46}$$

2.A.2 Liquid Market: HJB Equation

Proof. Assuming that the strategies for $\boldsymbol{\pi}_s$ and C_s are set and stay constant for the time interval $s \in [t, t+\Delta t)$ where $\Delta t \to 0$, we can write

$$V(t, W_t) = \sup_{(\boldsymbol{\pi}_s, C_s)} \int_t^{t+\Delta t} \mathrm{e}^{-\beta(s-t)} u(C_s) ds + \mathrm{e}^{-\beta \Delta t} E\Big[V(t+\Delta t, W_{t+\Delta t})\Big]$$

Multiplying both sides by $\frac{1}{\Delta t}\mathrm{e}^{\beta \Delta t}$ and rearranging:

$$\frac{\mathrm{e}^{\beta\Delta t}-1}{\Delta t} V(t, W_t) = \sup_{(\boldsymbol{\pi}_s, C_s)} \frac{1}{\Delta t} \int_t^{t+\Delta t} \mathrm{e}^{-\beta(s-t-\Delta t)} u(C_s) ds + \frac{1}{\Delta t} E\Big[V(t+\Delta t, W_{t+\Delta t}) - V(t, W_t)\Big]$$

Evaluating the above equation for $\Delta t \to 0$, applying the L'Hopital rule, the fact that $\frac{1}{\Delta t}\int_t^{t+\Delta t} f(s) ds = f(t)$ and using the definition of a drift term in a stochastic differential equation by denoting it as $E[dV(t, W_t]$ we get

$$\beta V(t, W_t) = \sup_{(\boldsymbol{\pi}_t, C_t)} u(C_t) + E\Big[dV(t, W_t)\Big]$$

Apply the Itô rule on the stochastic term $dV(t, W_t)$ and substitute in the budget constraint for dW_t

$$\begin{aligned} dV &= \frac{\partial V}{\partial t}dt + V_W dW_t + 1/2 V_{WW}[dW_t]^2 \\ &= \frac{\partial V}{\partial t}dt - C_t V_W dt + V_W W_t(rdt + \boldsymbol{\pi}_t'(\boldsymbol{\mu} - r\mathbb{1})dt + \boldsymbol{\pi}_t'\boldsymbol{\sigma}\boldsymbol{dZ}_t) + \frac{1}{2}V_{WW}W^2\boldsymbol{\pi}_t'\boldsymbol{\Sigma}\boldsymbol{\pi}_t dt \end{aligned}$$

Then use the fact that $E[dZ_j] = 0$, and that in the infinite horizon case $\frac{\partial V}{\partial t} = 0$. Consequently we can derive the drift term of the Bellman equation as

$$E\left[dV(t, W_t)\right]dt = \left[-C_t V_W + V_W W_t(r + \boldsymbol{\pi}_t'(\boldsymbol{\mu} - r\mathbb{1})) + \frac{1}{2}V_{ww}W^2\boldsymbol{\pi}_t'\boldsymbol{\Sigma}\boldsymbol{\pi}_t\right]dt$$

which after substitution yields the HJB equation of Proposition (2.20):

$$\beta V(t, W_t) = \sup_{(\boldsymbol{\pi}_t, C_t)} u(C_t) - C_t V_W + V_W W_t(r + \boldsymbol{\pi}_t'(\boldsymbol{\mu} - r\mathbb{1})) + \frac{1}{2}V_{ww}W^2\boldsymbol{\pi}_t'\boldsymbol{\Sigma}\boldsymbol{\pi}_t$$

□

2.A.3 Illiquid Market: Homogeneity of the Value function

The proof follows from Ang et al. (2014). For CRRA utility in particular, the value function $V(W_t, X_t)$ is homogeneous of degree $1-\gamma$, i.e. $V(kW_t, kX_t) = k^{1-\gamma}V(W_t, X_t)$ for any $k > 0$. This is a direct consequence of the fact that the budget constraint dynamics are linear in wealth and have constant moments (independent of the corresponding wealth states). Then it is reasonable to accept that for an optimal solution $\{W_s^*, X_s^*, dI_s^*, c_s^*, \theta_s^*\}$ also $\{kW_s^*, kX_s^*, kdI_s^*, c_s^*, \theta_s^*\}$ will be optimal as well for any $k > 0$, so that scaling both liquid and illiquid wealth up or down by the same number does not change the optimal investment and consumption rates given that we also scale the wealth transfers dI by the same number.

Then we can write

$$V(kW_t, kX_t) = \sup_{\theta, dI, c} E_t\left[\int_t^\infty \mathrm{e}^{-\beta(s-t)}\frac{(kc_s(1-\xi_s)W_s)^{1-\gamma}}{1-\gamma}ds\right] = k^{1-\gamma}V(W_t, X_t)$$

As a result

$$V(W_t, X_t) = \left(\frac{1}{k}\right)^{1-\gamma} V(kW_t, kX_t)$$

Setting $k = 1/(W_t + X_t)$ we get

$$V(W_t, X_t) = (X_t + W_t)^{1-\gamma} V\left((1-\xi), \xi\right) = (X_t + W_t)^{1-\gamma} H(\xi)$$

where ξ is the portion of total wealth invested in the illiquid asset x and $(1-\xi)$ is the portion invested in the liquid asset. The additional proof that $H(\xi)$ is concave is available in (Ang et al., 2014).

2.A.4 Illiquid Market: HJB Equation

Here we derive the equation (2.24). Starting with the continous time Bellman equation (2.14) with $\xi_t = \frac{X_t}{W_t + X_t}$ as the proportion of total wealth invested in the illiquid asset and θ_t as the proportion of liquid wealth invested in the illiquid asset we get:

$$\begin{aligned}
\beta V(X_t, W_t) &= \sup_{(\xi_t, \theta_t, C_t)} \{u(C_t) + E[dV(X_t, W_t)]\} \\
&= \sup_{(\xi_t, \theta_t, C_t)} \{u(C_t) + V_W W_t(r + (\mu_1 - r)\theta_t) - V_W C_t + V_X X_t \mu_2 \\
&+ \frac{1}{2} V_{WW} W_t^2 \theta^2 \sigma_1^2 + \frac{1}{2} V_{XX} X^2 \sigma_2^2 + V_{WX} W_t X_t \sigma_2 \sigma_1 \rho \theta + \eta(V^* - V(W_t, X_t))\}
\end{aligned}$$

where in the last line the Ito rule for jump processes (see (Shreve, 2004, Chapter 11)) is applied such that:

$$\begin{aligned}
E[dV] &= E\left[V_w dW^c + V_X dX^c + \frac{1}{2}\left(V_{WW}[dW^c]^2 + V_{XX}[dX^c]^2 + 2V_{WX}[dX^c dW^c]\right) + (V^* - V)dN\right] \\
&= \left(r + \theta_t(\mu_1 - r\mathbb{1})\right) V_W W_t + C_t V_W + \frac{1}{2} V_{WW} W_t^2 \theta_t^2 \sigma_1^2 \\
&+ V_X \mu_2 + \frac{1}{2} V_{XX} X^2 \sigma_2^2 + V_{WX} W X \theta \sigma_2 \sigma_1 \rho + \eta(V^* - V(W_t, X_t))
\end{aligned}$$

where we denote as dX^c and dW^c the continuous portion of the budget constraints and thus get the corresponding quadratic variation terms:

$$\begin{aligned}
dW_t^c &= (r + (\mu_1 - r)\theta_t)W dt - C_t + \theta_t \sigma_1 W dZ_{1t} \\
dX_t^c &= \mu_2 X_t dt + \sigma_2 \rho X_t dZ_{1t} + \sigma_2 \sqrt{1-\rho^2} X_t dZ_{2t} \\
[dW_t^c]^2 &= \theta_t^2 \sigma_1^2 W^2 dt \\
[dX_t^c]^2 &= \sigma_2^2 X_t^2 \\
[dW_t^c dX_t^c] &= \theta \sigma_2 \sigma_1 \rho W X
\end{aligned}$$

and denoting the jump size in the value function when a liquidity opportunity arises as $V^* - V(W_t, X_t)$, such that the expected jump size over a short period of time is $\eta(V^* - V(W_t, X_t))dt$. Note that whenever the Poisson jump process hits, we have inferred that the value function jumps to $V^* = (W_t + X_t)^{1-\gamma} H^*$.

This yields the given HJB equation

$$\mathcal{L}^C + \mathcal{L}^\theta + \mathcal{L} - \beta V(W_t, X_t) = 0$$

where

$$\begin{aligned}
\mathcal{L}^C &= \sup_{C_t} \Big\{ u(C_t) - C_t V_W \Big\} \\
\mathcal{L}^\theta &= \sup_{\theta_t} \Big\{ (r + \theta_t(\mu_1 - r\mathbb{1}))\, V_W W_t + \frac{1}{2} V_{WW} W_t^2 \theta_t^2 \sigma_1^2 + V_{WX} W X \theta_t \sigma_2 \sigma_1 \rho \Big\} \\
\mathcal{L} &= V_X \mu_2 + \frac{1}{2} V_{XX} X^2 + \eta(V^* - V(W_t, X_t))
\end{aligned}$$

To solve for consumption and liquid investment in terms the illiquid asset holdings, we apply the substitutions

$$\begin{aligned}
\xi_t &= \frac{X_t}{W_t + X_t} \\
V(W_t, X_t) &= (W_t + X_t)^{1-\gamma} H(\xi_t)
\end{aligned}$$

such that

$$\begin{aligned}
V_W &= (W+X)^{-\gamma}\left((1-\gamma)H(\xi) - \xi H'(\xi)\right)\\
V_X &= (W+X)^{-\gamma}\left((1-\gamma)H(\xi) + (1-\xi)H'(\xi)\right)\\
V_{WW} &= (W+X)^{-\gamma-1}\left(-\gamma(1-\gamma)H(\xi) + 2\xi\gamma H'(\xi) + \xi^2 H''(\xi)\right)\\
V_{XX} &= (W+X)^{-\gamma-1}\left(-\gamma(1-\gamma)H(\xi) - 2(1-\xi)\gamma H'(\xi) + (1-\xi)^2 H''(\xi)\right)\\
V_{WX} &= (W+X)^{-\gamma-1}\left(-\gamma(1-\gamma)H(\xi) - 2(1-\xi)\gamma H'(\xi) - (1-\xi)H''(\xi)\xi\right)
\end{aligned}$$

The partial derivatives are implied using the Chain Rule such that $\frac{\partial \xi}{\partial X} = \frac{1-\xi}{W+X}$ and $\frac{\partial \xi}{\partial W} = -\frac{\xi}{W+X}$, where $H'(\xi_t)$ and $H''(\xi_t)$ denote the first and second partial derivatives of $H(\xi)$ with respect to ξ.

Applying that in the HJB equation we can solve for optimal consumption

$$c_t^* = \Big((1-\gamma)H(\xi_t) - H'(\xi_t)\xi_t\Big)^{-\frac{1}{\gamma}}(1-\xi_t)^{-1}$$

and optimal liquid risk asset investment

$$\theta_t^* = -\frac{k_1 H(\xi_t) + k_2 H'(\xi_t) + k_3 H''(\xi_t)}{k_4 H(\xi_t) + k_5 H'(\xi_t) + k_6 H''(\xi_t)}$$

where $k_1, .., k_6$ are known constants defined by the market parameters and the agent's risk aversion such that

$$\begin{aligned}
k_1 &= -(1-\gamma)(\mu_1 - r) + \gamma(1-\gamma)\sigma_2\rho\sigma_1\xi\\
k_2 &= (\mu_1 - r)\xi - \sigma_2\sigma_1\rho\xi\gamma(2\xi - 1)\\
k_3 &= -\sigma_2\sigma_1\rho\xi^2(1-\xi)\\
k_4 &= -\gamma(1-\gamma)(1-\xi)\sigma_1^2\\
k_5 &= 2\gamma\xi(1-\xi)\sigma_1^2\\
k_6 &= \xi^2(1-\xi)\sigma_1^2
\end{aligned}$$

2.B Discretization & Numerical Solution Methods

2.B.1 The Bellman Equation

The discretized version of the Bellman Equation for the problem (2.21 - 2.22), with $\delta = \mathrm{e}^{-\beta \Delta t}$ as the discrete-time discount factor, can be written as [17]:

$$V(W_t, X_t) = \max_{(\theta_t, dI_t, c_t \in \mathcal{A})} \{u(C_t)\Delta t + \delta E_{W_t, X_t}[V(W_{t+\Delta t}, X_{t+\Delta t})]\} \tag{2.47}$$

The expectation is conditional on liquid wealth W_t and illiquid wealth X_t. Conditioning on t is not needed as we are looking at an infinite horizon problem and both wealth processes have the Markovian property. Using the homothetic properties of the CRRA utility function we can write

$$\begin{aligned} V(W_t, X_t) &= Q_t^{(1-\gamma)} H(\xi_t) \\ u(C_t) &= u(c_t W_t) \\ &= Q_t^{1-\gamma} u(c_t(1-\xi_t)) \end{aligned}$$

where $Q_t = W_t + X_t$ stands for total wealth and $W_t = (1 - \xi_t)Q_t$.

Knowing the function $H(\xi)$ allows us to find the optimal ratio of illiquid to total wealth ξ^* as the maximizing value for that function. Our goal is then to write the Bellman equation and to solve for the control variables in terms of ξ_t and $H(\xi_t)$.

The Bellman Equation can then also be written as:

$$\begin{aligned} V(Q_t, \xi_t) &= \max_{(\theta_t, \xi_t, c_t \in \mathcal{R})} \{u(c_t(1-\xi_t)Q_t)\Delta t + \delta E_{Q_t, \xi_t}[V(Q_{t+\Delta t}, \xi_{t+\Delta t})]\} \\ Q_t^{(1-\gamma)} H(\xi_t) &= \max_{(\theta_t, \xi_t, c_t \in \mathcal{R})} \{Q_t^{(1-\gamma)} u(c_t(1-\xi_t))\Delta t + \delta E_{Q_t, \xi_t}[Q_{t+\Delta t}^{(1-\gamma)} H(\xi_{t+\Delta t})]\} \end{aligned} \tag{2.48}$$

As illustrated on Figure 2.7, with probability p the agent will be able to trade the illiquid asset next period and will bring the portion of illiquid wealth to the desired optimal level $\xi^* = arg\max_\xi H(\xi)$. This is done by setting the transfer dI_t between liquid and illiquid wealth to accommodate optimal asset allocations, so:

$$dI_t = \begin{cases} \xi^* Q_t - X_{t-} & \text{with probab. } p \\ 0 & \text{with probab. } 1-p \end{cases}$$

where X_{t-} is the level of illiquid holdings just before rebalancing takes place. As a result, we eliminated the need to solve directly for dI_t by taking into account that the agent will set $\xi_{t+\Delta t} = \xi^*$ whenever trading is possible. If trading is not possible, the agent cannot rebalance and is stuck with sub-optimal levels of illiquid holdings $\xi_{t+\Delta t} = X_{t+\Delta t}/Q_{t+\Delta t}$, and the ratio will float away from last period's value as the prices of the two risky assets move erratically.

Combining those two states with the corresponding realization probabilities, we can use the law of iterated expectations[18] to drop the conditional expectation with respect to ξ.[19]

$$Q_t^{(1-\gamma)} H(\xi_t) = \max_{(\theta_t, dI_t, c_t \in \mathcal{R})} \left\{ Q_t^{(1-\gamma)} u(c_t(1-\xi_t))\Delta t + \delta \Big(p E_{Q_t}[Q_{t+\Delta t}^{(1-\gamma)}] H^* + (1-p) E_{Q_t}[Q_{t+\Delta t}^{(1-\gamma)} H(\xi_{t+\Delta t})] \Big) \right\}$$

Canceling out wealth from both sides of the equation yields desired form of the equation. Note, that by canceling out Q_t and by embedding it in the expectations the conditional part of the expectation with respect to Q has been dropped:

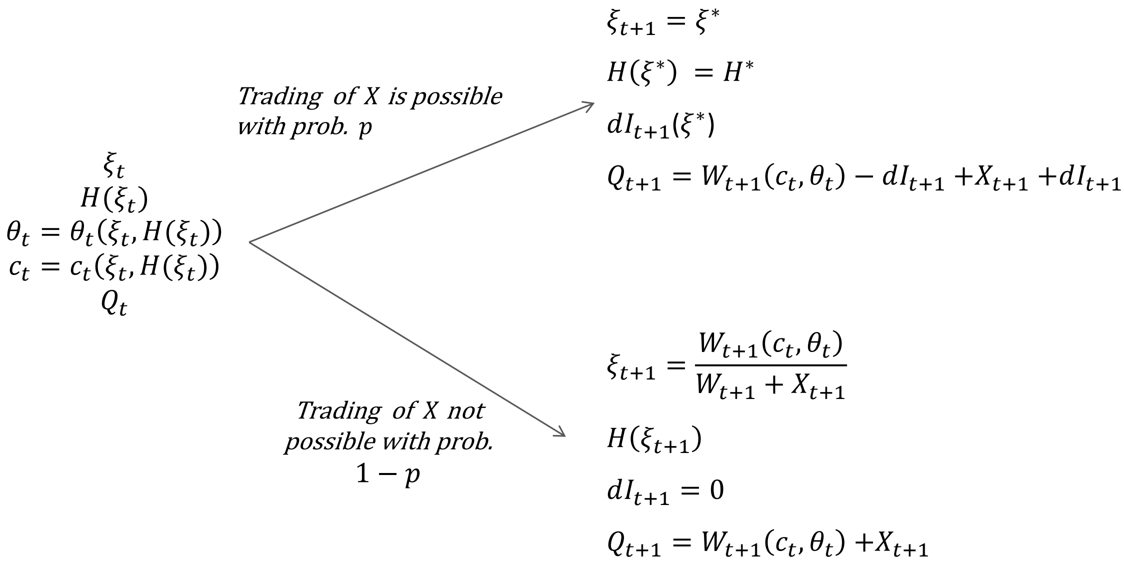
$$H(\xi_t) = \max_{(\theta_t, \xi_t, c_t)} \left\{ u(c_t(1-\xi_t))\Delta t + \delta \left(p H^* E_{\xi_t} \left[R_{q,t+\Delta t}^{1-\gamma} \right] + (1-p) E_{\xi_t} \left[R_{q,t+\Delta t}^{1-\gamma} H(\xi_{t+\Delta t}) \right] \right) \right\} \tag{2.49}$$

where $R_{q,t+\Delta t}$ is the growth of total wealth in the next period net of current consumption, and $\xi_{t+\Delta t}$ is the holdings ratio given that rebalancing is not possible. The laws of motion

[18] The law as applied here states that $E(X|Y_1) = E(E(X|Y_1, Y_2)|Y_2)$.

[19] A similar approach to decomposing the Bellman equation is used also by Moore and Young (2006). They apply it as part of a Markov Chain approximation method in order to solve a portfolio choice problem with insurable loss which occurs with a Poisson probability.

Figure 2.7: State Transition Dynamics



This chart illustrates the dynamics behind the optimal choice problem. In the coming period, the illiquid asset share floats freely to $\xi_{t+\Delta t}$. With probability p the agent can trade in the illiquid asset and can set it back to the optimal target of ξ^* by maximizing the known function $H(\xi)$. With probability $1-p$ the agent cannot trade and is stuck with the illiquid asset share of $\xi_{t+\Delta t}$.

then for these two variables are such that:

$$\begin{aligned} R_{q,t+\Delta t} &= \frac{Q_{t+\Delta t}}{Q_t} = \frac{W_t R_{w,t+\Delta t} + X_t R_{x,t+\Delta t}}{Q_t} \\ &= (1-\xi_t) R_{w,t+\Delta t} + \xi R_{x,t+\Delta t} \\ \xi_{t+\Delta t} &= \frac{X_{t+\Delta t}}{Q_{t+1}} = \frac{X_t}{Q_t} \frac{X_{t+\Delta t}}{X_t} \frac{Q_t}{Q_{t+\Delta t}} \\ &= \xi_t \frac{R_{x,t+\Delta t}}{R_{q,t+\Delta t}} \end{aligned} \tag{2.50}$$

Where $R_{w,t+\Delta t}$ and $R_{x,t+\Delta t}$ are the discretized gross returns on liquid and illiquid wealth respectively after factoring out consumption. Using Euler time-discretization of the continuous stochastic process we get:

$$\begin{aligned} R_{w,t+\Delta t} &= \frac{W_{t+\Delta t}}{W_t} = (r + (\mu_1 - r)\theta_t - c_t)\Delta t + \sigma_1 \sqrt{\Delta t} \Delta Z_{1,t} \\ 1 + R_{x,t+\Delta t} &= \frac{X_{t+\Delta t}}{X_t} = \mu_2 \Delta t + \sigma_2 \rho \Delta Z_{1,t} + \sigma_2 \sqrt{1-\rho^2} \sqrt{\Delta t} \Delta Z_{2,t} \end{aligned} \tag{2.51}$$

where $\Delta Z_{1,t}$ and $\Delta Z_{2,t}$ are two independent standard normal random variables. We use a Gaussian quadrature approach for the space-discretization of the normal distribution and for the evaluation of the expectations in the Euler equation.

The discretized Bellman equation can then be solved through value function iteration combined with standard numerical techniques. In Appendix 2.B.3 we show the algorithm used for the purpose.

2.B.2 Trading Probability

The Poisson process is a counting process, which itself is defined as a stochastic process $\{N_t, t \in [0,\infty)\}$ where N_t represents the total number of events that have occurred by time t, such that $N_t \geq 0$, $N(t)$ is integer valued and for $t_1 < t_2$, $N_{t_1} \leq N_{t_2}$ and for $t_1 < t_2$, $N_{t_2} - N_{t_1}$ equals the number of events that occur in the interval $(t_1; t_2]$. The time between Poisson events becomes stochastic and in fact can be shown to be exponentially distributed with rate $1/\eta$, where the exponential density is given by $f(x) = \frac{1}{\eta}e^{-x\frac{1}{\eta}}, \quad x \geq 0$. See e.g. Ross (2007) for further discussion.

We know that a Poisson process has stationary and independent increments, where the number of events that occur during any time increment of length Δt is Poisson distributed with mean $\eta\Delta t$. Formally, for any $t \geq 0, \Delta t > 0$, the probability of n events occurring can be written as:

$$P(N_{t+\Delta t} - N_t = n) = e^{-\eta\Delta t}\frac{(\eta\Delta t)^n}{n!}, \quad n = 0, 1, ...$$

[19] To provide some intuition on the discretization step, note that the discrete Bellman equation

$$V(W_t) = \sup_{(\pi_t, C_t)} \{u(C_t)\Delta t + e^{-\beta\Delta t}E[V(W_{t+\Delta t})]\}$$

can be shown to converge to its continuous counterpart for $\Delta t \to 0$. Multiply the equation by $e^{\beta\Delta t}$, subtract $V(W)$ from both sides and divide by Δt and this results in:

$$\frac{e^{\beta\Delta t} - 1}{\Delta t}V(W) = \sup_{(\pi_t \in \mathcal{R}^d, c_t \geq 0)} \{u(c_t) + \frac{1}{\Delta t}E_t[V(W_{t+\Delta t}) - V(W_t)]\}$$

Let $\Delta t \to 0$, then $e^{\beta\Delta t} \to 1$ and also by the L'Hôpital rule we have that $\frac{e^{\beta\Delta t}-1}{\Delta t} \to \beta$. As a result $\frac{1}{\Delta t}E_t[V(W_{t+\Delta t}) - V(W_t)] \to E_t[dV]$ which makes discretized equation equivalent to the continuous time version of (2.15).

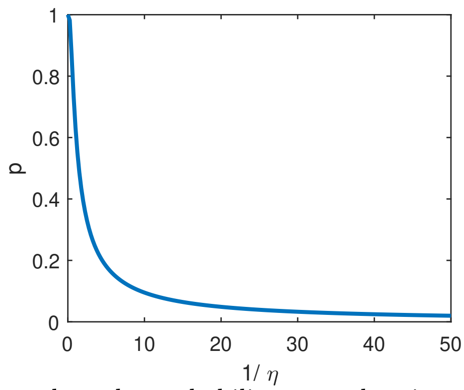
The probability of having a trading opportunity over the next period Δt can be calibrated through the intensity η of the Poisson process which determines the trading opportunities for the illiquid asset. Poisson arrivals are a common way to model search frictions following the seminal paper by Diamond (1982). Note one major difference between the Brownian Motion used to drive the asset return dynamics and the Poisson Process used to drive the liquidity dynamics: in the former, the size of the movement and thus its variance is time-dependent, while in the latter, the probability of occurrence is time-dependent.

We can then calibrate the illiquid asset's trading probability as the Poisson probability of having at least one trading event over time period Δt, given that the average time to wait for a trading opportunity over a year is $1/\eta$, and the average number of trading opportunities per year correspondingly is η. In that case:

$$p = P(N_{t+\Delta t} - N_t \geq 1) = 1 - P(N_{t+\Delta t} - N_t = 0) = 1 - \mathrm{e}^{-\eta \Delta t} \tag{2.52}$$

Figure 2.8 illustrates this functional relationship.

Figure 2.8: Relationship between p and η



The plot shows the connection based on the probability to trade p in an illiquid asset within a year and η, the average waiting time in years between trades.

2.B.3 Value Function Iteration

Based on the discretization from Section 2.B.1, we can solve the portfolio choice dynamic problem through value function iteration. Discretizing ξ and simulating the system one period ahead will allow us to iterate (2.49) until the iterative approximation of the value function $H(\xi)$ converges. The procedure will eventually yield a numerical approximation of the value function and the optimal solution for the policy functions $c(\xi)$ and $\theta(\xi)$. So, the goal is to find the vector $\tilde{h} = [h_1, h_2, ..., h_N]$ on a grid $\tilde{\xi}$ which best approximates the function $H(\xi)$. We use the following iterative procedure:

Initialization: Discretize the random space (the random variables ΔZ_1 and ΔZ_2 in (2.51)) through a simulation or strategically selected (Gaussian) quadrature points. Select an approximation grid of N gridpoints for ξ_t: $\tilde{\xi} = \{\xi_1, ..., \xi_N\} \in [0; 1)$, $j = 1, .., N$. We use $N = 20$. Select a class of approximating functions $h(\boldsymbol{a}; \tilde{\xi})$ which will approximate the true value function $H(\xi)$ and initialize the functional parameters $\boldsymbol{a}$. Select an initial maximum of the value function h^*.[20] As a starting point for the iteration we use the analytical two-asset Merton solution. We can then initiate the following iterative algorithm.

1. **Optimization:** For each ξ_j in the grid compute optimal consumption and liquid asset allocation:

$$c^{*,k}, \theta^{*,k+1} =$$
$$\arg\min \Big\{ u(c(1-\xi_j))\Delta t + \delta \Big(ph^* E\left[\hat{q}(c,\theta,\xi_j)^{1-\gamma}\right] + (1-p) E\Big[h\Big(\boldsymbol{a}^k; \hat{\xi}(c,\theta,\xi_j)\Big) \hat{q}(c,\theta,\xi_j)^{1-\gamma}$$

where $\hat{\xi}()$ and $\hat{q}()$ are next-period's dynamics calculated through (2.50), and $k = 1, ..., p$ is a counter measuring the iteration run.

2. **Update:** For each ξ_j and the optimal control policies found in the previous step update values of the value function that lie on the grid. Update the fit of the approximation function $h(\boldsymbol{a}^{k+1}; \xi)$ based on the new values. Update $h^{*,k+1} = \arg\max_\xi h(\boldsymbol{a}^{k+1}; \xi)$.

3. **Stopping:** The algorithms stops if $\left\| \ln\big(h(\boldsymbol{a}^{k+1}; \xi)\big) - \ln\big(h(\boldsymbol{a}^k; \xi)\big) \right\|^2 < \epsilon$, otherwise we go to **Step 1**. I use $\epsilon = 0.1^6$

The expectation operator in the second step is evaluated through multi-variate quadrature (Judd, 1998; Cai et al., 2013; Cai and Judd, 2014).

[20] We have successfully tried simple polynomials and cubic splines for the purpose. Eventually, the later turned out to provide more flexibility, so all the shown results are derived through this functional form.

3

Intergenerational Risk Sharing with Market Liquidity Risk

3.1 Introduction

There are well-known hurdles in a free market economy to sharing risk between generations that are born over distinct periods and thus are subject to different economic prosperity over their lifetimes. Due to the natural physical limitation of a finite lifetime, individuals cannot directly participate in risk that materializes before or after they become economically active. Combined with a lack of a strong bequest motive this creates a classical incomplete market inefficiency. A policy intervention that sets contingent transfers between young and old generations can improve social welfare by widening the risk-bearing pool in the economy and thus increasing its capacity to bear risk.[1]

However, any illiquidity that comes in the form of uncertain transaction cost to be paid when selling an asset from the lifetime savings mix of households has the potential to lower the benefits of such transfers. An increase in illiquidity, first of all, discourages individuals from holding risky assets, as the transaction cost that could be incurred when

[0]Acknowledgements: I would like to thank Roel Beetsma for the support and the valuable discussions which made the current paper possible. The paper draws extensively on those discussions. I am also thankful for the useful remarks and comments from Sweder van Wijnbergen, Christian Stoltenberg, Alessia Russo, Hans Schumacher, Agnieszka Markiewicz, Albert Jan Hummel and all participants in the MInt Macro Seminar at the University of Amsterdam, and the Tinbergen Institute Job Market seminars.

[1]See Merton (1981); Gordon and Varian (1988); Shiller (1999); Ball and Mankiw (2007); Gottardi and Kubler (2011); Lancia et al. (2020). Beetsma and Romp (2016) provide an overview of the growing literature of intergenerational risk sharing, its policy relevance, and institutional arrangements.

exiting the investment lowers the expected return from the asset. Second, it compresses the distribution of returns and thus lowers the range of asset returns that can be realized[2]. As a result, there is less financial risk to be shared between generations, and the potential benefits from sharing the risk between generations is lower. We find that higher illiquidity, measured as higher expected transaction costs in selling an asset, requires higher levels of risk sharing to compensate for the loss of sensitivity of the risk transfers to the asset variance. Nevertheless, the total improvement in welfare compared to a situation of no risk-sharing decreases with the level of illiquidity.

The fields of intergenerational risk sharing (IRS) and asset illiquidity overlap naturally when we consider the typical structure of lifetime savings and investments, and the institutions that manage them. First, long-term investors are often seen as well-poised to bear liquidity risk[3]. In the search for diversification and return potential, pension funds in the developed world tend to allocate significant portions of their portfolios to alternative asset classes such as hedge funds, infrastructure, real estate, and private equity funds[4]. These alternative investments impose a liquidity cost that can be a substantial source of investment risk. Second, housing wealth tends to account for a significant share of the retirement wealth of individuals worldwide and the marketability of housing is found to be a significant factor affecting the well-being of retirees[5].

Nevertheless, there is currently, to the best of our knowledge, no other study that explores the intersection between risk sharing (intergenerational or otherwise) and market illiquidity. This is a significant gap in the literature, given that the implications of illiquidity are well known for portfolio choice (Ang et al., 2014; Constantinides, 1986; Acharya and Pedersen, 2005), as well as for the conduct of fiscal (Kaplan and Violante, 2014) and monetary (Sousa, 2010; Chatziantoniou et al., 2017) policy. In this study, we connect the personal finance aspect of illiquidity to its policy relevance.

[2]Assume a fixed proportional liquidation cost $\bar{l} \in (0,1)$ exists such that the realized returns net of liquidation costs are measured as $R(1-\bar{l})$. The variance of the realized returns will be $(1-\bar{l})^2\mathbb{Var}(R) < \mathbb{Var}(R)$. The treatment of illiquidity here is more nuanced (see Section 3.3.2) but follows this line of thinking.

[3]Academically, the point has been made for example by Ang (2014); Amihud and Mendelson (1991); Gârleanu (2009)

[4]See, for example, data from OECD (2019); PensionsEurope (2018)

[5]Refer to Lusardi and Mitchell (2007); Crawford and O'Dea (2020); Shao et al. (2019); Nakajima and Telyukova (2020); Munk (2020)

We define illiquidity in an *ex-ante* sense as the expected proportional cost that needs to be paid to liquidate a risky investment. In terms of modeling the risk of illiquidity, we use a simple on and off shock that materializes with a given probability each period. This relates to the approach of Acharya and Pedersen (2005) who define illiquidity as a latent factor with a time-varying cost component in order to establish testable hypotheses of the way liquidity risk affects asset prices. The binary liquidity cost assumed in this paper can then be seen as a wedge between the fair value of the asset and its realizable market value in the spirit of Brunnermeier and Pedersen (2009).[6]

Our modeling approach relates to the fact that investments in alternative asset classes, such as private equity, often suffer large haircuts on their NAV when taken to the secondary market (Nadauld et al., 2019; Bollen and Sensoy, 2015; Albuquerque et al., 2018). It can also be seen as an expression of the latent costs associated with illiquidity, such as foregone earnings or diversification loss due to trading delays. Note that we look at the liquidity of the investor's portfolio around the time of switching from old age to young age. We ignore any illiquidity effects occurring before that. Expanding the model with more granular time periodicity, however, could also take that into account.

We develop a stylized framework with two overlapping generations (OLG) to consider the outlined problem. Wealth shocks arise from the returns of risky assets in the savings portfolio of individuals and from liquidation costs when the portfolio is sold to fund retirement consumption. Shocks each period occur before the current young have accessed the capital markets, and before they have made any investment decisions. The young start with labor endowment that is not affected by the current shock while the old bear financial risk on their savings. In a fully decentralized market economy, the young are making consumption, savings, and allocation decisions that optimize their lifetime utility, while the old consume from the accumulated retirement wealth.

This arrangement leaves room for an institutional designer to intervene and enforce transfers between the young and the old, which are contingent on the accumulated return of the old generation's savings portfolio. The transfers are designed from an ex-ante point

[6]Brunnermeier and Pedersen (2009) distinguish between *market liquidity*, the ease with which an asset can be placed on the market, and *funding liquidity*, the ease with which outside funds can be accessed once a liability shock hits on an agent's balance sheet. In this paper, liquidity is of the first kind as it concerns only the marketability of accumulated assets at a particular point in the lifetime of agents associated with retirement age.

of view and welfare in the economy is evaluated before any shocks materialize. The transfers are linear in the realized return on the individuals' retirement portfolio after paying out any liquidation costs and act as a partial insurance on retirement wealth, covered by the young. Once aware of the policy implementation, utility-optimizing individuals adjust their savings mix by factoring in the regulated transfer policy, giving rise also to indirect welfare effects.

We show that the link between optimal risk-sharing and risk itself can be split into two opposing effects. On one hand, a policy that engages the young in the shock that otherwise affects only the old widens the pool of people who can participate in that shock and increases the risk-bearing capacity of the economy. At the same time, this imports additional risk in the youth's labor endowment, thus extending the horizon over which individuals bear risk. Cumulatively, the latter effect also leads to more risk in their old age. The larger the variance of the asset is, the more the second effect dominates, and thus the lower optimal risk sharing needs to be. Similarly, the lower the asset variance, the more the first effect dominates, and the higher the optimal risk sharing should be.

We abstract from the particularities of the institutions through which IRS occurs. In reality, the contingent transfers between young and old, as modeled here, could be the result of several arrangements. First, risk sharing rules could be embedded in a collective pension system, for example, through indexation of the benefits received and contributions paid based on the funding ratio of the pension plan (Cai et al., 2013; Gollier, 2008). Alternatively, they could be implemented through counter-cyclical adjustments in the tax code in combination with adjustments to the public debt, through the pay-as-you-go pension system, or by some combination of each of these (Chen et al., 2016).[7]

As a benchmark risk-sharing case, we look at a planner solution, where the planner invests on behalf of the young and allocates consumption centrally between all young and

[7] One can refer to the existing literature on details about the optimal institutional arrangements of IRS. Beetsma and Romp (2016) provide an overview of the institutional side; Bovenberg and Mehlkopf (2014) review the literature on funded pension schemes, exploring the overlap between life-cycle investing and IRS, elaborating on commitment issues, problems of intergenerational fairness, and sustainability of the pension contract. Gollier (2008) (revisited by Schumacher (2021)) looks at a collective pension plan which allocates funds between a risky and a riskless asset and pays out benefits on a rolling-window basis. Similarly, Cui et al. (2011) look at risk sharing within funded plans with defined-benefit and hybrid structures, where IRS occurs through adjustments in the contribution and benefit levels.

old generations. In the spirit of Gollier (2008) this allows the clever use of wealth buffers to spread risk between generations that do not necessarily live in the same time periods.

Overall, we find that IRS mechanism increases the young's capacity to bear liquidity risk and allows them to allocate more wealth to illiquid assets compared to the case when those individuals are saving in isolation from the shocks that other cohorts are experiencing. We extend the results from earlier models which show that IRS increases the demand for risky assets (Gollier, 2008; Campbell and Nosbusch, 2007) by showing that the same effect holds for illiquid assets as well.

Quantitatively, we show that contingent transfers between two generations, as a second-best implementation of intergenerational risk-sharing (IRS) to what a central planner can do, can achieve a welfare improvement relative to the no-risk sharing case that is not too far from the benchmark first-best solution. For a reasonable parametrization based on global asset returns, we find that when the young can borrow, a policymaker should set risk sharing to 5% of the asset returns variation for risky liquid assets (and 2.1% for illiquid risky assets if agents are constrained), achieving 36% welfare improvement (17% improvement in the constrained case) relative to the no-risk-sharing case, when welfare is measured in the ex ante sense, i.e. before the realization of any shocks. Illiquid risky holdings by individuals increase by 61% on average after they adjust their portfolio to the policy. The benchmark planner case, on the other hand, realizes a welfare improvement of 48% by being able to spread risk among infinitely many generations.[8]

The paper continues as follows. Section 3.2 provides a short literature overview of two separate fields relating to the current paper, IRS and portfolio choice with illiquidity frictions, and puts the current paper in perspective. Section 3.3 provides the basic structure of the overlapping generations in the economy, defines the social welfare function, constructs the illiquid asset and discusses its properties. Section 3.4 defines the benchmark model with an infinitely lived planner. Section 3.5 redefines the problem for a decentralized economy where each generation solves its own savings-consumption-allocation optimization, while a policymaker determines the transfer policy between young and old. Section 3.6 explores the main mechanisms of risk sharing by exploring the analytical solutions to several simplified cases which illustrate the benefits of pooling risks versus the costs of

[8] Refer to tables (3.3) and (3.2) for further details.

compounding risky returns over time. This provides a rationale for role of asset variance and illiquidity in the context of IRS. Finally, Section 3.7 provides a welfare analysis and discusses the quantitative results.

3.2 Relation to the Literature

Several pioneering studies provide the theoretical backing for this paper. Gordon and Varian (1988) show the main mechanisms behind IRS and illustrate the possibility for welfare improvements for all generations together with the constraints that such a policy needs to handle. Even though their argument was initially developed to provide a non-Keynesian justification for the use of debt and social security transfers towards unlucky generations as a counter-cyclical policy, it also provided economic intuition for the existence of social security systems as risk-sharing mechanisms. We embed their arguments in a more formal OLG set-up, introducing a clear-cut welfare rule for the policymaker to set optimal policy, and add illiquidity risk to the investment asset.

Shiller (1999) argues that designing a social security system is a problem of creating a tool for optimal risk management, placing it naturally in the realm of theoretical finance and asset pricing. The planner problem is non-trivial compared to the standard problem of designing individual optimal asset allocation under risk. The risk-sharing system has to be implemented in a way that generations that are either not born yet or are not economically active, can participate in shocks currently occurring. We extend that point of view, arguing that the risk-sharing properties of the social security system should also be able to consider illiquidity of savings.

Within a general equilibrium framework, Ball and Mankiw (2007) develop the rationale for a funded social security trust that is sensitive to equity shocks in order to achieve an efficient allocation of risk across generations. Lancia et al. (2020) look at risk sharing between generations when the social planner policy cannot be enforced and needs to ensure that the participation constraints of each cohort are satisfied. In that case, a trade-off emerges between the efficiency of the policy and its sustainability over time. The current paper draws from their formulation of the policymaker welfare function while keeping the social policy mandatory and embedding it in a richer asset allocation context.

Merton (1981) develops the rationale that social security, if appropriately designed, can indirectly allow people to trade some of their human capital for partial old-age market-risk insurance. IRS policies thus allow agents to participate early in their lifetime in lotteries that otherwise materialize in old age. In aggregate, this widens the pool of risk participants each period and expands the risk-bearing capacity of the economy. The modeling framework in the current paper is different, but we also come to the same conclusion and extend the known results to the case where market illiquidity is present.

We relate also to the literature on portfolio choice with illiquidity and with transaction costs. In Ang et al. (2014), the illiquid asset is marketable only when liquidity materializes with the arrival of a Poisson shock. We relax this assumption by allowing access to a secondary market by accepting a price discount on the fair value of the asset, rather than barring trading altogether. This makes the properties of the asset more suitable to a two-period model of lifetime dynamics. Calibrating a period to 30-years, an individual should be able to always sell the illiquid asset within that time frame, what will vary is only whether a liquidation cost is paid or not. At the same, we keep the risk component that illiquidity has, deviating from the common assumption of fixed proportional cost in the transaction cost literature (Magill and Constantinides, 1976; Cai et al., 2013).

We also loosely relate to the macro literature of durable investments with life-cycle portfolio choice, and their policy implications. Kaplan and Violante (2014) in particular look at consumers' response to a fiscal stimulus when holding non-durable assets, and find that the ratio of housing to total wealth has significant implications on the effectiveness and timely response consumption demand to fiscal expansions. We add in the discussion the risk properties of liquidity and provide another regulatory perspective.

3.3 The Model

3.3.1 Assumptions

Time is discrete and indexed by $t \in \{0, 1, 2, 3 \dots\}$. There is a small open overlapping generations (OLG) economy, where each generation lives for a fixed duration of two periods (youth and old age), the two cohorts are of equal sizes, each cohort has homogeneous preferences and receives the same fixed endowment, there is no population growth, and there is no technological progress. All stochastic variables are defined by the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and all variables indexed by t are measurable w.r.t. the filtration $\mathcal{F}_t$ which defines all public information. Agents form expectations conditional on current information and there is no information asymmetry between agents and policymakers.

Individuals have time-separable discounted lifetime utility of consumption which can be written as:

$$u_y(C_{y,t}) + \beta \mathbb{E}_t u_o(C_{o,t+1})$$

where $u_y(.)$ and $u_o(.)$ stand for the utility of consumption of the young and the old, with positive and diminishing marginal utility of consumption, and $\beta \in (0, 1]$ is a subjective lifetime discount factor for the agent. In the general case, we assume that the Inada conditions for utility hold, even though in Section 3.6, we break this assumption for illustrative purposes.

At the beginning of their lifetime, individuals receive a fixed endowment Y. The endowment which is not consumed can be saved and transferred for consumption in the next period through several investment opportunities. First, agents can invest in a risk-free asset with fixed gross return of $R_f > 1$. Second, a frictionless market for N risky assets exists, where the price of each asset i follows a stochastic process $P_t^{s,i}$, and its gross return is defined as

$$\frac{P_t^{s,i}}{P_{t-1}^{s,i}} = R_t^{s,i} = \mu_{s,i} + \epsilon_t^{s,i}$$

Third, an illiquid market exists where an asset can be bought at a price P_t^x but can only be sold at the price $P_t^x(1 - l_t)$ where the liquidity cost l_t evolves independently of any asset shocks $\epsilon_t^{s,i}$ and ϵ_t^x, and follows an i.i.d. stochastic jump process such that

$$l_t = \begin{cases} 0 & \text{with probab.} \quad p \\ \bar{l} & \text{with probab.} \quad 1 - p \end{cases} \tag{3.1}$$

The proportional liquidity cost $\bar{l} \in (0, 1)$ stands for the price discount over the fair value of the asset at the time of the sale in case an illiquidity shock hits. Illiquidity is thus asymmetric and presents only downside risk. Short-selling of risky assets is not allowed.

3.3.2 Properties of the Illiquid Return

The gross return of the illiquid asset, excluding the effect ot illiquidity itself, is defined as $R_t^x = \mu_x + \epsilon_t^x$ and we have $\left[\epsilon_t^1, \ldots, \epsilon_t^N, \epsilon_t^x\right]' \sim IID(0, \Sigma)$. In a two period setting, assuming that agents buy the illiquid asset when young and sell it when old, we can write the after-liquidation return as:

$$\begin{aligned} \tilde{R}_t^x &= \frac{P_t^x(1 - l_t)}{P_{t-1}^x} \\ &= R_t^x(1 - l_t) = \mu_x - \mu_x l_t + \epsilon_t^x(1 - l_t) \end{aligned} \tag{3.2}$$

The illiquidity component is thus first of all a drag on the expected return of the asset. At the same time, the illiquidity shock interacts with the asset-specific risk component ϵ_t^x, and whenever a liquidity shock hits, it lowers the magnitude of the asset specific return.

Note that we can also write the gross return of the asset in a way that isolates the expected return from the noise term, where each of the two take into account the effect of illiquidity:

$$\tilde{R}_t^x = \tilde{\mu}_x + \tilde{\epsilon^x}_t \tag{3.3}$$

with

$$\begin{aligned} \tilde{\mu}_x &\equiv \mathbb{E}(\tilde{R}_t^x) = \mu_x(1 - \mathbb{E}(l_t)) \\ &= \mu_x\left(p + (1 - \bar{l})(1 - p)\right) \\ \tilde{\epsilon^x}_t &\equiv R_t(1 - l_t) - \tilde{\mu}_x \end{aligned} \tag{3.4}$$

The liquid asset then will be a special case with either p or $\bar{l}$ set to zero.

Formally, we define *ex-ante* illiquidity as the expected proportional cost that needs to be paid when selling the illiquid asset:

$$\mathbb{E}(l_t) = \bar{l}(1 - p) \tag{3.5}$$

As a result, *ex-ante* illiquidity will be increasing in the liquidation cost of the asset $\bar{l}$ and will be decreasing in the probability of incurring this cost p. *Ex-post* illiquidity, on the other hand, is quantified as $\bar{l}$ and measures the proportional transaction cost that needs to be paid given that the illiquidity risk has materialized. Going forward, unless specified otherwise, the illiquidity considered here refers to the ex-ante type.

We can then isolate several properties of the illiquid asset return. First of all, the expected return of the illiquid asset is monotonously decreasing with the severity of the liquidity friction (as $\bar{l}$ increases or p decreases). This is clear from (3.4). Taking first derivatives, we get $\frac{\partial \tilde{\mu}_x}{\partial \bar{l}} = \bar{l} < 0$ and $\frac{\partial \tilde{\mu}_x}{\partial p} = p - 1 > 0$.

Second, as the asset becomes more illiquid, the expected quadratic variations in the asset returns become smaller. To see that, note that the independence between the illiquidity shock l_t and the shock ϵ_t implies that

$$\begin{aligned}\mathbb{E}\left((\tilde{R}^x_{t+1})^2\right) &= \mathbb{E}\left((R^x_{t+1})^2(1-l_{t+1})^2\right) = \mathbb{E}\left((R^x_{t+1})^2\right)\mathbb{E}\left((1-l_{t+1})^2\right) \\ &= \left(\mu_x^2 + \sigma_x^2\right)\left(p + (1-\bar{l})^2(1-p)\right)\end{aligned} \tag{3.6}$$

Then, the partial derivatives of $\mathbb{E}(\tilde{R}_{t+1})^2$ are:

$$\begin{aligned}\frac{\partial \mathbb{E}\left((\tilde{R}^x_{t+1})^2\right)}{\partial p} &= \left(\mu_x^2 + \sigma_x^2\right)\bar{l}(2-\bar{l}) > 0 \\ \frac{\partial \mathbb{E}\left((\tilde{R}^x_{t+1})^2\right)}{\partial \bar{l}} &= 2\left(\mu_x^2 + \sigma_x^2\right)(1-\bar{l})(p-1) < 0\end{aligned} \tag{3.7}$$

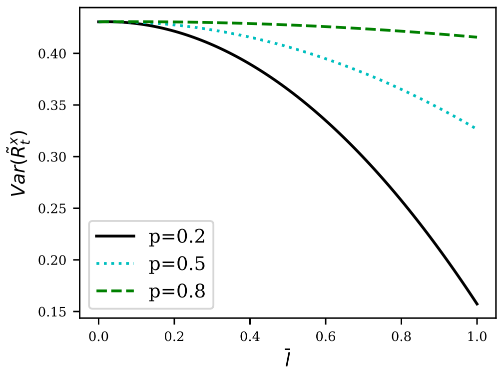
.

The effect of illiquidity on the variance of the asset returns is unclear in general, as both the expectation and the expected variation of asset returns are decreasing. To see that formally, note that

$$\begin{aligned}\tilde{\sigma}_x^2 \equiv \mathbb{Var}\left(\tilde{R}^x_t\right) &= \mathbb{E}\left((\tilde{R}^x_t)^2\right) - \left(\mathbb{E}\tilde{R}^x_t\right)^2 \\ &= \mathbb{E}\left((R^x_t)^2(1-l_t)^2\right) - \left(\mathbb{E}R^x_t\right)^2\left(\mathbb{E}(1-l_t)\right)^2 \\ &= (\sigma_x^2 + \mu_x^2)(p + (1-p)(1-\bar{l})^2) - \mu_x^2(p + (1-p)(1-\bar{l}))^2\end{aligned}$$

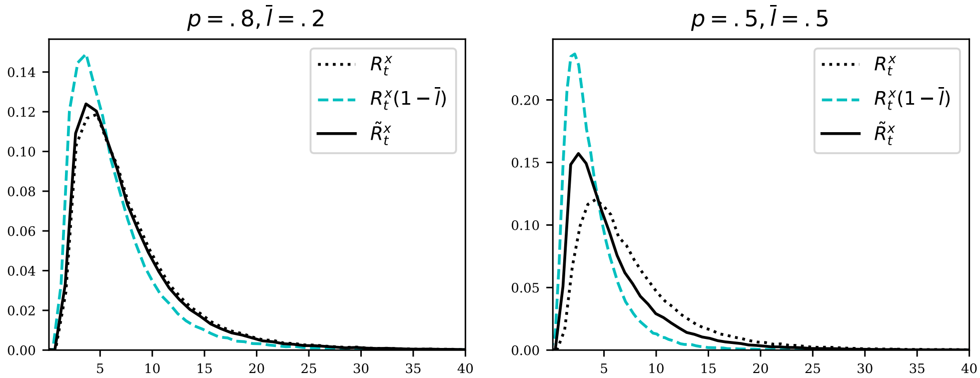
Even though in general the effect is ambiguous, Figure 3.1 illustrates that for a reasonable parametrization the variance will be monotonously decreasing in illiquidity. Figure 3.2 illustrates how the distribution of the illiquid asset return is formed by mixing the distribution of the *ex-post* liquid returns R_t^x and the distribution of the *ex-post* illiquid returns $R_t^x(1-\bar{l})$. The higher the liquidation cost $\bar{l}$, the more the distribution of $\tilde{R}_t^x$ (illustrated with the dashed-line distribution) is shifted to the left, and the lower is the resulting range of possible returns. At the same time, when the trading probability p is low, the *ex-post* illiquid returns distribution dominates when forming the distribution for $\tilde{R}_t^x$ and in the extreme case of p approaching one, the ex-post and ex ante distributions will merge.

Figure 3.1: Illiquid Return Variance



Note. This plot shows the effect of varying the liquidity parameters to the variance of asset return $\tilde{R}_t^x$ if R_t^x is log-normally distributed with annualized mean .061 and variance of .156, and the asset is held for 30 years.

Figure 3.2: Gross Return with Illiquidity



Note. This figure illustrates the effect of the stochastic illiquidity cost l_t on the gross return distribution of the illiquid asset. The dotted line shows the distribution of the ex-post liquid returns. The dashed line shows the ex-post illiquid returns, where the cost $l_t = \bar{l}$ is paid in all scenarios. The solid line in each case shows the ex-ante illiquid returns $\tilde{R}_t = R_t(1-l_t)$, where l_t is unknown in advance. Returns are log-normally distributed with $\mu_x = .061$ and $\sigma^x = .156$, and the asset is held for 30 years.

Finally, note that the combined term $\tilde{\epsilon}^x$, satisfies the same properties that are otherwise natural for a liquid asset, regardless of the liquidity parameters:

$$\begin{aligned} \mathbb{E}(\tilde{\epsilon}^x) &= 0 \\ \mathbb{E}((\tilde{\epsilon}^x)^2) &= \mathbb{E}(\tilde{\epsilon}^x \tilde{R}_t^x) = \mathbb{Var}(\tilde{\epsilon}_t^x) \equiv \tilde{\sigma}_x^2 \end{aligned} \tag{3.8}$$

3.3.3 Social Welfare

Welfare is quantified *ex-ante* through the unconditional expectation with respect to all generations' lifetime utilities in all possible states of the world, over all future time periods. The resulting social welfare is the discounted sum of the weighted expected utilities of all future young and old generations:

$$V_0 = \mathbb{E}\left(\sum_{t=1}^{\infty} \delta^{t-1}\left(\frac{\beta}{\delta} u_o(C_{o,t}) + u_y(C_{y,t})\right)\right) \tag{3.9}$$

where $\delta < 1$ is a policy-relevant discount factor and $\frac{\beta}{\delta}$ keeps the relative social weights between young and old utility fixed between time periods[9][10].

It is worth noting that in order to make the problem tractable, we abstract from some real-world complexity. First of all, we look at a partial equilibrium setting, justified by the assumption of a small open economy, such that world market returns are left unaffected by investment or consumption decisions within the home country. Thus, asset market returns are assumed to be exogenous and any possible general equilibrium effects on asset prices and on economic growth once the risk sharing system is implemented are ignored. Also, we ignore any spillover effects from the investment. In reality, potential investments in illiquid assets which finance for example infrastructure projects could have a positive spillover on social welfare. Second, we concentrate on risk coming from asset holdings and ignore labor income risk and possible correlations between labor and financial market

[9] Equivalently, we can also write the expectation as conditional on all information that the policymaker has available in period $t = 0$, as only consumption happening after period zero is policy relevant, and shocks happening in period one are independent from the realizations in period zero.

[10] This welfare specification is similar to Lancia et al. (2020) who use it in a social planner setting to develop an optimal intergenerational insurance rule under a voluntary scheme. The approach relates back to Ball and Mankiw (2007).

earnings[11]. Third, we focus on a purely utilitarian approach and ignore any political risks on keeping the policy. It is well known that risk sharing is welfare improving for all future generations on an ex-ante basis (before shocks materialize), and not necessarily beneficial for a particular generation on an ex-post basis, as paying compensation after the shock has materialized will make a particular generation worse off (Ball and Mankiw, 2007). Here, generations pre-commit to the scheme before they are born, and participation is mandatory. In reality, there is an incentive for the young to walk away from the arrangement if a negative asset return shock occurs, or the old to walk away if a positive asset return shock occurs. Finally, following the standard approach of a representative agent, we abstract also from any heterogeneity within cohorts.

3.4 Planner Problem

First, we consider a mechanism for optimizing social welfare, defined through an infinitely lived planner. The planner is taking over the young generation's labor endowment and is providing consumption for the young and the old every period. Any residual is invested on the market with allocations optimally set across the available assets. The resulting problem is in line with Gollier (2008), whose planner simultaneously optimizes over retirement benefits and the investment allocation for multiple overlapping generations. Two generations are used to illustrate the dynamics of the problem, even though in theory the model can be extended to cover multiple generations. In contrast to Gollier, we model consumption for the young in addition to retirement (old-age) consumption to capture more completely lifetime motives of investment. The illiquid asset here extends the investment universe of Gollier's problem and provides an additional dimension for asset allocation.

Using liquid and illiquid wealth as separate state variables is common in the portfolio choice literature whenever there is a transaction cost (Cai et al., 2013) or a liquidity friction (Ang et al., 2014) associated with one of the assets. We follow the same convention here. The planner allocates aggregate savings between liquid wealth W_t (consisting of a risk-

[11] Models relating labor income risks and investment shocks have been developed, for example, by Hemert (2005); Krueger and Kubler (2006); Boelaars and Mehlkopf (2018)

free liquid investment and a risky liquid investment S_t) and illiquid wealth X_t, which is managed by withdrawing amounts D_t^+ and investing amounts D_t^-. Note that modeling the flows into and out of illiquid wealth as separate choice variables allows the introduction of asymmetric liquidity costs while also allowing for differentiability of the objective with respect to all decision variables. Here whenever some amount is withdrawn from illiquid wealth, the stochastic proportional cost l_t has to be paid for being able to access the market.

This gives rise to the intertemporal wealth constraints:

$$\begin{aligned} W_{t+1} &= \left(W_t + Y - C_{y,t} - C_{o,t} - D_t^+ + D_t^-(1-l_t)\right) R^f + S_t' r_{t+1}^s \\ X_{t+1} &= \left(X_t + D_t^+ - D_t^-\right) R_{t+1}^x \end{aligned} \tag{3.10}$$

where $r_{t+1}^s = R_{t+1}^s - R^f \mathbb{1}$ is the vector of excess returns on the liquid risky assets, and $\mathbb{1}$ is a vector of ones.

Denoting $r_{t+1}^x = R_{t+1}^x - R^f$ as the excess return on the illiquid asset, we can see that total wealth $Q_{t+1} = W_{t+1} + X_{t+1}$ evolves as:

$$Q_{t+1} = (W_t + Y - C_{y,t} - C_{o,t})R^f + (D_t^+ - D_t^-) r_{t+1}^x + S_t' r_{t+1}^s + X_t R_t^x - D_t^- l_t R^f$$

Wealth is thus being destroyed each period when the liquidity shock hits through the term $D_t^- l_t R^f$, as the planner needs to pay the costs of withdrawing from illiquid wealth instead of earning the risk-free rate on this investment.

The solvency region $\mathcal{A}$ is defined by several constraints. First, borrowing is allowed up to a limited amount $L \geq 0$, so that aggregate consumption and investment do not exceed the available liquid wealth and income by more than the limit amount. Since the planner needs to stay solvent in all states of nature, the withdrawal amount is corrected by the maximum liquidity costs $\bar{l}$ that can be paid:

$$C_{y,t} + C_{o,t} + D_t^+ - D_t^-(1-\bar{l}) + S_t' \mathbb{1} \leq W_t + Y + L \tag{3.11}$$

Second, the illiquid asset cannot be set up as collateral, indicating that the amount withdrawn from illiquid wealth cannot be larger than illiquid wealth itself. The liquid risky asset cannot be short as well. These lead to the following constraints, respectively:

$$\begin{aligned} &D_t^- \leq X_t \\ &D_t^-, S_t, D_t^+ \geq 0 \end{aligned} \tag{3.12}$$

The planner is maximizing the ex-ante social welfare defined in equation (3.9). Following Bellman's principle of optimality, we can re-write it in recursive form as:

$$V(W_t, X_t) = \max_{C_{y,t}, C_{o,t}, S_t, D_t^+, D_t^- \in \mathcal{A}} \left\{ \tilde{u}(C_{y,t}, C_{o,t}) + \delta \mathbb{E} V(W_{t+1}, X_{t+1}) \right\} \tag{3.13}$$

with $\tilde{u}(C_{y,t}, C_{o,t}) = \frac{\beta}{\delta} u_o(C_{o,t}) + u_y(C_{y,t})$.

In optimality, as shown in the appendix (3.A.1) the planner will then set the consumption of the young and the old such that

$$u_y'(C_{y,t}) = \frac{\beta}{\delta} u_o'(C_{o,t}) = V_W(W_t, X_t) \tag{3.14}$$

The appendix derives also the first-order relations with respect to the investments in each risky asset.

3.5 Intergenerational Transfer Scheme

Now, we transition from an economy fully governed by a planner to one where generations make independent savings and asset allocation decisions. In the process, we introduce a policymaker, operating in that environment, who decides on welfare-improving transfers between the young and the old.

3.5.1 The Individuals' Problem

In a decentralized framework, agents decide how much of their endowment Y to save and how to allocate it across liquid and illiquid wealth (W_t and X_t respectively). Individuals are solving a similar problem to the planner, with the difference that now they face a limited horizon and have to liquidate all holdings before retirement, paying any liquidity cost if such arise. Unlike the planner, who can take advantage of illiquid wealth buffers over time, a single generation has to liquidate all wealth in the second period of their life to finance retirement consumption.

Between periods zero and one, transfer policy T_t is introduced between the young and the old. The policy is not anticipated before its introduction[12]. All future cohorts are obliged to participate without a walk-out option. The transfers can be either positive or negative for each cohort depending on the realization of the risky returns. The evolution of wealth from young age to retirement can then be written as

$$\begin{aligned} W_{t+1} &= (Y - C_{y,t} - D_t^+ - T_t)R^f + S_t' r_{t+1}^s \\ X_{t+1} &= D_t^+ R_{t+1}^x \end{aligned}$$

In old age, agents sell all accumulated assets paying any liquidation fees, and consume their retirement wealth net of the transfers T_{t+1} with the new-born cohort:

$$C_{o,t+1} = W_{t+1} + X_{t+1}(1 - l_{t+1}) + T_{t+1}$$

I ignore any bequests in the utility specification. This keeps the model tractable, avoiding any time path dependencies across generations. On an intuitive level, it can be expected that the stronger the bequest motive, the closer the decentralized solution will get to the planner solution defined earlier.

Denoting $M_t \equiv Y - C_{y,t} - S_t'\mathbb{1} - D_t^+ - T_t$ as the investment in the risk-free asset and combining the equations above, we get a simpler formulation of the problem. Individuals optimize consumption, taking the current state of the world and any transfer policy at the time they are born as given. This gives rise to the following optimization problem:

$$\begin{aligned} &\max_{M_t, S_t, D_t^+} \{u_y(C_{y,t}) + \beta \mathbb{E}_t u_o(C_{o,t+1})\} \\ \text{s.t.} \quad & C_{y,t} = Y - I_t'\mathbb{1} - T_t \\ & C_{o,t+1} = I_t' R_{t+1} + T_{t+1} \end{aligned} \tag{3.15}$$

[12] A more granular multi-period specification can incorporate anticipation effects. In a two-period setting, however, where each period represents 30 years the implementation of an unanticipated policy seems more realistic.

where $R_{t+1} = \left[R_f, (R^s_{t+1})', R^x_{t+1}(1 - l_{t+1})\right]'$ is a vector of asset returns net of any liquidation fees, and $I_t = \left[M_t, S'_t, D^+_t\right]'$ is a vector of investment amounts allocated across all available assets.

The transfers between generations that we consider are driven purely by risk sharing and thus are designed to be neutral in expectation. As the policymaker does not have a re-distributive objective, there is no sharing in the expected asset returns. This is ensured by considering transfers which are linear in the shock.[13] The transfers are thus given by $T_{t\geq 1} = T(\tau, R_t)$ where τ is a vector of parameters standing for the portion of the risk in each asset that is transferred from the old to the young.

To ensure that there is no systematic component and expected transfers are zero, the transfer function is constrained to be linear in the deviations of asset returns, net of any liquidation costs:

$$T(\tau, R_t) = Y\tau' \left(\mathbb{E}(R_t) - R_t\right) \tag{3.16}$$

To illustrate, assume that there is only one risky asset and that τ is set to be positive. A negative shock ($\epsilon_t = R_t - \mu < 0$) then implies that returns fall below the expectation, which in turn implies that the transfer $T_t = -\tau\epsilon_t$ will have a positive value for the old, so the young will partially reimburse the old for their losses. If the shock is positive on the other hand, this implies that the old will share a proportion of their excess return with the young.

We can then write the individual's first-order optimizing conditions with respect to each asset in the portfolio as a vector equation:

$$\begin{aligned} \mathbb{1}u'_y(C_{y,t}) &= \beta\mathbb{E}_t R_{t+1} u'_o(C_{o,t+1}) \\ \implies \mathbb{1}u'_y(Y - I'_t\mathbb{1} + T(\tau, R_t)) &= \beta\mathbb{E}_t R_{t+1} u'_o(I'_t R_{t+1} - T(\tau, R_{t+1})) \end{aligned} \tag{3.17}$$

where $\mathbb{1}$ is a vector of ones, $u'_i(.)$ stands for the marginal utility of young-age or respectively old-age consumption with $i = y, o$.

[13] While linear transfers are to some extent restrictive, this does capture first-order effects and significantly simplifies the consequent optimization problems. Non-linear transfers which will further increase welfare are possible, but the added insight relative to the added modeling complexity is likely to be low.

The system (3.17) implicitly defines the optimal investment amounts as a function of the policy instruments and of the realized random shocks, which can be written as $I_t = I(\tau, R_t)$. Substituting into the budget constraints of (3.15) we get the resulting optimal consumption policies:

$$\begin{aligned} C_y(\tau, R_t) &= Y - I(\tau, R_t)'\mathbb{1} - T(\tau, R_t) \\ C_o(\tau, R_t) &= I(\tau, R_t)'R_{t+1} + T(\tau, R_{t+1}) \end{aligned} \tag{3.18}$$

Note that, when a policymaker sets the risk-sharing parameters τ, this affects optimal consumption in two ways: first, through the transfers that are directly dependent on the policy parameters, and second, through the adjustment that individuals make on their savings and asset allocations mix in anticipation of the policy. This will then guide the marginal effect on the individuals' utility from changing the policy parameters.

As policy anticipation effects have been ruled out, the generation born in period zero will not factor in the possibility of a transfer in its optimal investment-consumption decision, but still will get to participate in the risk-sharing scheme once it is old:

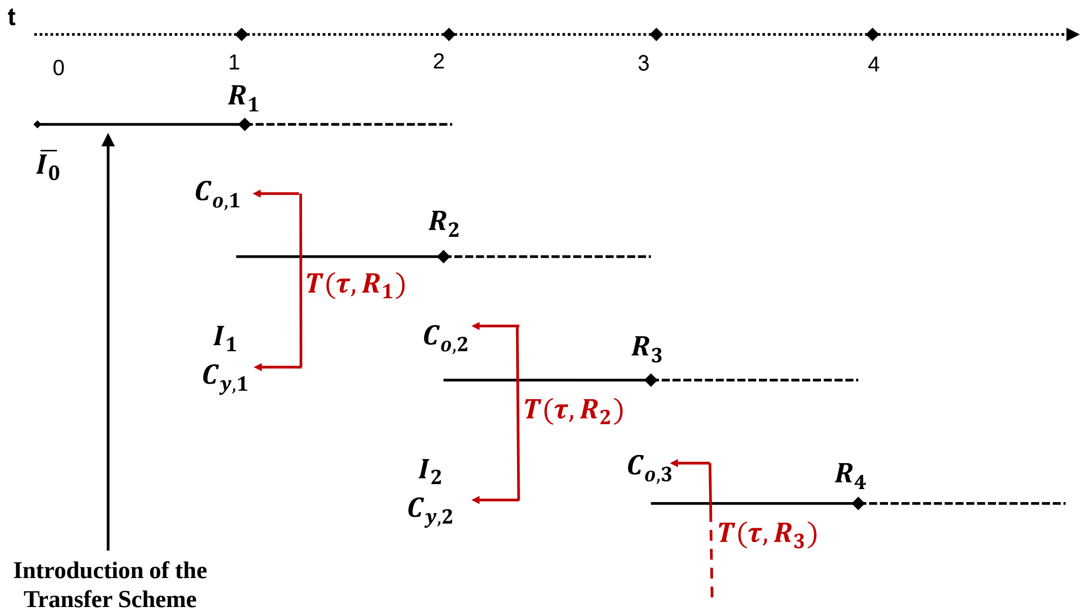
$$\overline{C}_o(\tau, R_1) = \overline{I}'R_1 + T(\tau, R_1) \tag{3.19}$$

where the investment amounts in the vector $\overline{I}$ are fixed before the risk-sharing policy is implemented. This implies that while $\overline{I}$ is set without anticipating the ensuing installment of a transfer scheme when the generation born in period zero reaches old age it gets to participate in the risk-sharing scheme, and they are compensated by the young born in period one if a negative shock is realized or get to transfer to the young some of the accumulated wealth if the shock is positive. Thus, only the direct channel of transfers will affect their lifetime utility.

3.5.2 Policymaker's Problem

The policymaker maximizes the welfare for current and future generations by implementing the IRS policy between periods zero and period one as illustrated in Figure 3.3.

Figure 3.3: Two-period OLG Model



This figure illustrates the timing of the intergenerational transfers in the presented model, the policy introduction and the overlapping structure of the generations.

She fine-tunes the transfers, knowing that the young will take the transfer policy into account when choosing consumption and asset allocation. In this context, the welfare function in (3.9) becomes an indirect utility that arises from summing up and weighting each generation's optimization problem as a function of the policy instrument τ.

To illustrate that, first note that we can write the social welfare function from (3.9) as:

$$V_0 = \mathbb{E}\frac{\beta}{\delta}u(C_{o,1}) + \mathbb{E}\left[u_y(C_{y,1}) + \beta\mathbb{E}_1 u_o(C_{o,2})\right] + \delta\mathbb{E}\left[u_y(C_{y,2}) + \beta\mathbb{E}_2 u_o(C_{o,3})\right] + \ldots$$

We can then substitute in the individuals' optimal consumption from (3.15). Denote the optimal lifetime utility of a generation as $v(\tau, R_t)$. Since asset returns are independent and identically distributed, and the optimal decision of each generation born after period zero is equivalent to the optimal decision of each consequent generation, the problem is

[13] The planner and the policymaker problems defined here fall under a Ramsey planner macro treatment where the policymaker has a restricted set of policy instruments at her disposal. The pension finance literature is relatively loose in defining both as social planner problems, even though in macro context there is a strict distinction between the two. Ball and Mankiw (2007) provide a link between the social planner and the Ramsey planner problems in the context of risk sharing with conditions on the social planner weights which ensure equivalence to a Ramsey solution.

stationary, and when looked at from period zero, $v(\tau, R_t)$ is identical in expectation for any t. We can factor it out of the sum, such that:

$$V(\tau) = \mathbb{E}\frac{\beta}{\delta}u\left(\overline{C}_o(\tau, R_1)\right) + \sum_{j=1}^{\infty}\delta^{j-1}\mathbb{E}v(\tau, R_t) = \mathbb{E}\frac{\beta}{\delta}u(\overline{C}_o(\tau, R_1)) + \mathbb{E}v(\tau, R_t)\sum_{j=1}^{\infty}\delta^{j-1}$$

which results in the indirect utility of the policymaker as a function of the policy instruments

$$V(\tau) = \mathbb{E}\frac{\beta}{\delta}u\left(\overline{C}_o(\tau, R_1)\right) + \frac{1}{1-\delta}\mathbb{E}v(\tau, R_t) \tag{3.20}$$

The policymaker then solves for the optimal transfer parameters τ^*:

$$\tau^* = \arg\max_{\tau}\left\{\frac{\beta}{\delta}\mathbb{E}u_o\left(\overline{C}_o(\tau, R_1)\right) + \frac{1}{1-\delta}\mathbb{E}u_y\left(C_y(\tau, R_t)\right) + \frac{\beta}{1-\delta}\mathbb{E}u_o(C_o(\tau, R_t))\right\} \tag{3.21}$$

The degree of risk sharing will then naturally depend on the individuals' optimal investments as determined in (3.17), or equivalently, the resulting optimal consumption from (3.18-3.19). This implies that the government needs to balance the utility of the old generation present immediately after the scheme is implemented with young age and old age utilities of future cohorts, weighted appropriately through the discount factors of the policymaker and the individuals. To reduce the notation overload going forward, we write $C_{o,1}$,$C_{y,t}$ and $C_{o,t}$ while keeping in mind that each of these satisfies the forms of (3.18) and (3.19).

Assuming that the expectation operator and the derivative can be interchanged, the optimality condition with respect to one of the instruments i can be written as:

$$\frac{\partial V(\tau)}{\partial \tau_i}: \quad \frac{\beta}{\delta}\frac{\partial \mathbb{E}u_o\left(C_{o,1}\right)}{\partial \tau_i} + \frac{1}{1-\delta}\frac{\partial \mathbb{E}v(\tau, R_t)}{\partial \tau_i} = 0 \tag{3.22}$$

As all consumption terms are assumed to satisfy the individual optimality conditions, relying on the Envelope Theorem for the individuals' optimal consumption sensitivity to τ, we can write:

$$\frac{\beta}{\delta}\mathbb{E}\left[u_o'(C_{o,1})\frac{\partial C_{o,1}}{\partial \tau_i}\right] + \frac{1}{1-\delta}\mathbb{E}\left[u_y'(C_{y,t})\frac{\partial C_{y,t}}{\partial \tau_i} + \beta u_o'(C_{o,t})\frac{\partial C_{o,t}}{\partial \tau_i}\right] = 0 \tag{3.23}$$

This can further be expanded by splitting the expectation-of-product terms into expectations and covariances, and noting that for generation zero $\mathbb{E}\frac{\partial C_{o,1}}{\partial \tau_i} = 0$, we have :

$$\begin{aligned}\frac{\beta}{\delta}\mathbb{C}\mathrm{ov}\left(u_o'(C_{o,1}), \frac{\partial C_{o,1}}{\partial \tau_i}\right) + \frac{1}{1-\delta}\left[\mathbb{E}u_y'(C_{y,t})\mathbb{E}\frac{\partial C_{y,t}}{\partial \tau_i} + \mathbb{C}\mathrm{ov}\left(u_y'(C_{y,t}), \frac{\partial C_{y,t}}{\partial \tau_i}\right)\right] \\ + \frac{\beta}{1-\delta}\left[\mathbb{E}u_o'(C_{o,t})\mathbb{E}\frac{\partial C_{o,t}}{\partial \tau_i} + \mathbb{C}\mathrm{ov}\left(u_o'(C_{o,t}), \frac{\partial C_{o,t}}{\partial \tau_i}\right)\right] = 0\end{aligned} \tag{3.24}$$

3.6 Main Mechanism with Quadratic Utility

In the general set-up, in optimality the individuals' condition (3.17) and the policymaker's condition (3.22) are simultaneously fulfilled. Now, we look at several special cases, which keep this set up but simplify the optimization conditions to make the resulting problem analytically tractable and to provide intuition in the dynamics of the model.

Assume for now that agents have the same quadratic period utility of the form $u(C) = C - \frac{\gamma}{2}C^2$ implying that their expected utility is a function of the mean and the variance of the random payoff C such that

$$\mathbb{E}u(C) = \mathbb{E}C - \frac{\gamma}{2}\left(\mathbb{V}\mathrm{ar}C + (\mathbb{E}C)^2\right) \tag{3.25}$$

where $\gamma > 0$ defines the degree of risk aversion, with higher γ implying higher aversion. Once the shock is realized, and consumption is deterministic, it also defines the marginal utility of consumption, evaluated as $u'(C) = 1 - \gamma C$. Consumption is assumed to stay below the satiation level of utility, such that $C < \frac{1}{\gamma}$ almost surely.[14].

Next, Section 3.6.1 looks at a simple setting when agents do not adjust their savings levels in response to the risk-sharing policy, Section 3.6.2 introduces endogenous savings, and Section 3.6.3 summarizes the main mechanisms at play and clarifies the intuition behind the observed mechanism. Appendix 3.A.4 looks at a setup with a risky and risk-free asset as individuals can consume in their youth and retirement. The analytical

[14]Loosely speaking, the quadratic utility assumption can be seen as a second-order approximation of the expected utility of a more complex utility function (Levy and Markowtiz, 1979; Buccola, 1982; Sharpe, 2007). For details on the use of quadratic utility in portfolio choice models see Brandimarte (2006), Cerný (2009), and D'Amato and Galasso (2010) who use it within an IRS context with a political game determining the optimal level of risk sharing with voting.

expressions become very complex, so we examine further the case in detail in the numerical section of 3.7 where additional complexity is added.

3.6.1 Exogenous Savings

Assume now that the savings are fixed to some $\bar{S}$ where $\bar{S} \in (0, Y)$. All shocks imported in the young age endowment wealth through the transfer scheme are completely absorbed through the consumption of the young. Assume for simplicity that only one risky asset is available, such that $T_t = \tau Y(\tilde{\mu} - \tilde{R}_t^x) = -\tau Y \tilde{\epsilon}_t^x$ where the risk-sharing parameter is τ. Optimal consumption defined in (3.18) then evolves as follows:

$$\begin{aligned} C_{y,t} &= Y - \overline{S} + \tau \tilde{\epsilon}_t^x Y \\ C_{o,t+1} &= \overline{S} \tilde{R}_{t+1}^x - \tau \tilde{\epsilon}_{t+1}^x Y \end{aligned} \tag{3.26}$$

Now, investments are fixed and do not react to changes in τ, so we have that $\mathbb{E}\frac{\partial C_{y,t}}{\partial \tau} = \mathbb{E}\frac{\partial C_{o,t+1}}{\partial \tau} = 0$, and the policymaker's optimality condition (3.24) simplifies. Furthermore, the policy surprise effect becomes irrelevant for the generation born in period zero, as neither they, nor by construction any future generation adjust their investments to the policy parameters. Also, the policy parameter τ is set once and for all before the realization of any shocks. As a result, the consumption streams for the old in period one and in any other period differ only in the realization of the shock. As a result, we can write condition (3.24) as

$$\begin{aligned} &\frac{\beta}{\delta}\mathbb{C}\text{ov}\left(u'(C_{o,t}), \frac{\partial C_{o,t}}{\partial \tau}\right) + \left[\frac{1}{1-\delta}\mathbb{C}\text{ov}\left(u'(C_{y,t}), \frac{\partial C_{y,t}}{\partial \tau}\right) + \frac{\beta}{1-\delta}\mathbb{C}\text{ov}\left(u'(C_{o,t}), \frac{\partial C_{o,t}}{\partial \tau}\right)\right] \\ &\qquad \equiv \underbrace{\delta \mathbb{C}\text{ov}\left(u'(C_{y,t}), \frac{\partial C_{y,t}}{\partial \tau}\right)}_{<0} + \underbrace{\beta \mathbb{C}\text{ov}\left(u'(C_{o,t}), \frac{\partial C_{o,t}}{\partial \tau}\right)}_{>0} \overset{!}{=} 0 \end{aligned} \tag{3.27}$$

The underbrackets show the signs of the covariance terms assuming[15] that $0 < \tau < \frac{\overline{S}}{Y}$. They hold for any utility function with a decreasing marginal utility of consumption. To illustrate why this is happening, imagine that there is a negative financial shock ($\epsilon_t < 0$).

[15] It can also be shown that in optimality τ cannot be negative when savings are positive, as then both covariances will be positive and the first-order condition would never hold. If $\tau > \bar{S}/Y$ on the other hand, both covariances are negative, and again the optimality condition cannot hold.

The IRS mechanism then transfers wealth from the young to the old. This has two effects. First, the marginal consumption of the old ($\frac{\partial C_{o,t}}{\partial \tau}$) increases and their marginal utility decreases for higher τ as they get compensated for the resource loss on their retirement savings. Second, the transfers induce a resource loss to the young as any compensation for the old is subtracted from their initial endowment, driving down the young's marginal consumption ($\frac{\partial C_{y,t}}{\partial \tau}$) and driving up their marginal utility. Since the transfers are linear, exactly the opposite effect occurs with a positive financial shock.

Overall, the IRS policy allows the old to trade negative retirement-wealth shocks with the currently young, who in turn in good times gain from the additional accumulated wealth of the old. The risk-sharing parameter τ drives the sizes of the trade-offs for each generation, so it needs to balance out the willingness of one generation to get protection in bad states of nature in lieu of smaller gain in good states with the willingness of the other generation to forego current consumption in bad states in lieu of higher consumption in good states. Overall, optimal τ needs to be set such that the two effects, as captured by the covariance terms, balance out.

In particular, for quadratic utility, we have

$$\mathbb{Cov}\left(u'(C_{y,t}), \frac{\partial C_{y,t}}{\partial \tau}\right) = \mathbb{Cov}(-\gamma\tau\tilde{\epsilon}^x_t, \tilde{\epsilon}^x_t) = -\gamma\tau\mathbb{Var}(\tilde{\epsilon}^x_t) \leq 0$$
$$\mathbb{Cov}\left(u'(C_{o,t+1}), \frac{\partial C_{o,t+1}}{\partial \tau}\right) = \mathbb{Cov}\left(-\gamma\left(\frac{\overline{S}}{Y} - \tau\right)\tilde{\epsilon}^x_{t+1}, -\tilde{\epsilon}^x_{t+1}\right) = \gamma\left(\frac{\overline{S}}{Y} - \tau\right)\mathbb{Var}\left(\tilde{\epsilon}^x_{t+1}\right) \geq 0$$

Substituting in (3.27), the asset return variance and risk preferences cancel out of the policymaker condition and do not play a role in determining optimality. The optimal proportion of the shock that will be shared across generations is proportional to the savings rate and depends on the discount rates:

$$\tau^* = \left(\frac{\beta}{\beta + \delta}\right)\frac{\overline{S}}{Y} \tag{3.28}$$

The more a generation values old-age relative to young-age utility (higher personal discount factor β), the higher the optimal level of IRS should be in order to allow generations to hedge the negative states of nature they could experience in retirement. Similarly, by construction a lower value for δ decreases the relative weight of the young in the welfare function (3.9) driving down the need for IRS.

3.6.2 Utility of Old-age Consumption Only

Now, assume that agents derive utility from old-age consumption only.[16] By construction, they will save all their young-age endowment and consume it when old. In contrast to the previous case, where young-age wealth shocks resulting from the transfers were fully absorbed by consumption, now the transfer shocks are fully absorbed by savings[17].

Formally, the indirect utility of consumption for each generation becomes

$$v(\tau, \tilde{\epsilon}_t^x) = \beta \mathbb{E}_t u_o(C_{o,t+1}) \tag{3.29}$$

where

$$C_{o,t+1} = S_t \tilde{R}_{t+1}^x - \tau Y \tilde{\epsilon}_{t+1}^x \tag{3.30}$$

For each generation born after implementation of the policy we have $S_t = Y + \tau Y \tilde{\epsilon}_t^x$, while in absence of anticipation effects generation zero has fixed savings $S_0 = Y$.

Substituting in the policymaker's optimizing condition (3.23) and simplifying we get:

$$\frac{\beta}{\delta} \mathbb{E}\left(u'(C_{o,1}) \cdot \frac{\partial C_{o,1}}{\partial \tau} \right) + \frac{\beta}{1-\delta} \mathbb{E}\left(u'(C_{o,t+1}) \cdot \frac{\partial C_{o,t}}{\partial \tau} \right) = 0$$

Simplifying further (Appendix (3.A.2) shows the derivation details), Y, γ and $\tilde{\sigma}^2$ cancel out, and we can solve for the optimal level of risk-sharing

$$\tau^* = \frac{1}{\delta \mathbb{E}\left((\tilde{R}_t^x)^2 \right) + 1} = \frac{1}{\delta \mathbb{E}\left(\tilde{\mu}^2 + \tilde{\sigma}^2 \right) + 1} \tag{3.31}$$

[16] This set-up is common in the pension literature where agents derive utility only from pension income and retirement consumption.

[17] A similar risk-sharing set up appears within the context of a political game for example in D'Amato and Galasso (2010); Ciurila and Romp (2015)

In contrast to the case with fixed savings, τ^* is now a decreasing function of the riskiness of the asset, quantified as the expected quadratic variation in the return of the asset after the risk of liquidity costs is covered. The discount rates do not appear here in the optimal term, as young-age consumption is not modeled and risk is not discounted over the lifetime of individuals.

The expected quadratic variation $\mathbb{E}\left((\tilde{R}_t^x)^2\right)$ is positively related to the probability to trade and negatively related to the liquidity cost, as shown in Section 3.3.2. Then it follows that as the illiquidity friction becomes more severe, the quadratic variations in the asset returns become smaller and an increase in the risk-sharing parameter is needed to ensure that enough risk is transferred across generations. Formally, we can show that

$$\frac{\partial \tau^*}{\partial p} < 0, \frac{\partial \tau^*}{\partial \bar{l}} > 0 \tag{3.32}$$

3.6.3 Risk Pooling vs. Compounding of Risk

Now, we explore the relationship between the level of IRS and the variance of the risky asset. We decompose the relationship into two counterbalancing effects. First, in aggregate, IRS expands the pool of people who can participate in a shock occurring in a given period by including the individuals who are not economically active in the risk-bearing pool. Second, it extends the time window over which individuals bear risk by forcing them to participate earlier in their lifetime in the realization of financial shocks, which are otherwise only affecting the wealth of the old. We know that the first effect enhances the overall risk-bearing capacity of the population. The second effect, however, on its own produces a welfare loss for the young and needs to be explored further.

It is well-known that the uncertainty in a risky asset's returns does not diversify with time, and that longer investment horizons do not lead to lower variance of the accumulated wealth. In essence, this is the fallacy of time diversification which states that aggregating shocks over time increases their cumulative variability (Samuelson, 1963; Ross, 1999). Gordon and Varian (1988) also refer to the *compounding of lotteries* and suggest that the time accumulation of variance, due to the random shocks transferred from one generation to the next, embeds a cost in the IRS mechanism. In our setting, as well, compounding of uncertainty makes it expensive in utility terms to transfer risks over to the young as the risk they will start bearing when young will accumulate through their savings and will lead to higher consumption variability in retirement. It is then natural that the larger the variance of the savings portfolio is, the more costly it is to transfer risk across generations.

Consider again the set-up of Section 3.6.2. The multiplicative shock which will appear in the old-age consumption equation (3.30) is the key driver of the inverse relationship between the level of optimal IRS and the magnitude of the asset variance. To illustrate, assume for the sake of argument that the savings asset is liquid. Then, old-age consumption is

$$C_{o,t+1} = S_t R_{t+1} + T_{t+1} = Y\left((1+\tau\epsilon_t)R_{t+1} - \tau\epsilon_{t+1}\right)$$

This means that the policy ($\tau > 0$) imports additional uncertainty into old-age consumption, having made the young-age starting wealth uncertain as $Y(1+\tau\epsilon_t)$. In old age, the variance of endowment is translated into additional variance of savings and in old age gets magnified by the variance of the accumulated asset return. Old-age wealth as a fraction of Y then becomes $(1+\tau\epsilon_t)R_{t+1} = \mu + \epsilon_{t+1} + \tau\mu\epsilon_t + \tau\epsilon_t\epsilon_{t+1}$. As a result, the variance of consumption becomes function of the risk sharing parameter.

We can decompose the total variance of old age consumption with IRS into the following two effects[18]:

$$\begin{aligned}\mathbb{Var}(C_{o,t+1}) &= \mathbb{Var}\left(S_t R_{t+1} - Y\tau\epsilon_{t+1}\right) \\ &= Y^2\left(\underbrace{\sigma^2(1-\tau)^2}_{\text{Pooling Effect}} + \underbrace{\mu^2\sigma^2\tau^2 + \tau^2\sigma^4}_{\text{Risk-Compoinding Effect}}\right)\end{aligned} \quad (3.33)$$

First of all, the IRS mechanism expands the pool of individuals that can participate in the risk, which is about to materialize in a given period. So, the old will bear only $1-\tau$ proportion of the risk occurring in their retirement, while the rest is transferred to the newly born. This drives the *pooling effect* of the variance.

The young consume their wealth when they retire. The risk that they have to participate in while young, proportional to τ, is then reinvested until retirement, when a new shock occurs and this amplifies the initial one. This results in the *risk-compounding effect.*

[18] See Appendix (3.A.3) for details on the derivation

As the variance of consumption is convex in the risk-sharing parameter, there will be a point after which the uncertainty imported in old age consumption through risk sharing will dominate over the reduction in uncertainty coming from the increased risk-bearing pool. The policymaker's problem then is to set τ such that the two effects are balanced. The variance of consumption is thus minimized for $\tau^* = \frac{1}{1+\sigma^2+\mu^2}$. Note that as the variance of the asset increases, the risk-compounding effect, having a higher order, dominates over the pooling effect, and to achieve optimal variance of consumption, the policymaker needs to reduce the level of risk sharing.

The individuals born in period zero, before the policy is implemented, get to experience solely the pooling effect, without being subject to the compounding cost. In particular, the variance of their consumption is

$$\mathbb{V}\mathrm{ar}(C_{o,1}) = Y^2\mathbb{V}\mathrm{ar}(R_1 - \tau\epsilon_1) = Y^2\mathbb{V}\mathrm{ar}(\mu + (1-\tau)\epsilon_1) = Y^2 \underbrace{(1-\tau)^2\sigma^2}_{\text{Pooling benefit}} \tag{3.34}$$

These individuals are thus privileged from an ex-ante point of view, as they benefit for each $\tau \in (0, 1]$ and in optimality will want to have it set to unity. With no risk sharing, they bear the full risk of their old-age consumption, and with complete sharing their old-age consumption risk is reduced to zero and any negative shocks are shifted to the newly born young generation at period one. For this generation, the risk reduction occurs free of the compounding cost, that other generations need to bear.

Note that sometimes the literature uses time-additive shocks to illustrates the pooling benefits of risk sharing. Crucially, this misses the time aggregation component of risk. To illustrate how additive shocks can mislead, assume first that the young bear τ portion of the shock while the old take proportion $1-\tau$. Additive shocks would imply that the young save $S_t = Y + \tau\epsilon_t Y$ and the old consume $C_{o,t} = S_{t-1} + (1-\tau)\epsilon_t Y = Y(1+\tau\epsilon_{t-1} + (1-\tau)\epsilon_t)$. It is clear that with *i.i.d.* shocks, we then have $\mathbb{V}\mathrm{ar}(C_{o,t}) = ((1-\tau)^2 + (\tau)^2)\,\sigma^2 Y^2$. The sharing parameter which minimizes the variance of consumption then is $\tau^*_{add.} = 1/2$ and it is clearly independent of the variance of the asset returns.

From that point of view, the argument can be made that averaging n independent shocks additive shocks, such that each generation gets a portion $\frac{1}{n}$ of each shock, leads to $\mathbb{Var}\left(\frac{1}{n}\sum_{i=0}^{n}\epsilon_{t-i}\right) = \frac{1}{n}\sigma^2$, and $n \to \infty$, the shocks will diversify away. Using this argument within the context of risk sharing can mislead that splitting shocks over many generations can make the risk disappear. When aggregating stochastic *i.i.d.* returns over time, however, which is in a way done by multiplying out the gross returns over time, the variance of the accumulated return will grow linearly with time:

$$\mathbb{Var}(R_t \cdot R_{t-1} \dots R_{t-n}) = \mathbb{Var}(\epsilon_t \cdot \epsilon_{t-1} \dots \epsilon_{t-n}) = \mathbb{E}\left(\epsilon_t^2 \cdot \epsilon_{t-1}^2 \dots \epsilon_{t-n}^2\right) + \mathbb{E}(\epsilon_t) \dots \mathbb{E}(\epsilon_{t-n}) = n\sigma^2$$

3.7 Quantitative Evaluation and Welfare Analysis with CRRA Utility

3.7.1 Set-up, Parameters and Initial Conditions

Now, assume that there are three assets available for investment: a risk-free liquid, a risky liquid, and a risky illiquid asset. The risky assets' gross returns follow a log-normal distribution, such that $R_t = \begin{bmatrix} R_t^s \\ R_t^x \end{bmatrix}$ where $\log(R_t) \sim N(\mu, \Sigma)$ with $\mu = \begin{bmatrix} \mu_s \\ \mu_x \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_x \\ \rho\sigma_s\sigma_x & \sigma_x^2 \end{bmatrix}$ and returns are scaled over a thirty-year holding period in line with the two-period OLG setting.

The risk-free asset is calibrated to the expected return of medium-term world government bonds. The risky liquid asset matches global equity's risk and returns characteristics. The risky illiquid asset is calibrated to match the properties of a portfolio of representative illiquid asset classes, where the weights are based on the relative sizes in a typical pension fund of private loans, equity, hedge funds, real estate and infrastructure holdings. Aggregated data on pension fund allocations are gathered from OECD (2019) for global funds and (PensionsEurope, 2018) for European funds. The data is summarized in Table 3.1. Expected asset returns, volatility, and correlations across the asset classes are based on the long term capital market forecasts in JP Morgan (2020). The costless trading probability (3.1) follows from a Poisson specification as the probability of having at least one trading opportunity during a time period Δt, such that[19]:

$$p = 1 - \mathrm{e}^{-\eta \Delta t} \tag{3.35}$$

Note that in this set-up $(1/\eta)$ is the average time one needs to wait for a costless trading opportunity to arrive. In calibrating η, we rely on data from Ang et al. (2014), who provide estimates of the average holding times and turnover of the illiquid asset classes considered here. Table 3.1 summarizes the data and the corresponding probability estimates for the variety of illiquid assets considered here. Further, Δt represents the period over which individuals would seek to sell their illiquid holdings. We assume five years as an approximation of the time before or after retirement during which individuals liquidate their asset holdings in order to fund retirement consumption. Overall, based on these considerations and based on the data, we assume $p = .8$ for the base scenario as a reasonable ballpark figure corresponding to the representative illiquid asset considered here.

The liquidity cost parameter $\bar{l}$ is calibrated through the average trading discount on the Net Asset Value (NAV) which needs to be accepted when selling private investments on the secondary market. Nadauld et al. (2019) explore the secondary market for private equity funds and find that they trade on average at a discount of 13.8%. The number varies significantly, typically in a range between 5% and 30%, depending on market conditions, fund age, and fund type (e.g., Buyout, Venture Capital, Real Estate). We assume $\bar{l} = 20\%$ as a ballpark base figure. Further on, we explore the solution's sensitivity by varying both p and $\bar{l}$ parameters. We set $\gamma = 5$ and $\beta = \delta = e^{-.03 \cdot 30}$.

[19]The costless trading opportunity for the illiquid asset arrives as a Poisson event with an expectation η. The probability of having n such trading opportunities over a period of time Δt then is

$$P(n) = \mathrm{e}^{-\eta \Delta t} \frac{(\eta \Delta t)^n}{n!}, \quad n = 0, 1, ...$$

Denoting N_t as the cumulative number of trading events which occurred up to time t, the probability p of being able to trade costlessly in the illiquid asset at least once over a given period can then be derived as

$$p = P(N_{t+\Delta t} - N_t \geq 1) = 1 - P(N_{t+\Delta t} - N_t = 0) = 1 - \mathrm{e}^{-\eta \Delta t}$$

Formally, this is equivalent to modeling liquidity events as a Poisson process, as in e.g. Ang et al. (2014).

Table 3.1: Parameters Calibration

	Holding Time	p	Weight	μ	σ	ρ
Liquid Risk-Free Asset						
- Mid-Term Gov Bonds			100%	0.002	-	
Liquid Risky Asset				-	-	1.000
- Global Equity			100%	0.061	0.156	
Illiquid Risky Asset				0.049	0.120	0.586
- Hedge Funds	1 - 2	0.92 - 0.99	16%	0.030	0.074	0.730
- Private Equity	4	0.71	23%	0.078	0.202	0.800
- Institutional Real Estate	8 - 10	0.39	39%	0.046	0.111	0.500
- Institutional Infrastructure	50 - 60	0.08	14%	0.047	0.105	0.550
- Private Loans	-	-	8%	0.017	0.045	0.150

This table shows the calibrated values for the return and risk properties of the three assets in the model. The average number of years it takes to trade on one of the illiquid assets is used as input to calculate the probability p that the asset can be sold over a five-year period. The Poisson probability formula links the two. The figures on μ and σ indicate the expected return and standard deviation respectively of the asset class, and ρ indicates the corresponding correlation to equity global. Data is used from JP Morgan (2020) adjusted for a fixed inflation rate of 2%. The weights indicate the asset proportions within the illiquid portfolio of a typical global pension fund as in OECD (2019) and they are used to aggregate the basket of illiquid assets into one representative illiquid asset.

The young and the old are assumed to have the same CRRA utility of consumption, where $\gamma > 1$ is the usual parameter of risk aversion:

$$u(C) \equiv \frac{1}{1-\gamma} C^{1-\gamma}, \quad C > 0$$

In the planner case, after the system is initiated (for $t \geq 1$), due to (3.14) in optimality we have:

$$C^*_{y,t} = \left(\frac{\beta}{\delta}\right)^{-\frac{1}{\gamma}} C^*_{o,t}$$

As a result, one only needs to solve for the youth age consumption in the planner case, and optimal old-age consumption will be proportional.

Using the social welfare formulation of (3.9), we can translate the utility units into Certainty Equivalent Consumption (CEC) units, where the CEC measures the stream of fixed risk-free future consumption that the current and all future generations would be willing to accept for the stream of risky consumption leaving the agent indifferent utility-wise between the two options. In the case of a CRRA utility, as Appendix (3.A.5) shows, the CEC for the whole population can then be written as:

$$CEC_0 = \left[(1-\delta)(1-\gamma)\frac{\delta}{\beta+\delta}V_0\right]^{\frac{1}{1-\gamma}} \tag{3.36}$$

where V_0 is a function of the starting wealth values $V_0 = V(W_0, X_0)$ in the case of the planner problem, while in the case of intergenerational transfers, it is a function of the policy instruments such that $V_0 = V(\tau^*)$.

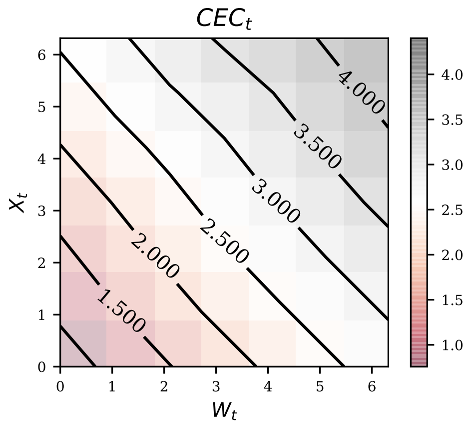
In the decentralized cases, either with or without risk sharing, all generations but generation zero are identical ex-ante and start with the same endowment and the same wealth, so this eliminates the need to establish initial conditions. In the autarky case, welfare is evaluated through equation (3.20), with all transfers set to zero, while in the case with transfers, the policymaker picks the optimal risk-sharing parameters.

In the planner economy, the initial conditions matter. To be consistent with the policymaker problem of section 3.5, we assume that the economy exists initially in a decentralized no-risk-sharing regime. Individuals from generation zero then freely pick their asset allocation and savings level ignorant of the following policy implementation. Between periods zero and one, the planner appropriates the accumulated savings in the economy. After period one, she starts optimally distributing consumption between the young and the old and starts picking the appropriate asset mix and the aggregate savings level, as indicated in section 3.7.2. The accumulated up to period one liquid and illiquid wealth then provide the initial conditions on which the planner relies. Consequently, we evaluate the welfare as the probability-weighted value-function level at accumulated liquid and illiquid wealth in period one, assuming that further on the planner is picking optimal values for the decision variables.

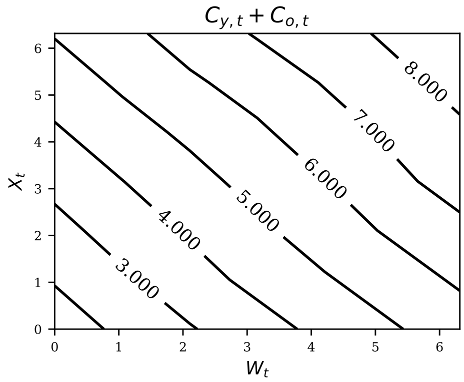
3.7.2 Planner Solution

First, in the planner economy, we look at how aggregate welfare, consumption and investment behave as functions of the stochastic state variables X_t and W_t. Figures (3.4a) and (3.4b) show that welfare and aggregate consumption, respectively, unambiguously increase if more liquid or illiquid wealth becomes available. The increase with respect to illiquid wealth follows from the fact that for $\bar{l} < 1$ and $p > 0$ the illiquidity friction

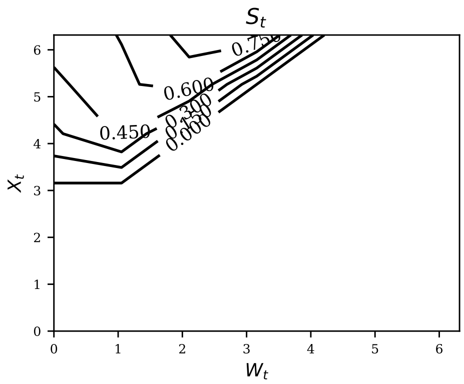
Figure 3.4: Planner Optimal Policies



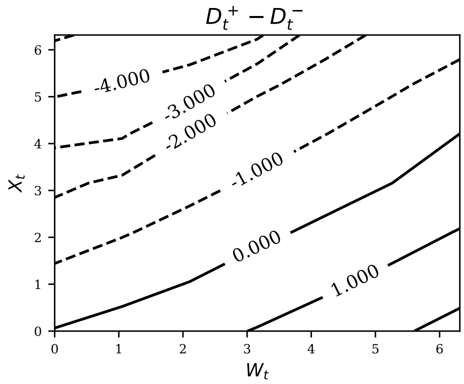
(a) Aggregate Welfare



(b) Aggregate Consumption



(c) Liquid Risky Investments



(d) Illiquid Wealth Withdrawals

This set of figures shows the solution of the planner dynamic programming problem. The charts show the value function (presented in CEC units), optimal consumption, optimal risky liquid investments, and the withdrawals (dashed, negative lines) and investments (solid, positive lines) into illiquid wealth, respectively. Borrowing is constrained with $L = 0$ in (3.11). The optimal solutions are presented as functions of the two stochastic state variables, W_t and X_t. In a dynamic setting, optimal consumption and investment for the coming period are set by knowing the two-state variables at the beginning of the period.

is only setting a potential cost to withdrawing funds for consumption but is not barring withdrawals entirely. Note, however, that due to the risk of incurring withdrawal costs, a proportional increase in liquid or illiquid wealth does not lead to an equal increase in welfare, so the slope of the iso-lines is not unity.

Figure 3.4c shows the optimal levels of liquid risky asset investments S_t given the beginning of period illiquid and liquid wealth. First, it can be seen that if the current liquid wealth is already high (the right half of the chart), it is not optimal to allocate resources to the liquid risky assets, as the regular endowment income provides enough liquidity for the coming period, and it becomes more profitable to allocate to the illiquid asset and reap the illiquidity premium embedded in it. At the same time, if the start-of-period illiquid wealth is too high (in the upper half of the chart) it becomes optimal to shift resources from the illiquid risky asset towards liquid risky assets, so S_t starts picking up, again to secure diversification.

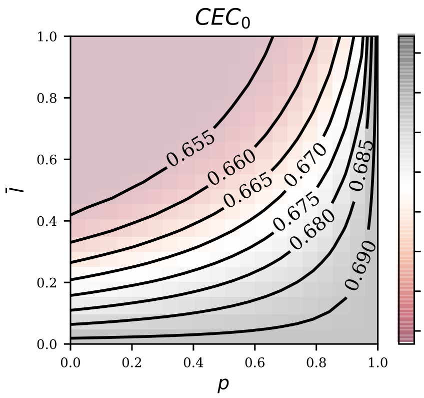
Figure 3.4d shows the total amount of withdrawals D_t^- (if negative) and investments D_t^+ (if positive) in illiquid wealth. This chart complements the findings so far. If a certain proportion of illiquid-to-liquid wealth is breached, indicated with the zero diagonal line, the planner will reallocate, withdrawing or depositing into illiquid savings in order to avoid over-concentrated holdings in one type of wealth. That relocation will either be consumed or invested in risk-free liquid holdings or risky-liquid holdings.

3.7.3 No Risk Sharing Solution

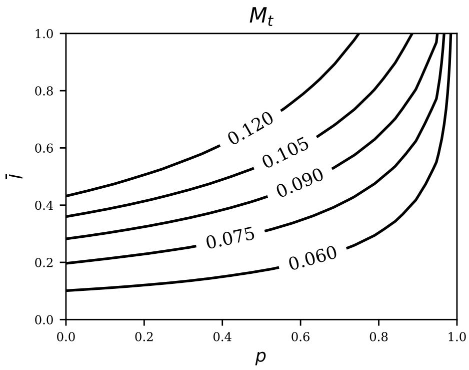
Now we look at an economy where generations themselves optimize the asset mix in their portfolio. First, they cannot share risks with other generations. We keep in line with the literature by referring to this as *autarky*, taking into account the fact that agents consume purely out of their endowments and do not have the technology to consume out of the endowment income of other generations (Beetsma and Romp, 2016; Gollier, 2008). Formally, the solution follows from the individuals' savings problem (3.15) where all transfers T_t are set to zero and where the policymaker does not play any role.

The set of figures (3.5) shows how welfare and investments are affected by illiquidity. It is clear that as the liquidity friction increases (either p increases or $\bar{l}$ decreases), the

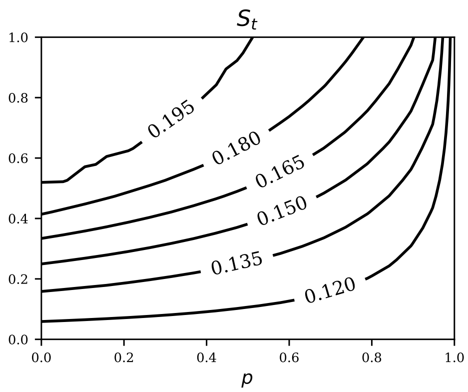
Figure 3.5: Illiquidity without Risk Sharing



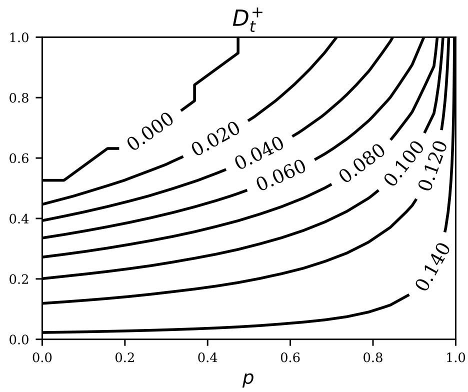
(a) Aggregate Welfare



(b) Risk-Free Investments



(c) Liquid Risky Investments



(d) Illiquid Risky Investments

This set of contour plots show the effect of the liquidity friction in the base case on the invested amounts in each of the assets. The lifetime CEC measures the certainty equivalent consumption of individuals over the two periods of their lifetime.

welfare in the economy (Figure 3.5a) goes down monotonically. This is driven by several factors. First, investment in the illiquid asset decreases (Figure 3.5d). Investment in the risky liquid asset increases (Figure 3.5c) to compensate, but overall there is a drop in total risky asset holdings as illiquidity rises[20]. Precautionary risk-free savings increase (Figure 3.5b). In this case, as the illiquidity friction increases, agents suffer from reduced diversification in their investment mix. At the same time, their capacity to bear market risk is reduced and they are not in a position to exploit fully the market risk premia.

3.7.4 Risk Sharing Transfers

Agents now solve the allocation-savings problem in a decentralized economy, where a policymaker administers the risk-sharing instruments by optimizing aggregate ex-ante welfare. The set of figures (3.6) shows the effect that risk sharing has on welfare and investment. The parameters τ_s and τ_x stand for sharing in the liquid and in the illiquid risky asset respectively. Individuals are allowed to borrow in their youth.

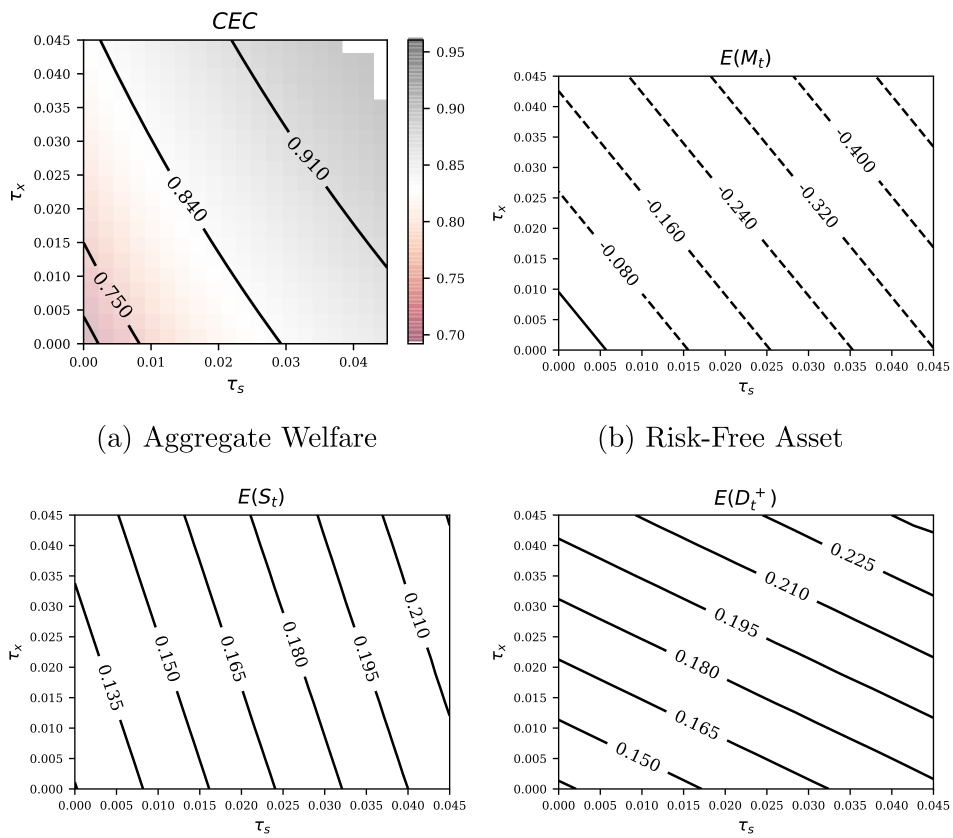
Figure 3.6a shows the aggregate welfare in the economy in period one as a function of the policy instruments. Welfare is measured in CEC units in line with (3.36). It is growing for the most part with the degree of risk-sharing, but it suddenly cuts off to zero at the chart's upper-right edge. This occurs when the risk-sharing parameters are set too high, which in combination with a large negative shock in either of the risky assets has the potential to produce scenarios where the endowment income of the young after transfers to the old leave nothing for young-age consumption and thus marginal utility becomes infinite.

Figures (3.6b), (3.6c), (3.6d) illustrate how the allocation to each of the three assets changes when either of the risk sharing instruments is varied. The charts, thus, show the expected investments over time, since (as shown in Section 3.5) risk sharing makes the amount available for investment random by being dependent on the realized financial shocks of the current period.

[20]For example, in a high liquidity case (bottom right corner of each chart, where $\bar{l}$ is low and p is high) total risky investments are around .26. In the low-liquidity case (the upper right corner of each chart), total liquid and illiquid risky investments drop to about .195

In line with studies done before (Gollier, 2008; Shiller, 1999; Campbell and Nosbusch, 2007), the increase in risk sharing between generations enhances the ability of individuals to bear investment risk. The average amounts allocated to the liquid (Figure 3.6c) and the illiquid asset (Figure 3.6d) increase with the degree of IRS, while the investment in the risk-free asset decreases (Figure 3.6b) and even becomes negative, as the individual leverages up by borrowing. In line with the diversification principle, the increase in risk sharing even in one of the assets increases the ability to bear risk in both of the risky assets. For example, an increase in τ_x from .009 to .016 increases the average optimal holdings in the liquid asset from .135 to .15. Still, an equivalent increase in one sharing parameter favors more the risky asset that it targets.

Figure 3.6: Illiquidity with Risk-Sharing Transfers



(a) Aggregate Welfare

(b) Risk-Free Asset

(c) Risky Liquid Asset

(d) Illiquid Risky Asset

This plot shows the effect of varying the risk-sharing parameters on welfare and the average levels of the optimal asset holdings. The white space in the upper-right edge of Figure 3.6a shows the values of τ_s and τ_x where the CEC cuts off to zero as scenarios appear in which transfers become larger than the endowment of the young.

Table 3.2 presents the values for consumption and optimal investment in an economy under optimal risk sharing. Two cases are considered for robustness: when agents are either able to borrow and when borrowing is restricted. As observed before, compared to autarky, risk sharing effectively increases the capacity to invest in the risky assets in both cases, even though the effect is more substantial without a borrowing constraint. Even in the case of restricted borrowing, the portfolio is on average fully invested in risky assets.

Going forward, we quantify the degree of welfare improvement in line with Beetsma and Romp (2016):

$$\left(\frac{CEC^i}{CEC^a} - 1\right) \cdot 100\%$$

where CEC^i stands either for the specific policy to be evaluated, and CEC^a stands for the aggregate welfare in the benchmark autarky economy.

Table 3.3 compares the welfare improvement from the two IRS mechanisms considered so far - from the intergenerational transfers and from the planner case. Naturally, the welfare achieved through a planner is higher, as a policymaker can share risk between two generations only, while a planner can share risk with infinitely many future generations. In addition, policymaker is restricted to apply only linear transfers of risk. Still, whether borrowing is restricted or not, the model projects that a policymaker is already able to realize a lot of the benefits possible through risk sharing, and the difference to what is achievable by the planner is not large. If borrowing is possible, the model projects a welfare improvement of 36% in the decentralized case with transfers vs. 48% for in the policymaker case. With borrowing, the relation is 17% vs. 21%.

A question of interest is to what degree the introduction of risk-sharing between the young and the old can help lower the welfare losses from increased asset risk. Next, we look at how changes in the risk profile of the illiquid asset, either in terms of increased variance or increased liquidity cost, affect welfare.

First, holding investment and risk-sharing fixed as in Figure 3.7a, we can see that for increased risk, there is, at least initially, more potential for welfare improvement in transferring part of it from the old to the young. The linear intergenerational transfers relocate wealth from the old to the young in states of nature where the marginal utility of the additional consumption of the old is low, and when asset returns are above their

Table 3.2: Risk Sharing vs. Base Case

	No Risk Sharing	Risk Sharing with borrowing	Risk Sharing without borrowing
$\mathbb{E}Cy$	0.694	1.008	0.805
$\mathbb{E}Co$	1.359	1.374	1.379
$\mathbb{E}M$	0.054	-0.446	0.000
$\mathbb{E}S$	0.119	0.224	0.110
$\mathbb{E}D^+$	0.133	0.214	0.085
τ_s^*		0.050	0.018
τ_x^*		0.021	0.030

Note. This table compares the policymaker risk-sharing and the autarky (no risk sharing) solution, looking at the case where a lower limit of zero on borrowing (investment M_t) is applied and when no such limit is applied. There is an internal optimal solution in the no sharing solution, so a constraint on borrowing does not change anything. The expected values are calculated for generations born in $t > 0$.

Table 3.3: Risk Sharing

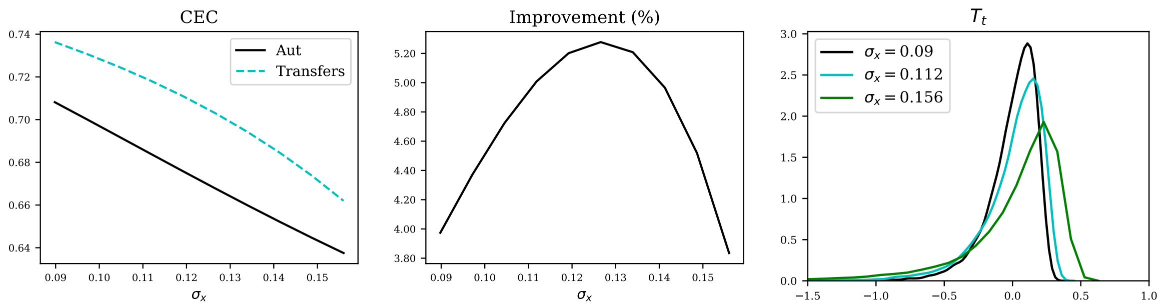
	No Risk Sharing	Policymaker	Planner
	With borrowing		
CEC	0.687	0.932	1.014
Improvement	-	36%	48%
	Without borrowing		
CEC	0.687	0.805	0.830
Improvement	-	17%	21%

Note. This table shows the welfare improvement over autarky when risk sharing is introduced either through a policymaker or through a planner approach. In the former case, individuals are allowed to borrow. In the latter case, borrowing is allowed only up to a level comparable to the one observed in the policymaker optimal solution.

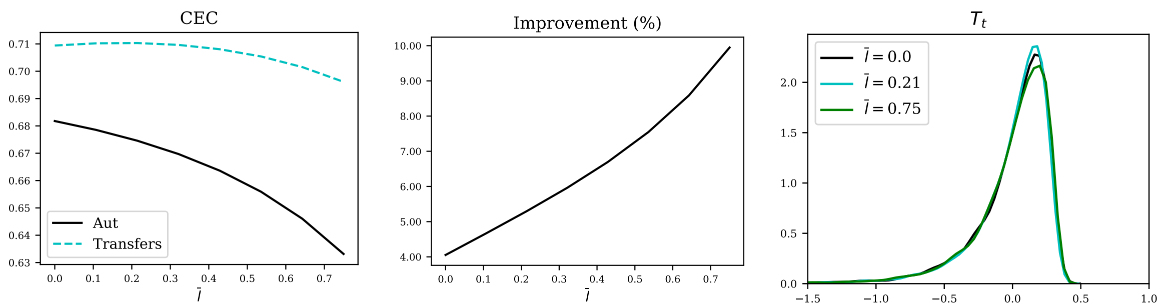
expected values. Furthermore, wealth is relocated in the opposite direction, from the young to the old, in states of nature when asset returns are low, consequently savings of the old are depreciated, and the old's marginal utility is currently high.

However, when the variance of the investment portfolio and consequently the variance of transfers increases, at some point the transfers may become too much for the young to bear. As scenarios of larger transfers from the young become more frequent, this may leave them with too little endowment to cover their young-age consumption, causing more scenarios where young-age utility drops more than the corresponding increase in the utility of the old, and thus also lowering the overall expected lifetime utility of the generation. As a result, at some point, this second push-back effect prevails and the welfare improvement starts declining with further variance. If τ_x is set too large, it may even happen that the risk-sharing policy comes at a disadvantage compared to the initial situation. Allowing τ_x to be adjusted avoids the problem of transferring too much risk from one generation to another, as shown in Figures (3.8).

Figure 3.7: Fixed Allocation and Risk Sharing



(a) Increased Asset Variance

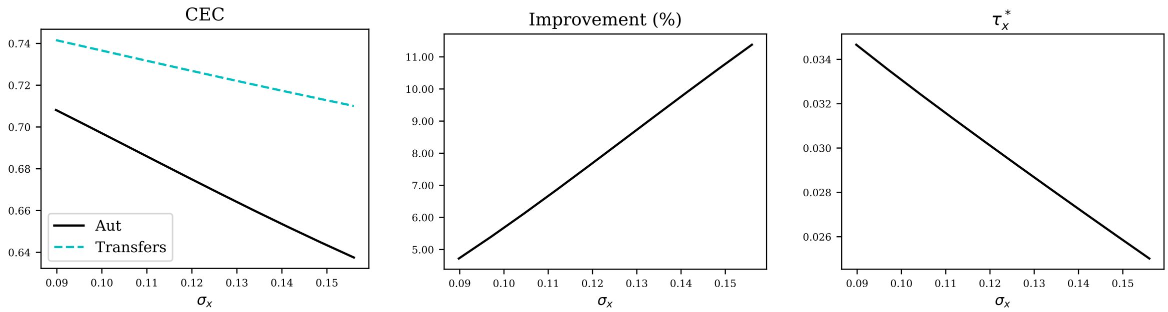


(b) Increased Liquidity Cost

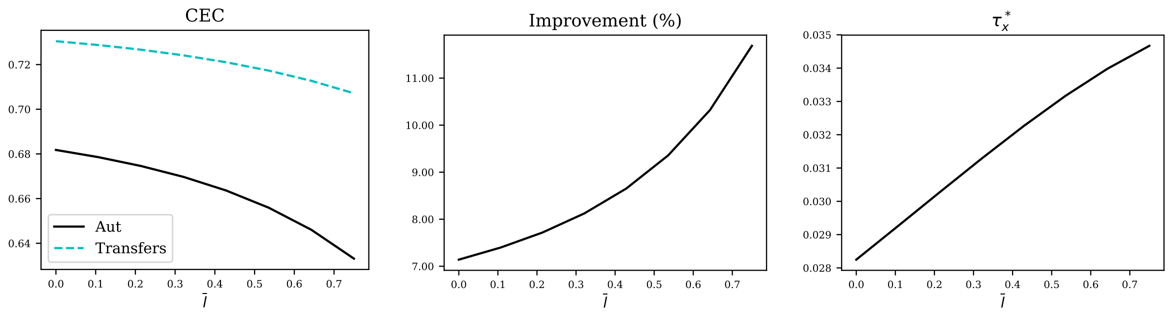
Note. This set of charts shows the effect of varying the volatility on the illiquid asset and respectively its liquidity cost. The investments are fixed in all states of nature to $M_t = S_t = D_t^+ = .1$, and risk sharing is fixed such that $\tau_s = 0$, $\tau_x = .05$.

Note however that increased liquidity risk does not increase the variance of transfers (3.7b). This is in line with the way illiquidity affects asset returns as shown in Section 3.3.2: higher illiquidity lowers the variance of wealth, thus lowering also the chance that more variance is transferred to the young than they could bear. As a result, improvement in welfare is monotonously increasing with rising illiquidity cost, in contrast to the effect observed for higher variance.

Figure 3.8: Fixed Allocation and Optimized Risk Sharing



(a) Increased Asset Variance



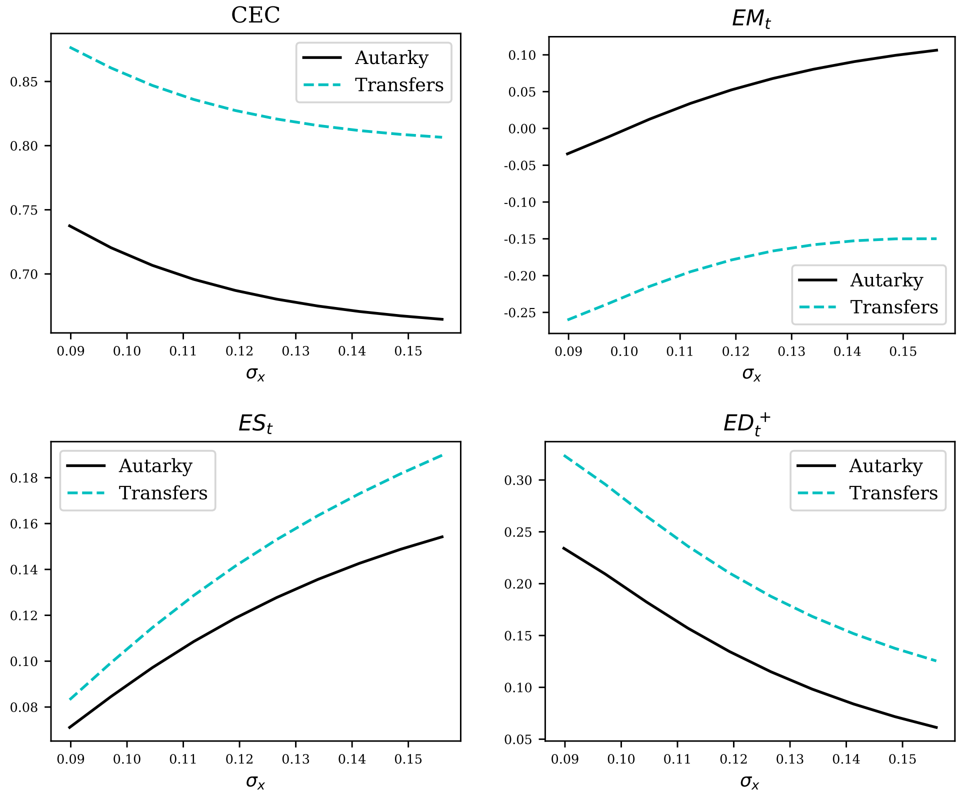
(b) Increased Liquidity Cost

Note. This set of charts shows the effect of varying the volatility on the illiquid asset and respectively its liquidity cost. The investments are fixed in all states of nature to $M_t = S_t = D_t^+ = .1$, and risk sharing for the illiquid asset is optimized by the policymaker, while $\tau_s = 0$.

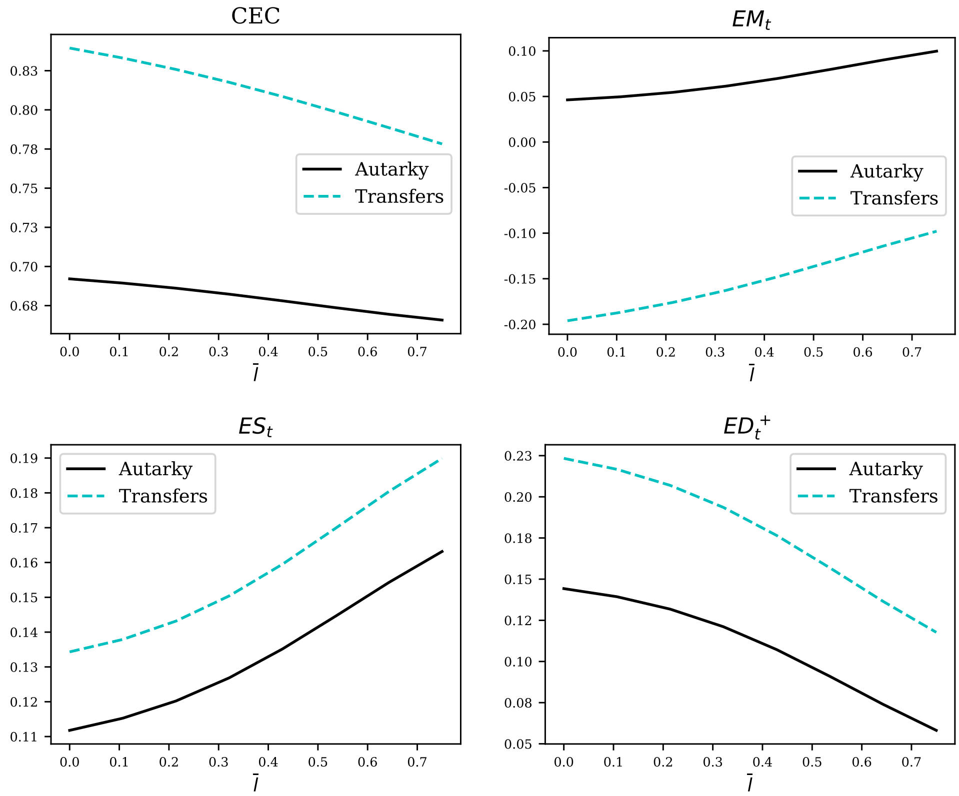
Second, as the risk of the illiquid asset increases, an investment substitution effect occurs, where individuals reduce holdings in that asset and increase their holdings in the risky liquid asset. This is clearly seen in Figures (3.9) where allocations are optimized while the risk sharing instruments stay fixed. The implementation of risk-sharing transfers shifts up the investments in each of the two risky assets, the increase being financed through lower holdings in the risk-free asset.

The plots in Figure 3.10 combine the effects of individuals optimizing their asset holdings and the policymaker simultaneously setting the optimal degree of risk sharing.

Figure 3.9: Optimized Allocation and Fixed Risk Sharing



(a) Increased Asset Variance



(b) Increased Liquidity Cost

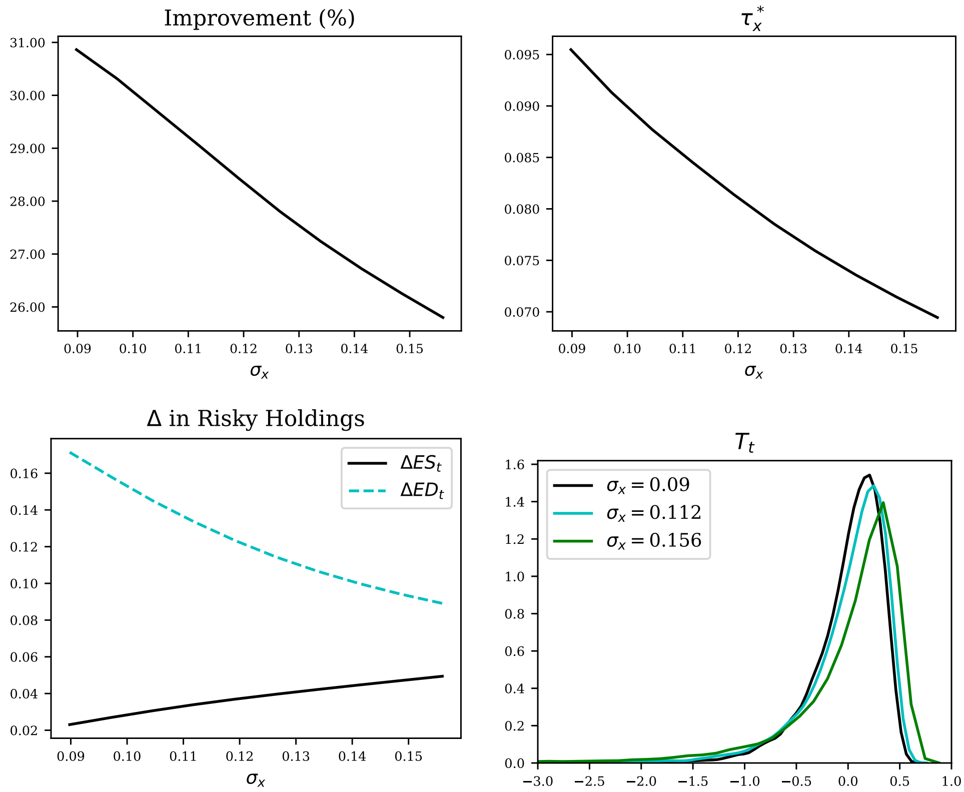
Note. This set of charts shows the effect of varying the volatility on the illiquid asset and respectively its liquidity cost. The investments are determined based on the individual optimality conditions while the risk sharing is fixed such that $\tau_s = 0$, $\tau_x = .05$.

First, as variance increases, the amount of optimal risk-sharing τ_x^* decreases to stabilize the variance of transfers. Second, risk sharing increases the desire of individuals to hold risky assets compared to autarky. Still, with higher risk, individuals hold less of the riskier illiquid asset and more of the substitute liquid asset. The investment in risky assets, in total, is diminishing with the increase in risk. In end effect, as the variance of the asset increases, the reduction of diversification in the savings portfolio held by the individuals results in lower welfare improvements as the variance of the illiquid asset increases. IRS cannot compensate for this effect, so overall, as the risk of the illiquid asset increases, the improvement in welfare is going down.

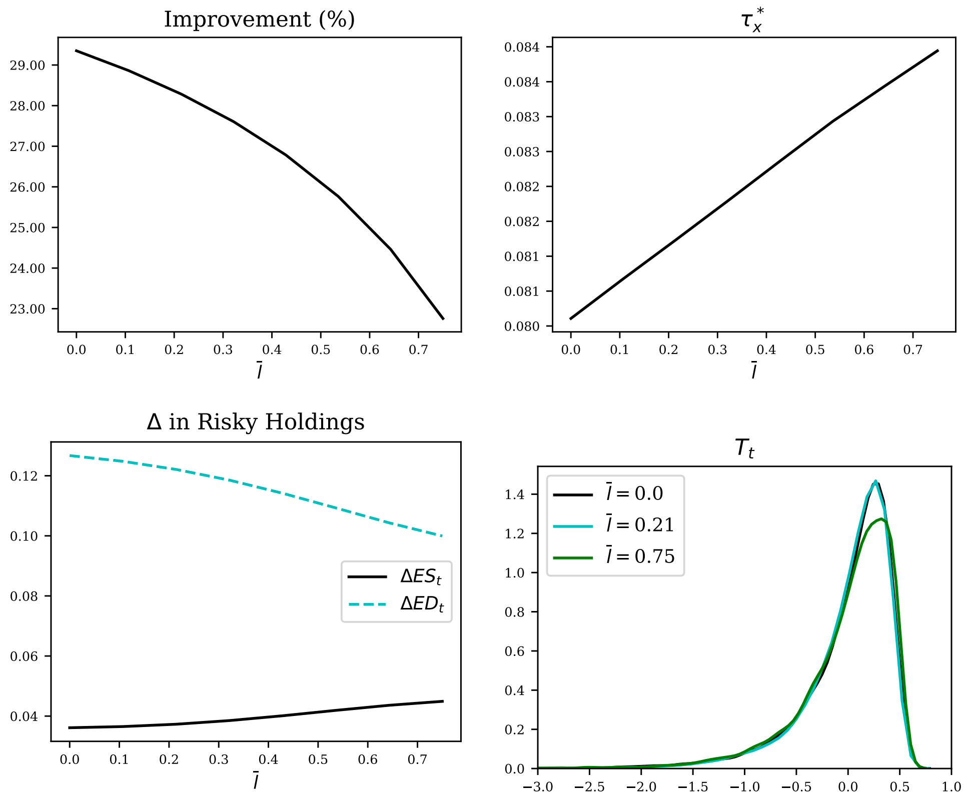
There is one notable difference in how an increase in σ_x versus an increase in $\bar{l}$ is affecting optimal risk sharing. Higher volatility justifies sharing a lower degree of IRS τ_x, while higher illiquidity calls for increased risk-sharing. The decrease in the variance of transfers caused by illiquidity needs to be compensated by a higher sensitivity of the transfers to the variance of the asset. This is in line with the arguments made earlier that with growing illiquidity, the variance of the savings portfolio decreases and sharing in the illiquid asset needs to be increased to compensate.

Figure 3.11 summarizes all cases considered and illustrates that the welfare increases with each relaxation of the constraints from the base case, defined as an individual with fixed investment shares without access to a risk sharing technology. The highest welfare occurs with case C5 when both IRS parameters are optimized by the policymaker while simultaneously individuals optimize their asset holdings.

Figure 3.10: The Effect of Risk on Welfare Improvements. Optimal Solution



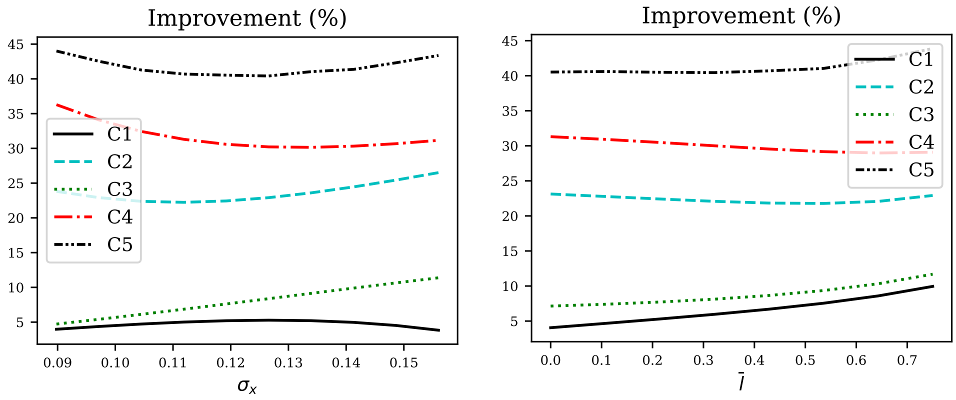
(a) Increased Asset Variance



(b) Increased Liquidity Cost

Note. This set of charts shows the effect of varying the volatility on the illiquid asset and respectively its liquidity cost. The level of inter-generational risk sharing is optimized over τ_x while $\tau_s = 0$. The model is parameterized for the base case, and in the first column the variance is varied, while in the second column the liquidation cost is varied.

Figure 3.11: Welfare Improvements vs. the Constrained No-Risk-Sharing Benchmark



Note. This set of charts shows the resulting welfare improvement vs. an autarky benchmark economy, in which agents cannot adjust their savings and $M_t = S_t = D_t^+ = 0.1$ in all states of nature. The lines represent the percentage CEC improvement resulting from switching to an intergenerational risk-sharing economy where C1: $\tau_s = 0$, $\tau_x = .05$, while the investments that individuals hold are fixed; C2: $\tau_s = 0$, $\tau_x = .05$, while individuals can adjust their asset holdings; C3: investments are fixed, $\tau_s = 0$ and τ_x is optimally adjusted; C4: investments are optimized, $\tau_s = 0$ and τ_x is optimized; C5: investments, τ_s and τ_x are optimized.

3.8 Conclusion

This paper examined the problem of optimally allocating risks across generations in the presence of market illiquidity within the asset mix of individuals' savings. We show in a stylized two-period overlapping generations framework that a contract of risk transfers between coexisting young and old cohorts enforced by a policymaker can improve welfare.

First, we show that optimal IRS is dependent on the variance of the savings portfolio of individuals. On one hand, the policymaker can create a mechanism that expands the pools of individuals who can bear the variance risk by including the young in the risk-sharing pool (pooling effect). On the other hand, introducing additional risk early on in individuals' lifetime savings accumulates higher variance in their old-age consumption (risk-compounding effect). The higher the variance, the more the latter dominates, and the lower the risk-sharing parameter should be.

From that point of view, illiquidity poses a friction to risk sharing between generations, as it reduces the variance over which the illiquid asset in the portfolio can be traded. To compensate for the loss of sensitivity of the IRS transfers to movements in the fair value of the asset, a policymaker needs to increase the level of risk sharing. In contrast, increases in the variance of a liquid asset, ceteris paribus, justify lower levels of risk sharing, as otherwise resources of the young will be destabilized, pushing them towards states of nature where the decline in young age utility of consumption is higher than the benefit of the elderly, thus lowering their lifetime utility and lowering the overall welfare in the aggregate economy.

Second, we find that risk sharing allows individuals to invest more into illiquid assets compared to the case when they are holding personal savings accounts without a mechanism to shift risks across generations. This is also in line with the literature which explores intergenerational risk-sharing mechanisms within funded pension plans (Gollier, 2008; Cui et al., 2011; Shiller, 1999).

The analytical results were justified by a quantitative welfare analysis with a realistic asset composition and utility specification. The framework highlights several policy-relevant implications. The tendency of pension funds to invest in illiquid asset classes, such as private equity, infrastructure projects, etc., may call for increased risk-sharing mechanisms between pension fund participants of different generations. Alternatively, increasing the liquidity of otherwise illiquid assets, for example with the development of attractive secondary markets for OTC traded assets, has the potential to also increase the benefits of intergenerational risk sharing, while also allowing for lower levels of risk sharing between cohorts.

3.A Appendix: Derivations

3.A.1 Planner Problem Derivations

Taking into account the constraints of the Bellman equation (3.13), we can write the optimization problem in Lagrangian form:

$$\mathcal{L} = \tilde{u}(C_{y,t}, C_{o,t}) + \delta \mathbb{E} V(W_{t+1}, X_{t+1}) - \sum_j \lambda_j g_j(W_t, X_t, C_{y,t}, C_{o,y}, S_t, D_t^+, D_t^-)$$

where each of the $g_j(.)$ functions, $j = 1, \dots 6$, represent one of the constraints written in a form such that $g_j(.) \leq 0$

$$\begin{aligned}
g_1(.) &= C_{y,t} + C_{o,t} + D_t^+ - D_t^-(1-\bar{l}) + S_t \mathbb{1} - W_t - Y - L;\\
g_2(.) &= D_t^- - D_t^+ - X_t;\\
g_3(.) &= D_t^- - X_t;\\
g_4(.) &= -D_t^-; g_5(.) = -D_t^+; g_6(.) = -S_t
\end{aligned}$$

and λ_j are the non-negative KKT multipliers subject to the standard interpretation as sensitivity of the optimal solution to relaxing the corresponding constraint. The standard first-order conditions apply together with complementary slackness and non-negativity:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= 0 \quad x \in \{C_{y,t}, C_{o,y}, S_t, D_t^+, D_t^-\}\\
\lambda_j g_j(.) &= 0\\
g_j(.) &\leq 0\\
\lambda_j &\leq 0, \quad j = 1, ...6
\end{aligned}$$

In particular, the first-order conditions with respect to consumption can be written as

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial C_{y,t}}: \quad & \frac{\partial \tilde{u}(C_{y,t}, C_{o,t})}{\partial C_{y,t}} = \delta R_f \mathbb{E} V_W(W_{t+1}, X_{t+1}) + \lambda_1 \\
& \implies u_y'(C_{y,t}) = \delta R_f \mathbb{E} V_W(W_{t+1}, X_{t+1}) + \lambda_1 \\
\frac{\partial \mathcal{L}}{\partial C_{o,t}}: \quad & \frac{\partial \tilde{u}(C_{y,t}, C_{o,t})}{\partial C_{o,t}} = \delta R_f \mathbb{E} V_W(W_{t+1}, X_{t+1}) + \lambda_1 \\
& \implies \frac{\beta}{\delta} u_o'(C_{o,t}) = \delta R_f \mathbb{E} V_W(W_{t+1}, X_{t+1}) + \lambda_1
\end{aligned} \tag{3.37}$$

As the optimal condition is symmetric w.r.t. the consumption of the young and the old, from (3.37) we have:

$$u_y'(C_{y,t}) = \frac{\beta}{\delta} u_o'(C_{o,t}) \tag{3.38}$$

Applying the Envelope Theorem on the constrained problem, we get:

$$\frac{\partial V(W_t, X_t)}{\partial W_t}: \quad V_W(W_t, X_t) = \delta R_f \mathbb{E} V_W(W_{t+1}, X_{t+1}) + \lambda_1 \tag{3.39}$$

and combining (3.39) with (3.37), we get this planner's version of the standard Ramsey equivalence of the marginal utility and the (liquid) wealth derivative of the value function:

$$\begin{aligned}
u_y'(C_{y,t}) &= V_W(W_t, X_t) \\
\frac{\beta}{\delta} u_o'(C_{o,t}) &= V_W(W_t, X_t)
\end{aligned} \tag{3.40}$$

Substituting (3.40) forward in (3.37), we can get the Euler relation, now corrected for a possible breach of the non-negativity borrowing constraint:

$$\begin{aligned}
u_y'(C_{y,t}) &= \delta R_f \mathbb{E} \tilde{u}'(C_{i,t+1}) + \lambda_1 \\
\frac{\beta}{\delta} u_o'(C_{o,t}) &= \delta R_f \mathbb{E} \tilde{u}'(C_{i,t+1}) + \lambda_1
\end{aligned} \tag{3.41}$$

The first-order condition w.r.t. the liquid risky investment in asset i is

$$\frac{\partial \mathcal{L}}{\partial S_t^i}: \mathbb{E} V_W(W_{t+1}, X_{t+1}) r_{t+1}^{s,i} - \frac{1}{\delta}(\lambda_1 - \lambda_6) = 0 \tag{3.42}$$

The first-order condition w.r.t. new investments in the illiquid asset D_t^+ and withdrawals from illiquid wealth D_t^-, respectively, can be written as

$$\frac{\partial \mathcal{L}}{\partial D_t^+} : R_f \mathbb{E} V_W(W_{t+1}, X_{t+1}) + \frac{1}{\delta}(\lambda_1 - \lambda_2 - \lambda_5) = \mathbb{E} V_X(W_{t+1}, X_{t+1}) R_{t+1}^x$$
$$\frac{\partial \mathcal{L}}{\partial D_t^-} : R_f \mathbb{E} V_W(W_{t+1}, X_{t+1})(1 - l_t) + \frac{1}{\delta}\left(\lambda_1(1-\bar{l}) - \lambda_2 - \lambda_3 + \lambda_4\right) = \mathbb{E} V_X(W_{t+1}, X_{t+1}) R_{t+1}^x$$

3.A.2 Utility of Old-Age Consumption

The generation born before implementation of the policy saves $S_0 = Y$, and consumes $C_{o,1} = Y\left(\tilde{R}_1 - \tau\tilde{\epsilon}_1\right)$ which implies the sensitivity of consumption to the risk-sharing parameter of $\frac{\partial C_{o,1}}{\partial \tau} = -Y\tilde{\epsilon}_1$.

Using the quadratic utility assumption, we get

$$\begin{aligned}
\mathbb{E}\left(u_o'(C_{o,1}) \cdot \frac{\partial C_{o,1}}{\partial \tau}\right) &= \mathbb{E}\left((1-\gamma C_1)(-Y\tilde{\epsilon}_1)\right) \\
&= \mathbb{E}\left(\left(1 - \gamma Y(\tilde{R}_1 - \tau\tilde{\epsilon}_1)\right)(-Y\tilde{\epsilon}_1)\right) \\
&= -Y\mathbb{E}(\tilde{\epsilon}_1) + \gamma Y \mathbb{E}\left((\tilde{R}_1 - \tau\tilde{\epsilon}_1)\tilde{\epsilon}_1\right) \\
&= \gamma Y^2 \tilde{\sigma}^2 (1-\tau)
\end{aligned}$$

where we use the fact that $\mathbb{E}(\tilde{\epsilon}_t) = 0$, $\mathbb{E}(\tilde{R}_t^x \tilde{\epsilon}_t) = \mathbb{E}(\tilde{\epsilon}_t) = \tilde{\sigma}^2$.

Generations born in periods $t \geq 1$, after the policy has been implemented, save $S_t = Y + \tau Y \epsilon_t$ and consume in old age $C_{o,t+1} = S_t \tilde{R}_{t+1} - \tau Y \epsilon_{t+1}$ with $\frac{\partial C_{o,t+1}}{\partial \tau} = Y\left(\tilde{\epsilon}_t \tilde{R}_{t+1} - \tilde{\epsilon}_{t+1}\right)$. As a result:

$$\begin{aligned}
\mathbb{E}\left(u_o'(C_{o,t+1}) \cdot \frac{\partial C_{o,t+1}}{\partial \tau}\right) &= \mathbb{E}\left((1-\gamma C_{t+1}) \cdot \frac{\partial C_{o,t+1}}{\partial \tau}\right) \\
&= \mathbb{E}\left(\frac{\partial C_{o,t+1}}{\partial \tau}\right) - \gamma \mathbb{E}\left(C_{t+1}\frac{\partial C_{o,t+1}}{\partial \tau}\right) \\
&= -\gamma Y \mathbb{E}\left(C_{t+1}\left(\tilde{\epsilon}_t \tilde{R}_{t+1} - \tilde{\epsilon}_{t+1}\right)\right) \\
&= -\gamma Y^2 \mathbb{E}\left((1+\tau\tilde{\epsilon}_t)\tilde{R}_{t+1} - \tau\tilde{\epsilon}_{t+1}\right)\left(\tilde{\epsilon}_t \tilde{R}_{t+1} - \tilde{\epsilon}_{t+1}\right)
\end{aligned}$$

Opening up the brackets and noting that due to time independence

$$\begin{aligned}
\mathbb{E}(\tilde{R}_{t+1}^2\epsilon_t) &= \mathbb{E}\tilde{R}_{t+1}^2\mathbb{E}\epsilon_t = 0 \\
\mathbb{E}\tilde{\epsilon}_t^2\tilde{R}_{t+1}^2 &= \mathbb{E}\tilde{\epsilon}_t^2\mathbb{E}\tilde{R}_{t+1}^2 = \tilde{\sigma}^2(\tilde{\mu}^2+\tilde{\sigma}^2) \\
\mathbb{E}(\tilde{\epsilon}_t^2\tilde{R}_{t+1}^2) &= \mathbb{E}\tilde{\epsilon}_t^2\mathbb{E}\tilde{R}_{t+1}^2 = \tilde{\sigma}^2
\end{aligned}$$

we get

$$\mathbb{E}\left(u_o'(C_{o,t+1})\cdot\frac{\partial C_{o,t+1}}{\partial\tau}\right) = \gamma Y^2\tilde{\sigma}^2\left(1-\tilde{\mu}^2\tau-\tilde{\sigma}^2\tau-\tau\right)$$

The policymaker maximizes welfare by reconciling the marginal expected benefits for all generations by applying condition (3.23).

$$\frac{\beta}{\delta}\mathbb{E}\left(u_o'(C_{o,1})\cdot\frac{\partial C_{o,1}}{\partial\tau}\right)+\frac{\beta}{1-\delta}\mathbb{E}\left(u_o'(C_{o,t+1})\cdot\frac{\partial C_{o,t+1}}{\partial\tau}\right)=0$$

Substituting in the derived terms for the generations born at period zero and after that, and canceling out γ, Y, β we get

$$\begin{aligned}
\frac{1}{\delta}\tilde{\sigma}^2(1-\tau)+\frac{1}{1-\delta}\tilde{\sigma}^2\left(1-(\tilde{\mu}^2+\tilde{\sigma}^2)\tau-\tau\right)&=0 \\
\implies \tau^* = \frac{1}{\delta\mathbb{E}\tilde{R}_t^2+1} &= \frac{1}{\delta(\tilde{\mu}^2+\tilde{\sigma}^2)+1}
\end{aligned}$$

3.A.3 Variance of Old-Age Consumption

In (3.33) we have:

$$\begin{aligned}
\mathbb{Var}\left(\frac{C_{o,t+1}}{Y}\right) &= \mathbb{Var}\left(\frac{S_t}{Y}R_{t+1}-\tau\epsilon_{t+1}\right) \\
&= \mathbb{Var}\left((1+\tau\epsilon_t)(\mu+\epsilon_{t+1})-\tau\epsilon_{t+1}\right) \\
&= \mathbb{Var}\left(\mu+\epsilon_{t+1}+\tau\mu\epsilon_t+\tau\epsilon_t\epsilon_{t+1}-\tau\epsilon_{t+1}\right) \\
&= \mathbb{Var}\left((1-\tau)\epsilon_{t+1}+\tau\mu\epsilon_t+\tau\epsilon_t\epsilon_{t+1}\right) \\
&= \mathbb{Var}((1-\tau)\epsilon_{t+1})+\mathbb{Var}(\tau\mu\epsilon_t)+\mathbb{Var}(\tau\epsilon_t\epsilon_{t+1}) \\
&= (1-\tau)^2\sigma^2+(\tau)^2\mu^2\sigma^2+(\tau)^2\sigma^4
\end{aligned}$$

where due to the zero-mean and *i.i.d.* properties of the shock we have $\mathbb{Cov}(\epsilon_t, \epsilon_{t+1}) = \mathbb{Cov}(\epsilon_t, \epsilon_t\epsilon_{t+1}) = \mathbb{Cov}(\epsilon_{t+1}, \epsilon_t\epsilon_{t+1}) = 0$ and $Var(\epsilon_t\epsilon_{t+1}) = \sigma^4$ or in more details:

$$\mathbb{Cov}(\epsilon_t, \epsilon_t\epsilon_{t+1}) = \mathbb{E}(\epsilon_t^2\epsilon_{t+1}) + \mathbb{E}(\epsilon_t)\mathbb{E}(\epsilon_t\epsilon_{t+1}) \overset{i.i.d.}{=} \mathbb{E}(\epsilon_t^2)\mathbb{E}(\epsilon_{t+1}) = 0$$

$$\mathbb{Var}(\epsilon_t\epsilon_{t+1}) = \mathbb{E}(\epsilon_t^2\epsilon_{t+1}^2) - \mathbb{E}(\epsilon_t)\mathbb{E}(\epsilon_{t+1}) \overset{i.i.d.}{=} \mathbb{E}(\epsilon_t^2)\mathbb{E}(\epsilon_{t+1}^2) = \sigma^4$$

Minimizing the variance of consumption for the generation born after $t > 0$, we get the first-order condition with respect to τ

$$\begin{aligned} \frac{\partial \mathbb{Var}(C_{o,t+1})}{\partial \tau} &= -(1-\tau)\sigma^2 + \tau\mu^2 + \tau\sigma^2 = 0 \\ &\implies \tau(\mu^2 + \sigma^2 + 1) = 1 \end{aligned}$$

This illustrates again that for an increase in the variance, τ needs to go down in order to keep the outcome optimal for a future generation.

3.A.4 Allocation Decision with Risk-Free and Risky Assset

Assume that generations have a nonzero utility of consumption when old and when young, such that $u_y(C) = u_o(C)$. Also, there is a risk-free asset with a fixed gross return of R_f and a risky asset. At the same time, the young have to decide how much to consume and save and how to allocate their savings between the risk-free asset and the risky asset. The two-period budget constraints of (3.15) simplify to

$$\begin{aligned} C_{y,t} &= Y - M_t - S_t + \tilde{\epsilon}_t\tau Y \\ C_{o,t+1} &= M_t R^f + S_t\tilde{R}^s_{t+1} - \tilde{\epsilon}_{t+1}\tau Y \end{aligned}$$

Applying the individuals' optimality conditions (3.17), it can be shown that

$$I(\tau, \epsilon_t) = \begin{bmatrix} M_t \\ S_t \end{bmatrix} = Y\left(\begin{bmatrix} -(a_1+b_1) \\ a_1 \end{bmatrix} + \begin{bmatrix} -(a_2+b_2) \\ a_2 \end{bmatrix}\tau + \begin{bmatrix} -(a_3+b_3) \\ a_3 \end{bmatrix}\tau\tilde{\epsilon}_t\right) \tag{3.43}$$

where all coefficients but b_1 are positive functions of the problem's primals. Optimal consumption for the young and the old then is

$$\begin{aligned} C_{y,t} &= Y\left[1 + b_1 + b_2\tau + (1-b_3)\tau\tilde{\epsilon}_t\right] \\ C_{o,t} &= Y\left[-\left((a_1+b_1) + (a_2+b_2)\tau - (a_3+b_3)\tau\tilde{\epsilon}_{t-1}\right)R_f + \left(a_1 + a_2\tau - a_3\tau\tilde{\epsilon}_{t-1}\right)(\mu+\tilde{\epsilon}_t) - \tau\tilde{\epsilon}_t\right] \end{aligned} \tag{3.44}$$

where $\overline{\mu} = \tilde{\mu} - R_f$ is the excess return on the risky asset the coefficients in equation (3.43) are

$$a_1 = \frac{\overline{\mu}\left(-R_f Y\gamma + R_f + 1\right)}{Y\gamma\left(R_f^2\beta\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2\right)}; \quad a_2 = \frac{\tilde{\sigma}^2\left(R_f^2\beta + 1\right)}{R_f^2\beta\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2}; \quad a_3 = \frac{R_f\overline{\mu}}{R_f^2\beta\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2}$$
$$b_1 = \frac{-R_f\beta\tilde{\sigma}^2 - Y\overline{\mu}^2\gamma - Y\gamma\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2}{Y\gamma\left(R_f^2\beta\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2\right)}; \quad b_2 = \frac{R_f\overline{\mu}\beta\tilde{\sigma}^2}{R_f^2\beta\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2}; \quad b_3 = \frac{\overline{\mu}^2 + \tilde{\sigma}^2}{R_f^2\beta\tilde{\sigma}^2 + \overline{\mu}^2 + \tilde{\sigma}^2}$$

The parameters have an appealing interpretation: a_1 captures the autonomous level of savings, independent of the realized shocks; a_2 is the increase in the share of (risky) savings which varies with the risk sharing; a_3 is the variation in savings due to shocks; b_1 and b_2 capture the autonomous change in young-age consumption after the policy is implemented while $1-b_3$ captures consumption variation with the realization of the shock.

As a result, with risk sharing, on average the individual starts investing more into risky assets and less into risk-free assets. So, the standard conjecture that the capacity of individuals to bear risk increases with risk sharing, holds here as well. In addition, now we can see that the realization of a positive shock leads to a decrease in risky asset holdings and either increase in risk-free assets or an increase in young-age consumption.

3.A.5 Certainty Equivalent Consumption

Equating the cumulative utility for the two, we get:

$$\sum_{j=1}^{\infty}\delta^{t-1}\left(u(CEC) + \frac{\beta}{\delta}u(CEC)\right) = \sum_{j=1}^{\infty}\delta^{t-1}\mathbb{E}\left(\frac{\beta}{\delta}u(C_{o,t}) + u(C_{y,t})\right) = V_0$$

where in the case of the planner problem of Section 3.4, V_0 is a function of the starting wealth values $V_0 = V(W_0, X_0)$, while in the case of optimal transfers of Section 3.5.2, welfare is a determined by the optimal transfers, such that $V_0 = V(\tau^*)$.

The left-hand side can then be expanded so that we get

$$\frac{1}{1-\delta}\left(u(CEC) + \frac{\beta}{\delta}u(CEC)\right) = \frac{1}{1-\delta}\left(1 + \frac{\beta}{\delta}\right)u(CEC)$$

which produces the equation

$$CEC_0 = I_u\left[(1-\delta)\frac{\delta}{\beta+\delta}V_0\right] \tag{3.45}$$

where I_u is the inverse of the utility function.

3.B Appendix: Numerical Algorithms

3.B.1 Solution Algorithm for the Planner Problem

I solve the problem numerically through value function iteration. Several authors among other examine the tools needed to set up the solution algorithm (Judd, 1998; Miranda and Fackler, 2002; Cai and Judd, 2014; Rust, 1996; Cai et al., 2013). We use the following approach:

0. **Initialize**
   - Set up a grid for the two state variables $\{W_t\}_{j_w}, \{X_t\}_{j_x}$
   - Set up an initial guess on the function values V_0 on the grid.
   - Set-up interpolating splines for $V(W, X)$ to approximate the value function. We use in particular cubic splines, which allows us to evaluate also the first-order derivatives of the function.
   - Get bivariate quadrature points and weights for R^S and R^X (see Section 3.B.3)
1. **Maximization:** For each W_{j_w} and for each X_{j_x} find the consumption and investment vector which maximize the RHS of the Bellman equation (3.13) subject to the constraints. Sequential Quadratic Programming (SLSQP) is used to solve numerically the resulting problem.
2. **Approximation:** Find the left-hand side of (3.13) and fit a new spline approximation to the value function.
3. **Evaluation** Evaluate distance from previous-run value function $dist = ||V^i - V^{i-1}||^1$ and iterate until $dist < tol$ for some error tolerance level.

3.B.2 Solution Algorithm for the Policymaker Problem

Solving the risk-sharing problem accounts to solving two optimization problems. We have to solve the unconstrained optimization of the policymaker (3.21) by providing as input the optimal solution of the individual portfolio choice problem (3.15). We use the

Nelder-Mead method to find numerically the solution of the former and the SLSQP to find the optimum of the latter. Again, a two-dimensional quadrature (Section 3.B.3) is used to evaluate expectations.

Also, note that this specification of the liquidity factor allows me, using the law of iterated expectations, to evaluate individuals indirect utility function as

$$\mathbb{E}v(\tau, R_t) = p\mathbb{E}\left(v(\tau, R_t)|l_t = 0\right) + (1-p)\mathbb{E}\left(v(\tau, R_t)|l_t = \bar{l}\right)$$

I also split the expectation of the value function of the planner in the same way.

3.B.3 Quadrature

I use quadrature to evaluate the expectation terms in the numerical section of this paper. In low dimensions, the quadrature provides a fast and reliable approximation.[21] Multi-dimensional quadrature methods are less common, so I provide here explicitly the approach taken.

In the uni-variate space, the Gaussian quadrature performs the following approximation:

$$\int_a^b f(x)w(x)dx \approx \sum_{i=1}^m w_i f(x_i)$$

for some quadrature nodes x_i, and some positive quadrature weights w_i, and m is the number of quadrature points used.

When working with a normally distributed random variable, it is usefule to apply in particular the Gauss-Hermite (GH) quadrature which selects a weight function of the form $w(x) = e^{-x^2}$.

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx \approx \sum_{i=1}^m w_i f(x_i) \tag{3.46}$$

with x_i and w_i as the GH nodes and weights, respectively.

[21] For more details see (Judd, 1998; Cai et al., 2013; Cai and Judd, 2014).

Assume for example that the random variable y is normally distributed with $y \sim N(\mu, \sigma)$. We can evaluate the expectation of $f(y)$ as

$$\mathbb{E}(f(y)) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} f(y) e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

To reconcile this with the Gauss-Hermite approach of (3.46), we use a change of variable $y = \sqrt{2}\sigma x + \mu$ such that

$$\begin{aligned}\mathbb{E}(f(\sqrt{2}\sigma x + \mu)) &= (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} f(\sqrt{2}\sigma x + \mu) e^{-x^2} \sqrt{2}\sigma dx \\ &\approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{m} w_i f(\sqrt{2}\sigma x_i + \mu)\end{aligned}$$

We can apply this approximation in asset pricing context for an asset whose log returns R are normally distributed such that $\ln(R) \equiv y \sim N(\mu, \sigma)$. We can then find the expectation of a function of R as $\mathbb{E}f(R)$ as

$$\mathbb{E}f(e^y) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} f(e^y) e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{m} w_i f(e^{\sqrt{2}\sigma x_i + \mu})$$

In the multi-dimensional space, we can use the quadrature product rule, which approximates

$$\int_{R^d} f(x) dx \approx \sum_{i_1=1}^{m} \cdots \sum_{i_d=1}^{m} w_{i_1} w_{i_2} \cdots w_{i_d} f(x_{i_1}, x_{i_2}, \cdots, x_{i_d})$$

For example, assuming that a two-dimensional vector process $Y \sim N(\mu, \Sigma)$ where μ is a 2×1 vector of expectations and Σ is the (positive semi-definite) covariance matrix. To reconcile this with the Hermite-Gauss approximation, we follow the same approach as before. We perform a Choleski decomposition $\Sigma = LL'$ and do a change of variable

$Y = \sqrt{2}LX + \mu$ and as a result, we get

$$\begin{aligned}
\mathbb{E}(f(x)) &= 2\pi|\Sigma|^{-1/2}\int_{R^2} f(y)e^{-\frac{(Y-\mu)'\Sigma(Y-\mu)}{2}}dY \\
&= 2\pi|\Sigma|^{-1/2}\int_{R^2} f(X)e^{-X'X}2|L|dX \\
&\approx \frac{1}{\pi}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} w_{i_1}w_{i_2}f(\sqrt{2}L_{11}x_{i_1}+\mu_1, \sqrt{2}(L_{21}x_{i_1}+L_{22}x_{i_2})+\mu_2)
\end{aligned}$$

where x_i and w_i are the nodes and weights from the Gaus-Hermite quadrature, μ_1,μ_2 are elements of the vector of expectations, and L_{11},L_{22} are elements of the L matrix.

4

Quantifying Systemic Risk in the Presence of Unlisted Bank

4.1 Introduction

The canonical approach to measuring various aspects of systemic risk in banking relies on equity return correlations to assess interdependencies between banks' losses above Value at Risk (Adrian and Brunnermeier (2016)). But in many countries this approach is thwarted by the presence of state-owned and/or co-operative banks. To circumvent this problem we extend the Adrian-Brunnermeier approach (and the related Marginal Expected Shortfall (MES) approach introduced by Acharya et al. (2017)) by relying on CDS contracts rather than equity returns to extract the required information on covariance structure. We extend the portfolio-of-loans approach suggested by Huang et al. (2009, 2012) by explicitly focusing on tail risk and by modelling tail dependencies in distress. We look at key private institutions in the Dutch financial sector (insurance and banking) and develop a valuation-of-loans approach to measure systemic risk and identify and rank the systemic players on this market. Our approach is appropriate whenever potentially systemic institutions are not publicly traded on the equity market. Our analysis con-

[0]Acknowledgements: We are grateful to DNB for the financial and operational support during the realization of this project, Saskia de Vries of DNB for her support, Jeroen Huiting, Kenny Martens, Laura Izquierdo Rios from DNB Financial Stability Division for earaly discussions. However, the views expressed in this paper are purely thosue of the authors and do not necessarily correspond to views held by the DNB. We would also like to thank Franc Klassen for his valuable feedback, as well as participants in the research seminars at the University of Amsterdam, and DNB for the valuable discussions.

firms that financial institutions need to be monitored in the context of other financial institutions, as Adrian and Brunnermeier (2016) also argue. In particular, we show that important linkages exist between banking and insurance firms that need to be explored further and taken into account when measuring systemic risk.

Systemic linkages arise naturally through various channels. A direct link stems from the channels by which banks operate on the interbank market. Banks and insurers also tend to trade directly with each other on the derivative markets. Finally systemic dependencies may also arise indirectly, due to common exposure of the key institutions to the same risk sources - either on the liability side, when funding sources are similar, or on the asset side, when the institutions hold similar or correlated asset portfolios (Moore and Zhou, 2012; de Haan et al., 2019). We present a framework that does not require a particular view on what is causing systemic losses, but rather offers an approach that can identify the potential for high joint distress based on observed dependencies between traded credit protection on the market whatever the underlying channels of interdependencies are.

First, we show that monitoring the financial risk of an institution in isolation of the risks of its counterparties, and the system as a whole, may offer a misleading ranking between systemically important financial institutions (SIFIs). Second, we illustrate that high-frequency data from the credit default swap (CDS) market can be used to monitor *ex-ante* the build-up of systemic risk and systemic dependencies. This is particularly valuable in the context of the Dutch financial sector, where key institutions are privately held, and market data on their equity value is not available. Third, we link systemic risk to the potential for joint distress between institutions by evaluating the tail dependencies in their losses if a default of one institution were to occur.

We define systemic risk both through the prospect that several key institutions become distressed at the same time, and through the prospect that the common losses they generate may have a large social impact. To quantify such risk, our model relies on several building blocks. First, we use a contingent balance sheet approach (Merton, 1974) and define distress as the situation in which the market value of a firm's assets falls below a default barrier. The observed CDS spreads allow us to estimate the probability of such distress occurring. Second, a latent factor is assumed to drive common changes in the asset values of firms. Systemic risk will thus have two related components: first, the possibility that several companies realize a credit event at the same time; and second, the magnitude and the dependency in the losses that are generated among the financial institutions once a default occurs. We aggregate the two components using a credit portfolio approach and estimate the MES for the institutions in the portfolio (Acharya et al., 2017). The MES measures the average potential loss of an institution if the system as a whole realizes a tail event, thus quantifying the sensitivity of an institution to other institutional losses in the system. In addition, we relate the liability-weighted MES to the share of systemic risk that can be attributed to a single institution.

To the best of our knowledge, we are the first to model empirically, in a systemic risk context, dependencies between default occurrences and the potential losses given a default. Such dependencies are crucial for a number of reasons. First, there is sound empirical evidence that realized losses tend to rise in periods when risk probabilities also increase (Artzner, 1999). Second, the potential default of a SIFI by definition will have a strong impact on other players in the industry by increasing their default risk and at the same time lowering the value of the assets backing up their liabilities below fair value as industry-wide distress triggers fire sales. From that point of view, we argue that reliable systemic risk estimation should cover the potential for LGD (Loss Given Default) dependencies.

We use a flexible modeling approach which allows for factor exposure heterogeneity fitted on CDS data. This is an improvement over the well-known Vasicek credit model (Vasicek, 1987) which assumes a single correlation parameter driving the dependencies in the whole portfolio.

We look at seven Dutch financial institutions: two insurers (Aegon and NN) and five banks (ING Bank, ABN, VB, Rabobank, and NIBC). Our model allows us to rank the companies by their contribution to risk, where risk is quantified by the Expected Shortfall of the systemic portfolio.

The current paper continues as follows. In Section 4.2 we review the relevant literature. Section 4.3 describes the structural credit model we employ to describe co-dependencies between institutions in the system. Section 4.4 discusses the credit risk approach used to quantify the sensitivity and the contribution of each institution to systemic risk. Section

4.5 reviews the dataset and defines the regulatory portfolio. Sections (4.6) and (4.7) discuss the results and respectively their policy relevance, while Section 4.8 concludes.

4.2 Literature Review

Our paper is part of the wider literature using high-frequency asset prices to inform central bank policies. Examples are Hattori et al. (2016); Olijslagers et al. (2019) who use option-implied asset volatilities and risk-neutral distributions to evaluate the effectiveness of central bank stabilization policies. Market-implied views have also been seen as a valuable tool for monitoring financial stability and for advising on macro-prudential policies (Jayaram and Gadanecz, 2016). Acharya et al. (2014) use co-movements in CDS rates of sovereigns and local banks during the Euro sovereign debt crisis to show how a *doom-loop* channel evolves, in which a bail-out of a local bank in trouble, because it is deemed systemically important, leads to a deterioration in the creditworthiness of the government, which then further depresses the credit-worthiness of the bailed-out bank due to its large exposure to local sovereign bonds; after which there are further hits to the solvency of the government and so on.

Also, our paper relates closely to the literature of systemic risk which utilizes equity market information. In many cases, especially in Europe and certainly in the sample of Dutch institutions that we consider, the major challenge in exposing market-implied views is that some of the key players in the financial sector are not publicly traded. Approaches that rely on equity price co-movements (like Adrian and Brunnermeier (2016)) then cannot encompass the full system, cannot be used to track the systemic impact of those institutions, and may in fact not be usable at all if too few of the quantitatively important institutions have an equity market listing. For this reason, we develop a structural approach that utilizes information from the CDS market.

The intuition behind the mechanism that we employ is simple. We know through Merton (1974) that the market value of a company's assets is related both to the market value of its equity and of its liabilities. The level of the firm's CDS spread at any particular instance relates to the chance that the value of its assets may drop and that it may experience distress in the form of a credit event captured by the CDS contract[1]. What is more important for us however, is that co-movements in default probabilities can provide information on the tendency of the institutions to become distressed at the same time. Tarashev and Zhu (2006) also follow this line of reasoning. Rather than estimating the unobservable asset values, as is done for example in Duan (1994, 2000) and in Lehar (2005), we add a model of the losses in case of default that allows us to quantify the distribution of systemic losses and the potential for large losses by several institutions at the same time.

A CDS is in essence an insurance contract, which is traded over-the-counter (OTC), and in which the protection buyer agrees to make regular payments, the CDS spread rate over a notional amount, to the protection seller. In return, the protection seller commits to compensate the buyer in case of default of the contractually referenced institution. The value of a CDS contract thus provides information on the fair value spread that should be used to discount the company's debt[2].

The CDS market has several features that make it an attractive source of information for the financial sector. It is more liquid and has fewer trading frictions compared to credit traded directly through the corporate bonds market. In terms of information transmission, CDS spreads have been shown to lead bond markets, especially in distress periods, and have an edge over credit rating agencies (Bai and Collin-Dufresne, 2019; Avino et al., 2019; Culp et al., 2018; Annaert et al., 2013). Some evidence exists that they may even lead equity markets, especially in revealing negative credit news. This relates to the fact that in contrast to conventional asset markets, the CDS market almost by definition is composed of insiders (Acharya and Johnson, 2005). Furthermore, liquidity and transparency in the market have increased substantially in recent years. After the Financial Crisis of 2008/09, OTC derivatives, and as such also CDS contracts, became subject to increased regulatory scrutiny through the EMIR framework in Europe and the Dodd-Frank Act in the US. To

[1]For a similar line of thinking, cf Carr and Wu (2011) who provide a link between the value of a CDS contract and deep out-of-the-money put options on a company's stock.

[2]Apart from a hedging opportunity, CDSs are used to arbitrage away any relative mispricing between the equity and bond prices of the reference entity (capital structure arbitrage (Kapadia and Pu, 2012)), or to exploit mispricings in the value of traded debt (Augustin and Schnitzler, 2021). There is a large empirical literature dealing with such possibilities and the limits to arbitrage, which we do not consider in the current study.

cope with systemic risk issues, central clearing was introduced with increased contract standardization and transparency was improved by introducing reporting mandates for counterparties[3].

Furthermore, CDS prices trade on standardized terms and conditions and do not have be bootstrapped or interpolated as do bond yields. Also, comparison between firms is easier, because unlike corporate fixed income securities, single-name CDS contracts do not contain additional noise from issue-specific covenants, such as seniority, callability or coupon structure (Zhang et al., 2009; Culp et al., 2018).

Several general concerns regarding CDS prices need to be mentioned as well however. First, CDS rates also price in the risk of default of the protection seller and not only the reference entity. The size of this extra premium, however, has been shown empirically to be economically negligible (Arora et al., 2012), and with the recent rise of Central Clearing for OTC derivatives it is likely to have decreased further (Loon and Zhong, 2014, 2016). Second, single-name CDS contracts are not as liquid as public equity and this raises concerns that the spreads could be overstating default risk by confounding it with an illiquidity premium. Even though the argument is valid, it misses two important points. Illiquidity risk tends to be correlated with default risk, as protection dries up at times when it is most needed (Kamga and Wilde; Augustin and Schnitzler, 2021). Also, strong illiquidity in the CDS contract may be indicative of the market's unwillingness to fund a particular financial institution due to fears that a possible future fire sale could push it into insolvency (cf Diamond and Rajan (2011)), and may well reflect a private cost of leverage as in Shleifer and Vishny (1992). Overall, we take the view of Segoviano and Goodhart (2009), backed up by empirical evidence, that even though in magnitude CDS spreads may be overreacting to bad news in certain situations, the direction is usually justified by information on the reference institution's creditworthiness. Thus, we use the CDS mid quotes without correcting them further for non-credit related premia.

Part of the literature on bank distress relies on reduced-form statistical modelling to link bank CDS movements to periods of financial adversity. Avino et al. (2019), for example, look at the spreads of single-name CDS contracts for European and US banks

[3]For an overview of the market microstructure, and recent regulatory reforms of the CDS market see Aldasoro and Ehlers (2018) and Paddrik and Tompaidis (2019).

and evaluate the propensity of spread changes to predict bank distress in the form of recapitalization or nationalization. One standard deviation increase in the CDS spread changes, is estimated to correspond to a 7% to 14% increase in the (physical) probability of financial distress of a bank. Annaert et al. (2013) look at the determinants of CDS spread changes for a universe of European banks and separate them into a firm-specific credit risk component, a trading liquidity component, and a business cycle components capturing common variation linked to the business environment.

On the methodological front, Oh and Patton (2018) link bank distress to large upticks in the CDS prices of the reference banks, and measure the probability of joint distress through a factor copula dependency model. Billio et al. (2012) offer an early econometric model which quantifies interconnectedness through Granger-causality networks. Bräuning and Koopman (2016) extend the idea with time-varying heterogeneity in the link formation between banks using CDS spreads of US and European institutions. The goal is to capture the dynamic formation of potential core-periphery clusters, which are natural for the financial sector. Moratis and Sakellaris (2021) on the other hand use a panel VAR model to decompose the transmission of systemic shocks across a universe of global banks. These studies offer preliminary evidence that CDS fluctuations can serve as an early warning signal of bank risk, supplementing data from the stock market, credit rating agencies, and accounting data. Our contribution to this literature is to embed CDS spreads into a structural model of the firm's capital, which allows for non-linear relationships to form naturally.

An earlier branch of the empirical literature also uses structural firm models to imply bank fragility (Gropp et al., 2006; Chan-Lau and Sy, 2007; Bharath and Shumway, 2008). Most notable is the distance-to-default (DD) measure (Merton, 1974; Crosbie and Bohn, 2002) which compares the current market value of assets to the default barrier of the firm[4]. While the foundation in our study is similar, we aim to evaluating cross-linkages and the impact each bank has on the system as a whole, rather than on modelling the individual default risk of each bank in isolation.

[4]Various extensions of the DD measure exist, capturing for example volatility clustering (Nagel and Purnanandam, 2019), and asymmetric volatility shocks (Kenc et al., 2021).

Most of all, we relate to the broader literature on measuring and quantifying systemic risk through asset price co-movements (Lehar, 2005; Segoviano and Goodhart, 2009; Zhou, 2010; Huang et al., 2012; Adrian and Brunnermeier, 2016; Brownlees and Engle, 2017; Acharya et al., 2017; Engle, 2018). Some of the approaches developed in that area can be seen as largely model-free since they do not rely on particular capital structure assumptions of the individual firms. The CoVaR approach of Adrian and Brunnermeier (2016) for example, along with an earlier study by Baur and Schulze (2009), relies on a quantile regression on equity prices to determine tail co-dependencies and risk contributions. Wang (2021) adapt this approach by embedding a neural network. Most of these studies rely on high-frequency data on equity prices.

Another strand of the systemic literature, most notably Lehar (2005), relies on Merton's theory of contingent claims to imply the market value of firm assets and the correlations between institutions as a measure of systemic risk. In contrast, our approach does not aim to imply the value of assets themselves. Rather, we directly focus on the potential for systemic events to materialize and evaluate the potential losses for the systemic portfolio when defaults occur. Using a structural model in combination with copula default dependencies, Segoviano and Goodhart (2009) comes to the PAO measure, the probability of at least one more bank defaulting given a default in particular bank. We develop the idea further by also calculating the probability of two or more defaults given that at least one has occurred. This allows us to concentrate specifically on periods of financial contagion.

We view the regulatory space as a portfolio of risky loans, similar to Chan-Lau and Gravelle (2005); Huang et al. (2009, 2012); Puzanova and Düllmann (2013); Kaserer and Klein (2019). In that approach, systemic losses arise when an institution defaults and cannot cover the value of its liabilities. The tendency of particular institutions to drive systemic losses will result in a higher contribution to systemic risk.

From this perspective, the modeling tools developed by the securitization literature, typically used to value n-th to default derivatives on loan portfolios, can be applied (Hull and White, 2004; Tarashev and Zhu, 2006). In particular, Tarashev and Zhu (2006) link the correlation structure embedded in CDS prices to the correlation between asset values in the Merton capital structure framework. A latent factor model driving the asset return variations can then be used to connect the default probabilities of the different institutions.[5]

Our innovation is to also embed a model of correlated losses between the institutions, overcoming a modeling deficiency in earlier studies, which typically assume a fixed Loss Given Default (LGD) (Puzanova and Düllmann, 2013) or assume that Recovery Rates (RRs) are random but sampled independently from each other (Huang et al., 2012; Kaserer and Klein, 2019). In a tail scenario, a SIFI's default can be expected not only to raise the default risk of other participants in the sector, but also to simultaneously decrease the value of the assets backing up their liabilities. From that point our approach of endogenizing the LGD relates to the literature on fire sales. See for example Shleifer and Vishny (1992) who argue that in times of industry-wide distress and increased default rates, assets tend to go to industry outsiders who may lack the necessary skills to manage them and will thus be willing to buy them only at a discount to fair value. As a result, LGDs will tend to rise with the drop in liquidation prices. This has been empirically observed among others by Acharya et al. (2007).[6]

We finish by quantifying systemic risk through a Monte Carlo simulation of the possible scenarios over the coming year by evaluating the average loss of an institution if the portfolio as a whole is its tail (Acharya et al., 2017).

4.3 A Structural Model of Defaults and Losses

We begin by defining the structural credit risk model behind the occurrence of systemic losses. Key here will be the assumptions driving asset value correlations and loss correlations. These asset value processes will be at the core of the data-generating processes that we define in sections (4.3.1) and (4.3.2), dealing respectively with default correlations and correlations of Losses Given Default. These data-generating processes will then guide factor model estimation in Section 4.3.3, and tail-risk estimation later in Section 4.4.

[5] The approach here can be traced back to an early latent factor credit model developed by Vasicek (1987) to price loan portfolios. In general, using a factor model to drive the correlations structure between portfolio positions is referred to as a factor copula model.

[6] See also IJtsma and Spierdijk (2017) for a discussion of fire sales, endogenous LGDs and the relation to systemic risk.

4.3.1 Default and Asset Correlations

We start from Merton (1974) and describe the evolution of the value of assets of each institution $i = 1, ..., n$ under the risk-neutral measure through the process

$$d \ln V_{i,t} = rdt + \sigma_{v,i} dW_{i,t} \tag{4.1}$$

Note that we can write (4.1) as $dW_{i,t} = \frac{d \ln V_{i,t} - rdt}{\sigma_{v,i}}$ which gives the statistical interpretation of $dW_{i,t}$ as the standardized excess asset returns under the risk-neutral measure.

We assume that the risk component of asset value changes is driven by a common factor component M_t and an idiosyncratic component $Z_{i,t}$:

$$dW_{i,t} = A_i M_t + \sqrt{1 - A_i A_i'} Z_{i,t} \tag{4.2}$$

where $M_t = [m_{1,t}, \ldots m_{f,t}]'$ is the vector of stochastic latent factors and $Z_{i,t}$ is the firm-specific factor. $A_i = [\alpha_{i,1}, ..., \alpha_{i,f}]$ is the vector of factor loadings, such that $A_i A_i' \leq 1$. All factors are assumed to be mutually independent with zero mean and a standard deviation of one. Note that if one assumes $A_i = A_j$ for all i, j, one gets the well known Vasicek loan pricing model which assumes the same averaged-out factor exposure across all loans. In the approach used here, we allow for exposure heterogeneity. One could interpret the factors as independent economy-wide and industry-wide shocks affecting the uncertainty in the firm's asset value.

In Merton's setting[7], default occurs at maturity ($T = t + \Delta t$) when assets fall below the face value of debt:

$$\begin{aligned} PD_t &= \mathbb{P}(V_{t+\Delta t} \leq D) \\ &= \mathbb{P}\left(V_t \exp\left((r - \frac{\sigma_v^2}{2})\Delta t + \sigma_v W_{t+\Delta t} \right) \leq D \right) \end{aligned}$$

[7] See Appendix (4.A) for presentation of Merton's firm value model and the role spreads play in it.

Consider next Distance-to-Default DD_t[8]:

$$DD_t = \frac{\ln \frac{V_t}{D} + \left(r - \frac{\sigma_v^2}{2}\right) \Delta t}{\sigma_v \sqrt{\Delta t}}$$

which allows us to rewrite the expression for the probability of default as:

$$PD_t = \mathbb{P}\left(\frac{W_{t+\Delta t}}{\sqrt{\Delta t}} \leq -DD_t\right)$$

As a result, we get:

$$PD_t = \Phi(-DD_t) \tag{4.3}$$

where $\Phi(.)$ is the cumulative standard normal distribution.

We can then write the discrete first difference of DD_t as:

$$\Delta DD_t = \frac{\Delta \ln V_t}{\sigma_v}$$

The correlation between asset returns can be written as:

$$\begin{aligned}\rho_{i,j} &= \mathbb{C}\mathrm{orr}(\Delta \ln V_{i,t}, \Delta \ln V_{j,t}) \\ &= \mathbb{C}\mathrm{orr}(\sigma_{v,i} \Delta DD_{i,t}, \sigma_{v,j} \Delta DD_{j,t})\end{aligned}$$

Correlations are invariant to linear transformation, so we can drop the σ_v term. Then after substituting in the inverted relationship (4.3), the asset correlations can be implied from the correlations between the transformed probabilities of the default:

$$\rho_{i,j} = \mathbb{C}\mathrm{orr}\left(\Delta\Phi^{-1}(PD_{i,t}), \Delta\Phi^{-1}(PD_{j,t})\right) \tag{4.4}$$

Equation (4.4) is of crucial importance because it relates the co-dependencies in the probabilities of default (PDs) to the asset correlations of the underlying institutions. This allows us to use PDs that can be derived from observed single-name CDS prices to pinpoint

[8]The DD measure has a wide application to risk management as a predictable indicator of bank fragility (Gropp et al., 2006; Chan-Lau and Sy, 2007).

values for the correlations between institutions. In Section 4.3.3 we discuss in detail how these asset correlations can be used to estimate the parameters of the latent factor model in (4.2).

Our reliance on the Merton (1974) framework implies that we assume default to occur when a fixed default barrier is crossed at debt maturity. Further refinements have been developed to relax this assumption, of which we mention in particular Leland (1994) who endogenizes the default barrier and defines it as the boundary beyond which equity holders refuse to supply new equity to avoid default. Even though the Merton framework maybe conceptually restrictive, it is a widely used as a raw approximation of default. The related Merton-based DD has also been shown to be predictive of actual defaults (Bharath and Shumway, 2008) and has certain robustness against model misspecification (Jessen and Lando, 2015). As a result, we do not pursue any of the structural extensions in this study.[9]

Next, we use the asset correlations implied from co-movements in the default probabilities to solve for the loadings of the latent factor model (4.2). We pick the coefficients which minimize the squared error between the target correlations and the correlations implied by the factor loadings:

$$\min_{A_1,\ldots,A_n} \sum_{i=2}^{N} \sum_{j=1}^{N} (\rho_{ij} - A_i A_j')^2 \tag{4.5}$$

An efficient algorithm that solves the minimization problem is provided by Andersen and Basu (2003). It operates through an iterative principal component analysis rather than by brute force numerical optimization. Appendix (4.B) clarifies the algorithm.[10]

In Section 4.3.3 we discuss in detail how observed CDS rates can be used to imply default probabilities for the period, how the target correlations $\rho_{i,j}$ are set and in turn how the factor model driving asset returns is estimated, but before doing that we have to specify the processes driving losses conditional on default.

[9] See Sundaresan (2013) for a review of structural credit models and their applications.

[10] An alternative is to use Kalman Filtering techniques. As shown by Tarashev and Zhu (2006), the two produce very similar results.

4.3.2 A Model of Loss Correlations

The next step is determining the size of the potential losses if a default were to occur. A common deficiency in the systemic risk literature which uses the portfolio-of-loans approach is that the realized recovery rate RR is assumed to be either fixed (Puzanova and Düllmann, 2013) or stochastic but independent across firms and from the realization of default (Huang et al., 2009, 2012; Kaserer and Klein, 2019).[11] Relying on strong assumptions on default losses is inevitable, as defaults, especially of SIFIs, are rarely observed. Yet, we try to addresses the empirical evidence that as default rates in the economy increase, the recovery values on assets decrease (Altman et al., 2004; Acharya et al., 2007). Therefore in an extension of the existing literature we allow default losses to be dependent on the latent factors driving asset correlations. Accounting for this is likely to have significant consequences for the quantification of systemic risk which inevitably depends on the tail risk dependencies between institutions.

To do so we follow Frye (2000) and Andersen and Sidenius (2005) and model the RRs based on the value of a stochastic collateral process $C_{i,t}$ which backs up liabilities. Dependency is achieved by making the value of the collateral dependent on the same set of factors that drive the asset value processes. In particular, we define the value of collateral per euro of liabilities as:

$$d \ln C_i = \sigma_c dW_i^c \tag{4.6}$$

where dW_i^c is a term driving the total recovery risk and σ_c is a scaling parameter.

We assume that common collateral value variation is driven by the same common factors defining the asset correlations in (4.2). Z_i^c defines an independent factor capturing possible firm-specific discrepancies between the underlying assets of the firm and the value of recovered collateral, which could be due to a loss on the value of intangible assets, or any other restricting costs due to, liquidation, or delay costs. The same factor loadings

[11] This relies on a modelling approach often used in the credit risk literature to sample simulations of the random RRs independently from triangular or beta distributions (Hull, 2018).

determined in (4.2) are assumed to hold here as well. Formally, we therefore have

$$dW_i^c = A_i M + \sqrt{1 - A_i A_i'} Z_i^c \tag{4.7}$$

Finally, in case of default, the recovery rate (RR_i) as a proportion of liabilities is never larger than 100% of the recovered liabilities:

$$\begin{aligned} RR_i &= \mu_{c,i} \min(1, C_i) \\ &= \mu_{c,i} \min\left(1, \exp\{\sigma_c dW_i^c\}\right) \end{aligned} \tag{4.8}$$

where $\mu_{c,i}$ is calibrated to match the assumption of the expected RR.

We do not have a reliable way to estimate σ_c for each institution in the portfolio, so we match it to the VSTOXX index (Figure 4.14) as a way of generating time variation which is tied to the willingness of investors to take risks and thus to the overall asset valuation sentiment in the economy. VSTOXX, similarly to its US counterpart VIX, measures the implied volatility derived from near-term exchange-traded options on the Euro Stoxx 50 index. The options are widely used by investors for hedging purposes, so the two composite indices constructed from their prices are indicative of the risk appetite prevalent in the economy. A low appetite for risk relates to a greater cost of capital, lowering investments, and driving down prices, while a high appetite relates to credit and asset price bubbles, increasing the chance for future recessions and stress in the financial system (Illing and Aaron, 2005; Gai and Vause, 2006; Aven, 2013).

4.3.3 Extracting Default Probabilities from observed CDS prices

We now proceed with the estimation of the factor loadings of equation (4.2). The first step is to find the institutions' default probabilities over time. Once we have these time series, relation (4.4) allows us to pin down the asset correlation matrix between the various institutions under consideration. These will serve as target correlations against which the model is fitted when estimating the factor loadings.

So, first, we extract the (risk-neutral) default probabilities needed in Equation (4.4) from the observed CDS rates. Following Duffie (1999) we assume, in this subsection only, that RRs are constant over the horizon of the contract, setting aside the equation (4.8) we use in analyzing correlated LGDs. We do not try to identify expected recovery rates separately from the observed CDS data. There are alternative and more sophisticated approaches; for example Pan and Singleton (2008) identify separately the RR and the default intensity of the credit process exploiting the term structure of the CDS curve constructed from contracts with different maturities. Christensen (2006) models jointly the dynamics of the RR, the default intensity, and interest rate by breaking away from the standard Recovery of Market Value (RMV) approach of Duffie and Singleton (1999) according to which at default the bondholder receives a fixed fraction of the prevailing market value of the firm. Under the RMV approach the default intensity only shows up within a product with the recovery rate, so the two cannot be identified separately. Having one collateral model when assessing LGD correlations and another one when extracting default probabilities from observed CDS spreads comes down to an inconsistency that is well known in the literature (see Tarashev and Zhu (2006)'s discussion of precisely this issue). Yet, the simplifying assumption we employ in estimation is widely used in the literature and is hard to improve on given the identification problem we just discussed.

With this in mind we can proceed with the pricing equation of the CDS contract. By market convention, at the initiation date t of the contract the spread CDS_t is set to ensure that the value of the protection leg and the premium leg of the contract are equal, such that the contract has a zero value:

$$\underbrace{CDS_t \int_t^{t+T} e^{-r_\tau \tau} \Gamma_\tau d\tau}_{\text{PV of CDS premia}} = \underbrace{(1 - ERR_t) \int_t^{t+T} e^{-r_\tau \tau} q_\tau d\tau}_{\text{PV of protection payment}} \tag{4.9}$$

r_τ is the risk-free rate, CDS_t is the observed CDS spread for the day, q_τ is the annualized instantaneous risk-neutral default probability , $\Gamma_\tau = 1 - \int_t^\tau q_s ds$ is the risk-neutral survival probability until time τ, and ERR_t is the expected recovery rate in case of default, assumed to be constant over time.

For simplicity we assume that the risk-free rate r and the annualized default rate q are fixed over the horizon of the contract. Then the default probability q at time t follows from equation (4.9):

$$q_t = \frac{aCDS_t}{a(1 - ERR_t) + bCDS_t} \tag{4.10}$$

with $a = \int_t^{t+T} e^{-r\tau} d\tau$ and $b = \int_t^{t+T} \tau e^{-r\tau} d\tau$. Setting $T = 5$ to capture 5 year CDS contracts, we can imply the annualized default probabilities.[12] We can then substitute the implied risk neutral probability q_t for PD_t in Equation (4.4), which then allows us to fix the asset correlations between all pairs of institutions.

Finally we should point out that we are ignoring correlation risk premia. We rely on evidence provided by Tarashev and Zhu (2006) that such premia, if they exist at all, are quantitatively very small in CDS prices.

4.4 Measuring Systemic Risk: A Credit Portfolio Approach

We now have the machinery in place to start modeling systemic risk. We model the space of institutions falling under the regulator's supervision as a structured credit portfolio. An institution becomes distressed if a credit event occurs in its subordinated debt. Each institution's liability can be seen as a loan from the public and amounts to the total Exposure at Distress (EAD). A loss occurs whenever an institution defaults and cannot deliver the full promise of its outstanding liabilities to its counterparties.

Formally, the systemic loss L_{sys} is the sum of the individual losses of each institution in case of distress over the following year; the sum is scaled by the total liabilities in the system:

$$\begin{aligned} L_{sys} &= \sum_{i=1}^{n} w_i L_i \\ L_i &= \mathbb{1}_{d_i}(1 - RR_i) \end{aligned} \tag{4.11}$$

[12]In credit risk (and more generally in survival analysis), the variable q relates to the *hazard rate*, the constant arrival rate (in a Poisson sense) of a credit event. At any instant, given that a default has not yet occurred, the time until it does is exponentially distributed with parameter q. For a small Δt and small q, the probability of default is then $\Delta t.q$. See Duffie (1999) for details.

where each loss L_i stands for the percentage losses in default as a proportion of the own liabilities of institution i, and $w_i = \frac{B_i}{\sum_{j=1}^{N} B_j}$ is the relative weight of the institution's liabilities (B_i) in the systemic portfolio. $\mathbb{1}_{d_i}$ is a default indicator function, where in line with the structural assumptions made so far default occurs when $dW_i \leq -DD_i$, in line with the expression from equation (4.3).

We define systemic risk as the potential for large default losses in the financial system. A single entity's contribution to systemic risk then will be measured as its propensity to increase that potential. Several elements can thus drive the systemic risk contributions of an institution. First of all, both increases in the default probability and decreases in the proportion that can be recovered in case of default will lead to a higher contribution. Second, the size of the institution, measured by its outstanding liability relative to the size of others, will determine how important the institution's potential losses are for the system as a whole. Third, the propensity of the institution to become distressed or to realize large losses whenever other institutions in the portfolio are distressed will also affect its systemic risk contribution.

Formally, we quantify downside risk through Expected Shortfall[13] (ES), which measures the average losses of an institution, or, where relevant, the portfolio as a whole, in the worst q-th percentile of its potential loss distribution:

$$ES_i = \mathbb{E}(L_i | L_i \geq VaR_i) \tag{4.12}$$

where VaR_i stands for the Value-at-Risk of the institution at confidence level (CL) $1-q$:[14]

$$\mathbb{P}(L_i \geq VaR_i) = q$$

The ES thus measures the average loss once the VaR-threshold of an institution has been exceeded. An appealing feature of this measure is that it is coherent, in the sense

[13] The measure is often referred to as Expected Tail Loss or Conditional Value at Risk (Rachev et al., 2008)

[14] Typically, q stands for the tail probability and takes value of e.g. $5\%, 1\%, .01\%$ depending on how far in the tail we want to measure the potential for extreme losses. Then, given the potential loss distribution, we are $(1-q)\%$ certain that losses will not exceed the corresponding VaR estimate.

of Artzner (1999), and thus allows for capturing diversification in an intuitive way when the losses of a system are aggregated.[15]

The VaR and the ES of a financial institution quantify the potential losses that could occur if an institution is distressed. These measures however do not take into account the fact that banks operate in a network and each institution's failure may trigger failures of other institutions.

As a measure of tail codependency, we follow Acharya et al. (2017) to define Marginal Expected Shortfall (MES) as the average loss of institution i given that the system is in the worst q-th percentile of its distribution of potential losses:

$$MES_i = \mathbb{E}\left(L_i|L_{sys} \geq VaR_{sys}\right) \tag{4.13}$$

Note that the weighted sum of all MESs in the portfolio provides the ES of the system. This follows from (4.12) and (4.11):

$$\begin{aligned} ES_{sys} =& \mathbb{E}\left(\sum_i w_i L_i|L_{sys} \geq VaR_{sys}\right) \\ =& \sum_i w_i \mathbb{E}\left(L_i|L_{sys} \geq VaR_{sys}\right) \\ =& \sum_i w_i MES_i \end{aligned} \tag{4.14}$$

This additivity property allows us to break down the total ES of the portfolio into percentage contributions due to each institution as

$$PC \text{ to } ES_i = \frac{w_i MES_i}{ES_{sys}} \tag{4.15}$$

which will be a useful metric further on in attributing risk across institutions and ranking them by systemic importance. Note that (4.14) implies also that the MES measure can

[15] The set of coherent risk measures are defined axiomatically through a number of intuitive properties: (1) *Monotonicity*: comparing several random payoffs, lower losses in all states of nature imply lower risk; (2) *Positivide homogeneity*: scaling a portfolio random payoff by a positive factor also scales its risk by the same factor; (3) *Sub-additivity*: the risk of the portfolio is not greater than the sum of the risks of the assets which comprise it; (4) *Invariance*: adding cash to a portfolio reduces its risk by the amount added. ES covers all of the properties, while VaR fails at sub-additivity. In fact, functionals which satisfy (2) and (3) are convex, a feature that defines mathematically the concept of diversification in modern portfolio theory (Rachev et al., 2008).

be interpreted as the sensitivity of the system's tail risk to the weight of the institution in the portfolio as we have $\frac{\partial ES_{sys}}{\partial w_i} = MES_i$.

4.5 Data

4.5.1 Note on the Dutch Financial Sector

The Dutch banking sector is comparatively large relative to other EU countries: the total cumulative balance sheet value of all banks accounts to about 400% of GDP in the beginning of 2013, a rise from about 100% in the 1970s. For comparison that figure amounted to about 300% in the EU and in Germany according to figures by DNB (2015). By 2018, The Netherlands is in the top 5 countries ranked by the ratio of value of bank assets to GDP (DNB, 2019). The sector is highly concentrated, and domestic banks are dominating the market. We look at the five largest Dutch banks:

- **ING Bank**: privately held by ING Group, it is the most internationally oriented Dutch bank with operations also in Mexico, Taiwan, etc.
- **ABN AMRO**: mostly focused domestically with some operations abroad. Equity on the company is publicly traded, providing a little less than half of the capital of the bank. The remainder is Government held, down from 100% after a nationalization/rescue in 2008. It is the only bank in our sample whose equity is publicly traded.[16]
- **Rabobank**: a cooperative bank, largely focused on the agricultural and consumer sector with certain activities abroad.
- **NIBC Bank N.V.**: a commercial bank, subsidiary of NIBC Holding N.V. which is publicly traded.

[15] In the risk management jargon (Hull, 2018), the weighted MESs are often referred to as *Component* Expected Shortfall.

[16] ING Group's equity is publicly traded, but it owns a large number of subsidiaries operating worldwide that are operationally and legally disjoint from the Dutch subsidiary. NIBC's equity has been de-listed since February, 2021.

- **De Volksbank**: a bank holding operating exclusively in the Netherlands owned by the Dutch state since the nationalization of its predecessor SNS in 2012.

Out of the 5 institutions, ABN AMRO, ING, and Rabobank are designated as systemically important by the European Banking Authority.

In addition, we look at two insurance companies, which have CDS swaps traded on their name:

- **NN Group**: one of the largest insurance holdings in the Netherlands. It is active in life and non-life insurance, and also has an asset management branch. NN was part of the ING Group and was split off from it between 2013 and 206. Its equity is currently publicly traded.

- **Aegon NV**: a holding company engaging in insurance, pensions, and asset management services. It is globally active in its operations.

The goal is to check if the market perceives dependencies between the insurance sector and the banking sector in the Netherlands and to check if any of the insurance firms will show up as systemically important when looked at in the context of the total financial sector. Also, we want to capture any potential interlinkages between insurers and banks that could drive systemic losses. The equity of both insurers is publicly traded on the equity market.

4.5.2 Dataset and Data Assumptions

We use weekly data for ISDA'14 compliant CDS mid prices on subordinated debt. The data is collected from Bloomberg. Figure 4.12 in Appendix (4.C) shows the evolution of CDS rates for each institution and Figure 4.13 (also in Appendix (4.C)) shows a scatter matrix and distribution plots for the CDS rate log changes. This gives an initial view of the possible dependencies in the occurrence of credit events between institutions.

We evaluate systemic risk in a cross section and over time. First we use the period September 9th, 2019 to September 13th, 2021 to evaluate and rank the institutions by their contribution to systemic risk. Then we use a rolling window backtest in the period January 1st, 2010 up to September 13th, 2021 to evaluate the evolution of the systemic risk measures. In the spirit of Lehar (2005), the time window consists of 2 years of weekly observations to evaluate the model and project the risk metrics, after which the window is shifted forward by a week and the model is re-evaluated. This produces a series of out-of-sample metrics.

Annual balance sheet data is collected from FactSet and from publicly available financial statements of the firms, whenever the data provider has a gap. The annual numbers are interpolated to weekly with a cubic spline to avoid jumps at year-end, driven by accounting standards rather than the arrival of new market information.

The structure of the liabilities of each company is used to induce the expected recovery rate (ERR) in case of default (Figure 4.15 and Table 4.1). Following Kaserer and Klein (2019), an expected recovery rate of 80% is assumed on deposits (in case the institution is a bank) or policy insurance liabilities (in case the institution is an insurer) and 40% on other liabilities. The reasoning is that the collateral on the former type of liabilities is regulated to be more liquid and low-risk and thus higher recovery in case of default can be expected. Kaserer and Klein (2019) provides an overview of the empirical studies which underpin the numbers on the expected recovery rates ERR.

Figure 4.15 (in Appendix (4.C)) shows for each institution the value of the deposits and insurance policies, the value of other liabilities, and the resulting ERR assumptions. There is little variation in the ERR over time but diversity across institutions is large. Table 4.1 below shows the average liability weights (LW), the weight of the institution in the systemic portfolio; the liability ratio (LR), the ratio of deposits and policy liabilities to other types of liabilities, and the ERR per institution. Ranked by LW, the largest institutions are ING Bank, Rabobank, and Aegon. In terms of ERR, Volksbank has the highest value, while Aegon has the lowest (note that RR is a ratio).

Table 4.1: Recovery Assumptions

	LW	LR	ERR
ABN	0.15	0.64	0.66
INGB	0.35	0.72	0.69
RABO	0.23	0.61	0.64
NIBC	0.01	0.57	0.63
VB	0.02	0.83	0.73
AEGO	0.16	0.31	0.52
NN	0.09	0.72	0.69

Note. This table shows the average Liability Weights (LW) over the period 2010-2021 in the regulatory portfolio, the Liability Ratio (LR) as the average ratio of deposits (or respectively policy weights) in the company's balance sheet, and average recovery rate (RR) per company.

4.6 Results: an Overview

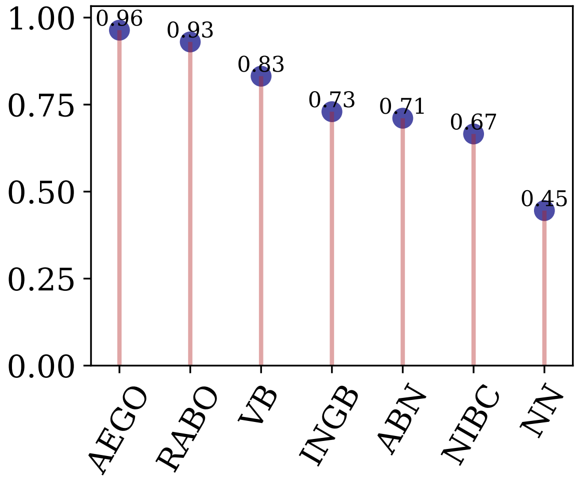
4.6.1 Factor Exposures and Asset Correlations

The first building block for evaluating the potential systemic losses driven by individual institutions consists of the estimation of the latent factor model, as showin in Section 4.3.1. The latent factor, synthesized from the common asset return variation in the sample, is often interpreted as a market driver of risk. From that point of view, the factor loadings on their own provide a useful interpretation as market exposures. They measure the sensitivity of an institution to market movements, and indicate the number of standard deviations the asset return of an institution will fall below the mean in response to one standard deviation drop in the return of the market.

Figure 4.1 shows the exposure values for each institution. The factor loadings are estimated based on the observed CDS prices over the considered time window. We can already see that three groups of institutions start to form - those with high sensitivity to the common factor (Aegon, Rabo, and VB), those with median exposure (ING, ABN, NIBC), and those with low factor sensitivity (NN).

For interpretation purposes, it is also useful to translate the exposure figures into the share of total asset return variation due to market risk vs. the share due to idiosyncratic variation. In fact, squaring the loadings provides the share of market risk:

Figure 4.1: Common Factor Loadings



Note. This figure shows the estimated exposure (loading) of each institution to the common latent factor.

$$\begin{aligned}\frac{\mathbb{Var}(\mathit{\Delta}\ln V_i)}{\sigma_i} &= a_i^2\mathbb{Var}(M_i) + (1-a_i^2)\mathbb{Var}(Z_i) \\ &= \underbrace{a_i^2}_{\text{Factor Risk Share}} + \underbrace{(1-a_i^2)}_{\text{Idiosyncratic Risk Share}}\end{aligned} \tag{4.16}$$

In the same line of thought, cross-multiplying the loadings of two institutions provides the implied correlation between the return of their asset holdings, since we have:

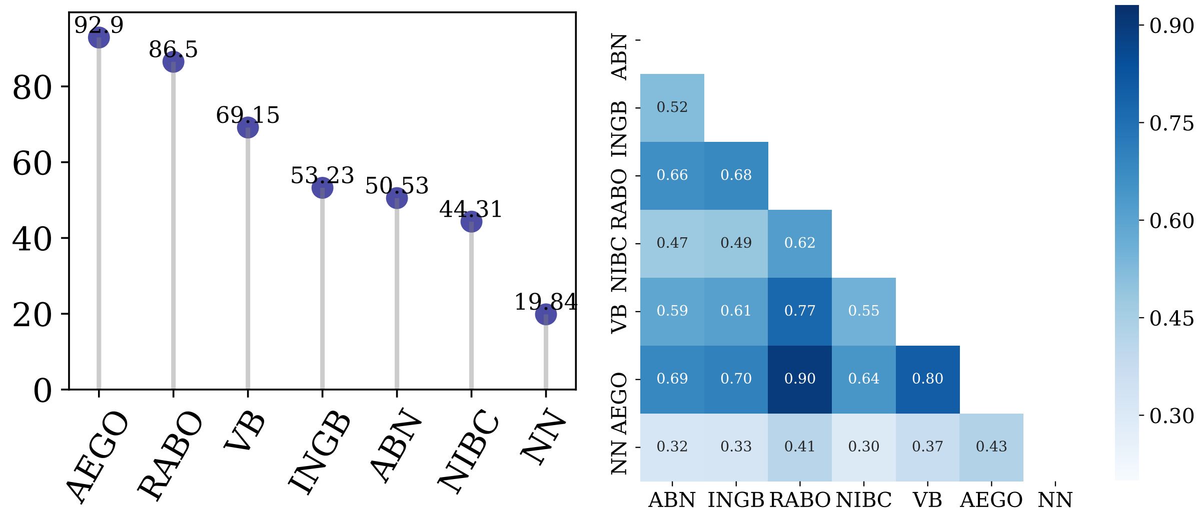
$$\mathbb{Corr}(\mathit{\Delta}\ln V_i, \mathit{\Delta}\ln V_j) = a_i a_j$$

Figures (4.2) below illustrate the results. Naturally, the smaller the factor loading of an institution is, the more the risk of that institution is purely idiosyncratic, and the less correlated it is with all other institutions in the market.

4.6.2 Probabilities of Joint and Systemic Defaults

The next building block of the model is the default simulation based on a fixed default barrier in line with the Merton firm model. To do this we draw 500K independent Monte Carlo simulations for the idiosyncratic and the common factors. Based on each

Figure 4.2: Common Factor Loadings



(a) Share common factor risk (% of Total)

(b) Implied Asset Correlations

institution's factor exposures, outlined in the previous section, these can be translated into scenarios of (standardized) asset value changes over the coming year. The default probability implied through the observed CDS rate for the period provides the default boundary for the institution, as indicated in equation (4.3). Subsequently, in each simulated scenario of asset value drops, we can evaluate whether the barrier would be crossed and whether a default would occur. The common factor provides co-variation in the occurance of defaults, which will guide the probability of multiple defaults occurring at the same time.

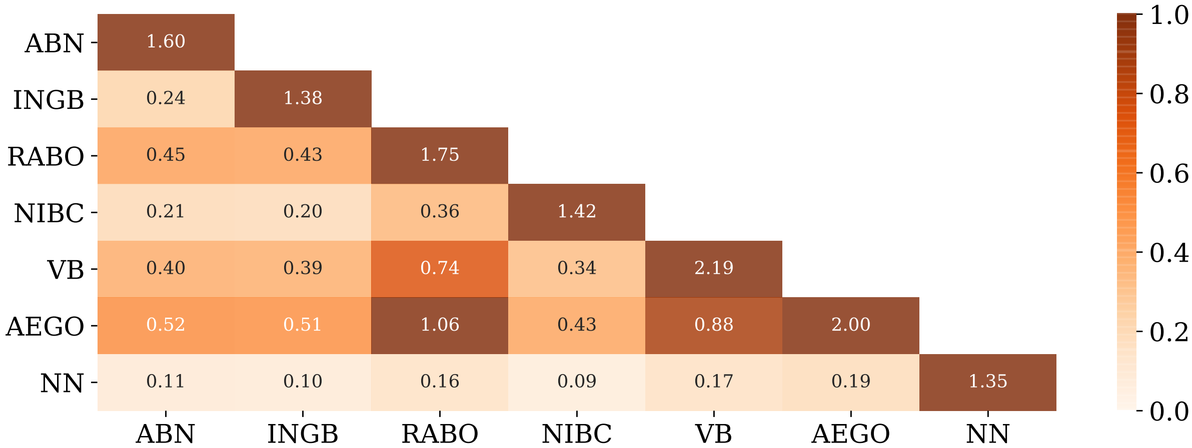
In aggregate, this allows us to estimate the average share of default scenarios per institution, matching the estimated individual default probability from the observed CDS spread. More importantly, we can find the average share of joint defaults, illustrating the tendency of institutions to become distressed at the same time. Figure 4.3a shows these numbers. The diagonal corresponds to the standalone default probabilities. VB, Aegon, and Rabo are ranked highest. The off-diagonal terms show the probability of joint defaults. The three highest pairs here are, maybe not surprisingly, again among the group of institutions that have the highest common factor exposure: Aegon with Rabo, Aegon with VB, and VB with Rabo.

Next, we can translate the joint default probabilities into conditional probabilities of one institution's default conditional on a default of another institution using the definitional relationship:

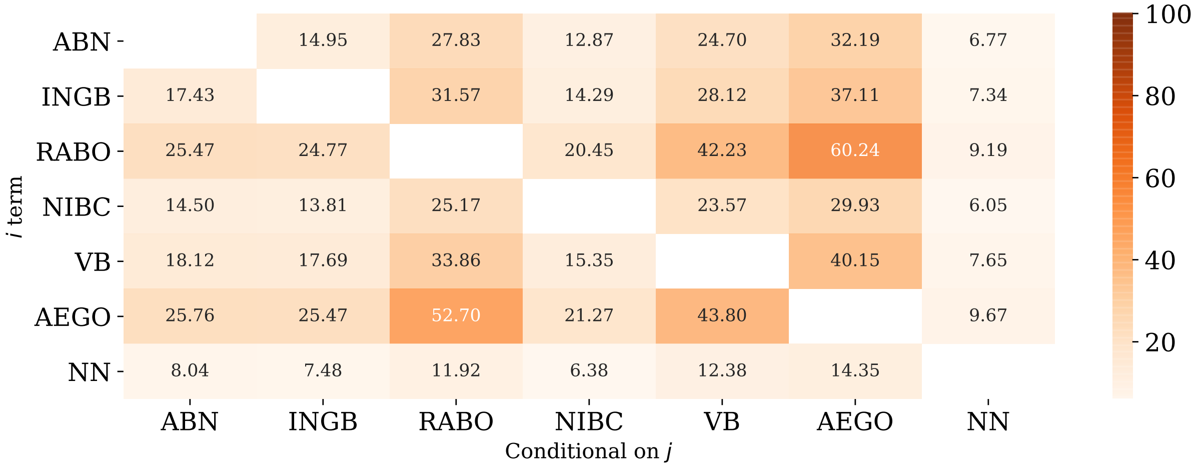
$$\mathbb{P}(\mathbb{1}_{d_i} = 1 | \mathbb{1}_{d_j} = 1) = \frac{\mathbb{P}(\mathbb{1}_{d_i} = 1, \mathbb{1}_{d_j} = 1)}{\mathbb{P}(\mathbb{1}_{d_j} = 1)}$$

where in each case $\mathbb{P}(\cdot)$ indicates the probability of default. Figure (4.3b) below shows the results. We can see that the high asset correlations also translate into high joint and conditional default probabilities.

Figure 4.3: Default Probability Matrix



(a) Joint Probability of Default



(b) Conditional Probability of Default

Note: This set of charts shows (a) the probability that two institutions may default together over a one year horizon; (b) the probability that institution i may default, conditional on j being in default.

We also want to look at the potential for a systemic event to trigger cascading defaults. For the purpose, we define the random variable N_d which will measure the number of defaults that will materialize over the coming period as:

$$N_d = \sum_{i=1}^{N} \mathbb{1}_{d_i} \tag{4.17}$$

The factor-based simulation, outlined in Section 4.3, allows us to evaluate the proportion of cases where more than $k = 1, 2, 3, 4$ defaults happen at the same time. This produces $\mathbb{P}(N_d \geq k_1)$, as summarized in column two in Table 4.2. There is about 7% chance that at least one of the considered institutions may default over the next year. The probability is relatively high, but the overall trend, as Figure 4.4a indicates, has been decreasing since the 2008/09 financial crisis.

We also compute the probability that there are more defaults, given that one or two have already materialized. Using the law of conditional probabilities, these can be computed as

$$\frac{\mathbb{P}(N_d \geq k)}{\mathbb{P}(N_d \geq \overline{k})}$$

where k is the total number of defaults given that at least $\overline{k}$ have already happened. The results are reported in column three and column four of Table 4.2 for $\overline{k}$ equal to one or two, respectively. If a default occurs, there is a substantial chance (about 34%) that other defaults may follow. Examining the trend of conditional defaults over time, Figure 4.4b shows that the cyclical pattern here is different - the probability decreases after the Euro government debt crisis in 2010-2011, increases in 2016, possibly due to Brexit concerns, and spikes suddenly in March 2020, which is when the first Covid waves came up in Europe.

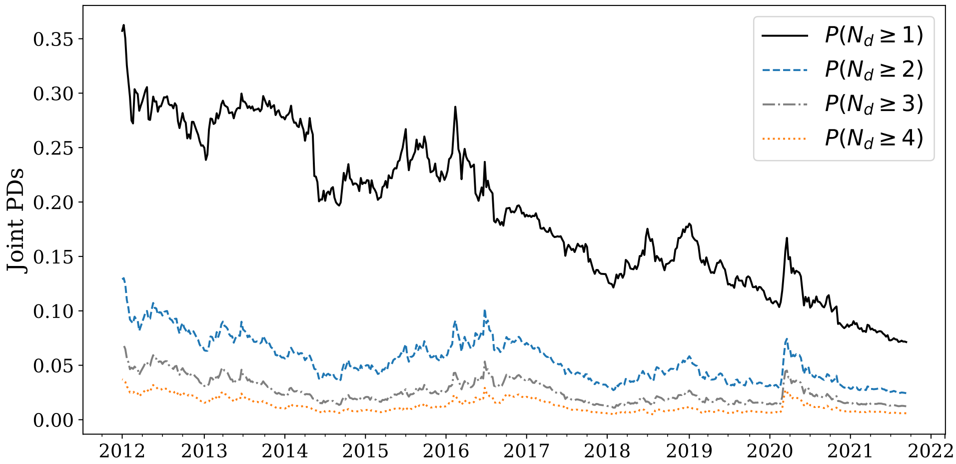
It is worth noting that, since our model is not identifying causality in any form, the conditional probability of additional defaults could stand either for potential spillovers from one distressed bank to another, or could represent a common external shock affecting multiple institutions.

Table 4.2: Probability of Systemic Defaults

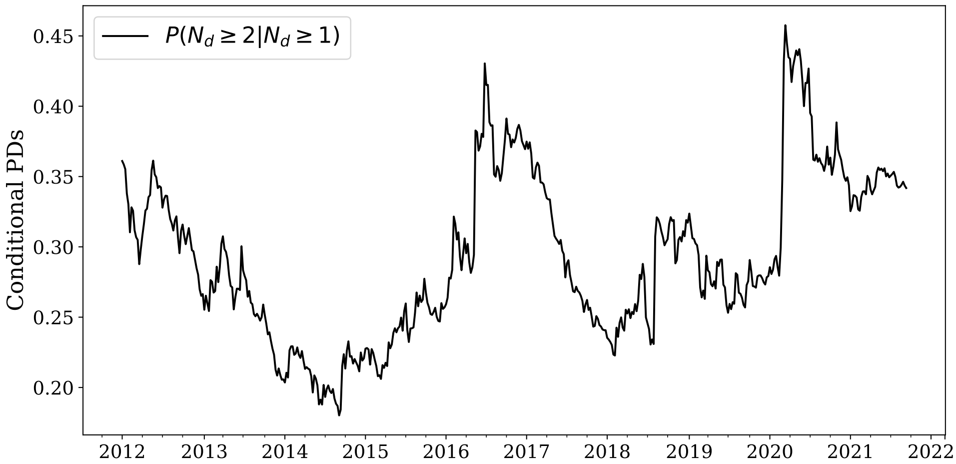
k	$\mathbb{P}(N_d \geq k)$	$\mathbb{P}(N_d \geq k \| N_d \geq 1)$	$\mathbb{P}(N_d \geq k \| N_d \geq 2)$
1	7.17	-	-
2	2.50	34.88	-
3	1.24	17.27	49.50
4	0.61	8.52	24.43

Note. The first column in the table indicates the threshold number of defaults k. The second column shows the unconditional probability that more than k out of 7 firms default at the same time. The third and fourth columns show respectively the conditional probability that more than one or two additional institutions will default given that at least one or respectively two have already defaulted.

Figure 4.4: Probability of Systemic Defaults



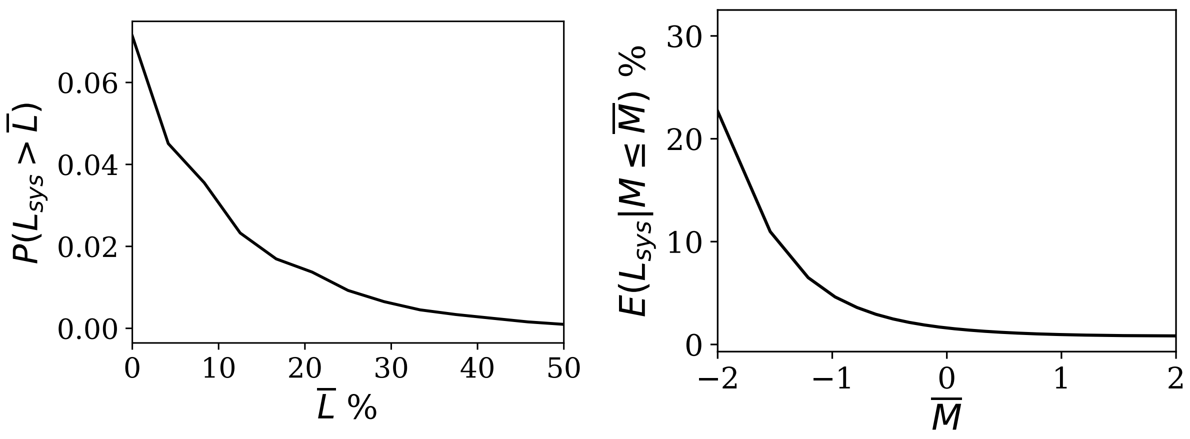
(a) Unconditional



(b) Conditional on One Default

Note. This set of charts shows (a) the probability that more than one, two, three or four defaults occur at the same time; (b) the probability that more than two defaults could occur if one has already happened.

Figure 4.5: Systemic Default Loss Distribution.



(a) Systemic Losses

(b) Factor-driven Expected Loss

Note. The first figure presents the probability that systemic losses (as percent of total liabilities in the system) will be larger than a threshold $\overline{L}$ as a function of that threshold. The second figure shows the expected systemic loss, as percent of total liabilities in the system, conditional on the factor dropping by more than $\overline{M}$ standard deviations away from the mean.

4.6.3 Marginal Expected Shortfall

Evaluating only systemic default probabilities as was done in Section 4.6.2 does not take into account the fact that the default of some institutions may have a much larger impact than that of others. Everything else fixed, bailing out a larger institution, will be more costly for the regulator, and its default and the possibility that it cannot cover its liabilities will have a wider impact on the economy. A proper systemic risk appraisal should also capture the size of the potential losses given that joint distress occurs. So our next step is to assess the size of the expected losses if tail risk events do in fact happen. Now we need to incorporate the stochastic nature of expected losses and the way they are correlated across institutions and, equally important, the way they depend on default probabilities, following the approach outlined in Section 4.3.2.

The charts in Figure 4.5 show an initial view into the potential size of the systemic losses and the probability that such losses could be realized. First, Figure 4.5a shows the distribution of the simulated cumulative losses as a percent of the total outstanding liabilities in the system. Consistent with the earlier estimates in Table 4.2, in about 93% of the cases (corresponding to $1 - \mathbb{P}(L_{sys} > 0)$), the system is resilient and does not encounter any distress which would lead to default losses for any of its composite institutions. In about 3% of the cases, systemic losses could be above 10%.

Figure 4.5b on the other hand, relates the size of the aggregate losses that can be expected to the size of a potential systemic shock driven by a drop in the common factor M. By our estimates, for example, a drop of more than two standard deviations in the latent factor (e.g. a shock of the magnitude of the 2008/09 Financial Crisis) can be expected to bring losses of about 20% of the size of aggregate liabilities. In other words, given the defaults that this large systemic shock would generate, some of the financial institutions will experience asset value drops by such a magnitude that they will not be able to deliver about 20% of their outstanding liability commitments.

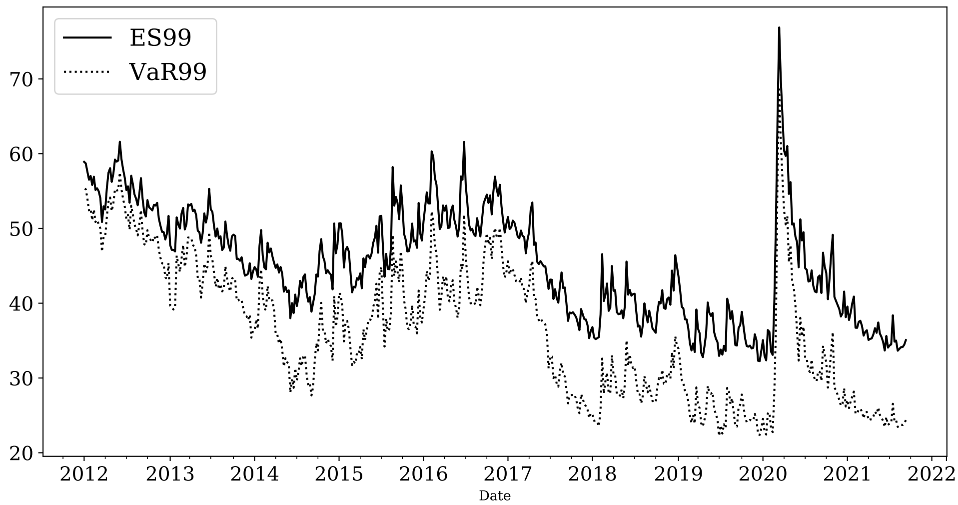
A standard approach to quantify the risk of losses within a portfolio in a single number is to employ the downside risk measures defined in Section 4.4. In particular, we use the Expected Shortfall to measure the average of the worst 1% of the possible outcomes for the coming year. Using the simulated systemic losses, we evaluate the systemic risk by the ES of the portfolio of institutions, and arrive at an estimate of 35.05% (cf Table 4.3). Table 4.3 summarizes the risk for the system and for each individual institution, where the ES will be indicative of standalone risk.

For the system as a whole, Figure 4.6 puts the risk evaluation in context, plotting over time the ES and the VaR of the systemic portfolio. Note that in contrast to the downward trend of Figure 4.4a, the tail risk of the systemic portfolio is not on a downward trend over time but seems to be more in line with the cycles in Figure 4.4b.

Overall, these estimates allow regulators not only to track the resilience in the system and to look for increased probability of large systemic losses, but also to verify whether the buffers currently in the system are enough to cover the potential losses once a systemic event materializes. If the buffers are not sufficient, regulators could either require higher buffers to be set aside, or could look for ways to increase the resilience of the system, by closely examining the institutions which are the highest contributors to risk.

In order to attribute the potential systemic losses to the individual institutions comprising the system, we employ the MES measure suggested by Acharya et al. (2017) and defined in (4.13). In particular, we look at the percentage contributions defined in (4.15).

Figure 4.6: Systemic Risk (ES, VaR)



Note. This plot shows the tail risk of the systemic portfolio quantified by the ES and VaR.

Table 4.3 summarizes the results. First of all, it points to a certain discrepancy between bank vulnerability rankings in relation to banks' own risk, as measured by $ES99$, and in relation to the vulnerability of the system as a whole, as measured by $MES99$. The two riskiest institutions on their own are Aegon and Rabo. They are also the most sensitive to shocks in the system, followed by VB which is number five when ranked by ES.

Second, taking institutional size into account, the top three contributors to systemic risk shift. Ranking by *PC to ES*, Rabo becomes first, owning about 32% of the total systemic risk, followed by ING with about 28%, Aegon, an insurer, with 26% and ABN with about 9%.

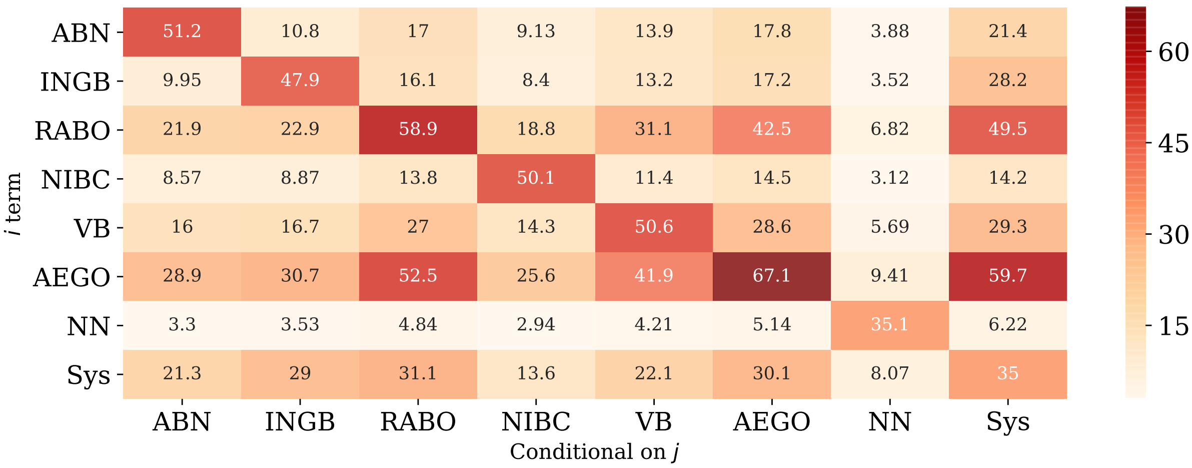
We can also define a so-called network relation based on ES as:

$$NES_{i,j} = \mathbb{E}\left(L_i | L_j \geq VaR_j\right) \tag{4.18}$$

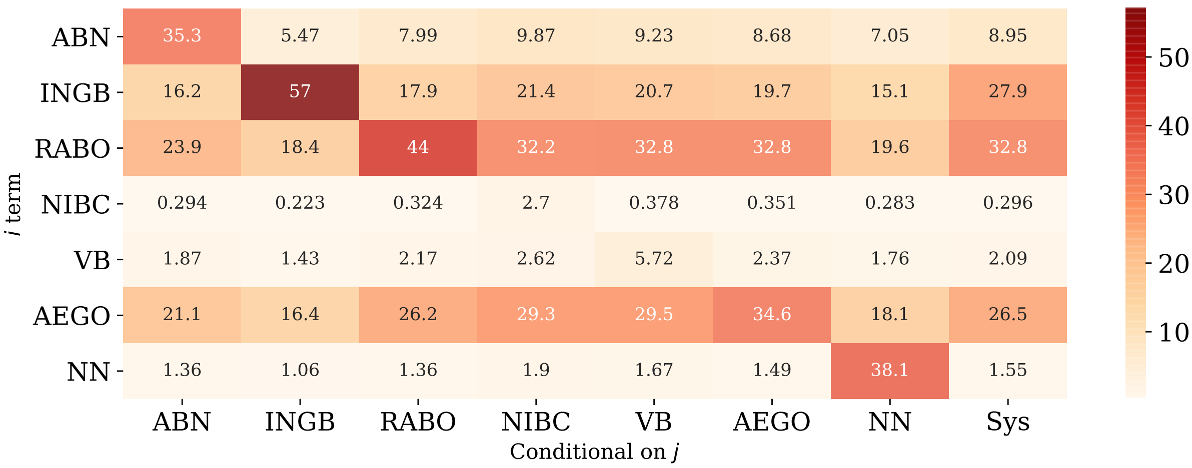
to measure the average losses of institution i when institution j is in distress. In contrast to the asset correlations and the conditional default probabilities, this distress dependency metric also takes into account the size of the losses.

Figure 4.7a then shows the expected loss of the row entry given that the column entry is below its 99% VaR, where the diagonal of the table corresponds to the ES of each entity, and the last column corresponds to the MES of each institution. More interesting

Figure 4.7: Network Expected Losses, $q = .99$



(a) Network ES



(b) Network PC to ES

Note: These set of charts illustrate the network effects of tail losses. The first chart shows the expected loss of the i entry, conditional on the j entry. The last column and the last row, labeled Sys, stand for Systemic losses. The diagonal of the table corresponds to the ES of each entity, and the last column corresponds to the MES of each institution, the last row measures the $CoES$ of the column item, and the off-diagonal terms measure $NES_{i,j}$ with i as the row entry and j as the column entry.
The second chart shows the percentage contributions to systemic losses given that the column item is in its tail. Column items sum up to 100% and can be interpreted as percentage contributions to $CoES$.

Table 4.3: Systemic Risk Statistics

	EL		w		ES99		MES99		PC to ES99	
ABN	0.73	(4)	14.67	(4)	51.21	(3)	21.38	(5)	8.95	(4)
INGB	0.60	(6)	34.57	(1)	47.87	(6)	28.24	(4)	27.86	(2)
RABO	0.97	(2)	23.24	(2)	58.88	(2)	49.45	(2)	32.80	(1)
NIBC	0.65	(5)	0.73	(7)	50.07	(5)	14.22	(6)	0.30	(7)
VB	0.94	(3)	2.50	(6)	50.56	(4)	29.33	(3)	2.09	(5)
AEGO	1.29	(1)	15.54	(3)	67.15	(1)	59.65	(1)	26.46	(3)
NN	0.47	(7)	8.75	(5)	35.09	(7)	6.22	(7)	1.55	(6)
System	0.81		100.00		35.05		35.05		100.00	

Note. This table shows the Expected Loss, Liability Weight, ES, MES, and Percentage Contribution to ES statistics. All statistics are in percentage loss format. The numbers in the brackets provide the ranking relative to the group.

are the off-diagonal entries, which quantify loss co-dependencies between the different institutions, corresponding to the $NES_{i,j}$ measure defined earlier: the average loss of institution i given that j is in distress. In contrast to the asset correlations and the conditional default probabilities, this distress dependency metric takes into account the size of the losses. The largest co-dependent loss here occurs between Aegon and Rabo. In particular, if Rabo is in its tail, Aegon would lose 52.5%, which is not far from the loss it would realize in its tail, an ES of 58.9%.

Note that the last row of Figure 4.7a shows, the average loss of the system given that the institution is in its tail, a measure which we can call the $CoES$. This is an inverted version of the MES presented in (4.13). The additivity property of the expectation combined with common conditioning, allows us to break down the $CoES$ of an institution into its weighted network components:

$$\begin{aligned} CoES_j &= \mathbb{E}(L_{sys}|L_j \geq VaR_j) \\ &= \mathbb{E}(\sum_i w_i L_i | L_j \geq VaR_j) = \sum_i w_i NES_{i,j} \end{aligned} \tag{4.19}$$

One possible interpretation of the $CoES$ is as a stress scenario, measuring the expected systemic loss if one institution becomes distressed. Note that all other institutions in that scenario will not be held fixed but will react following their tail dependencies with the distressed institution. A higher value for the metric means that either the distressed entity has more impact on the system on its own, or that all institutions are strongly correlated when generating losses. In the extreme, if either an entity is driving all the losses in the system, or the entity's losses are fully correlated to those of other players, its $CoES$ will be equal to the ES of the system. From that point of view Rabo, Aegon, and ING Bank respectively have the highest potential to impact the system.

Weighting up and adding up all $NES_{i,j}$ over i, i.e. all row entries over a column entry j, generates the $CoES$ of institution j, as indicated earlier in (4.19). This allows us to determine the percentage contribution each institution will bring about to systemic losses when one of its peers is in its tail. Formally, we have

$$\text{Network } PC \text{ to } ES_{i,j} = \frac{\sum_i NES_{i,j}}{CoES_j}$$

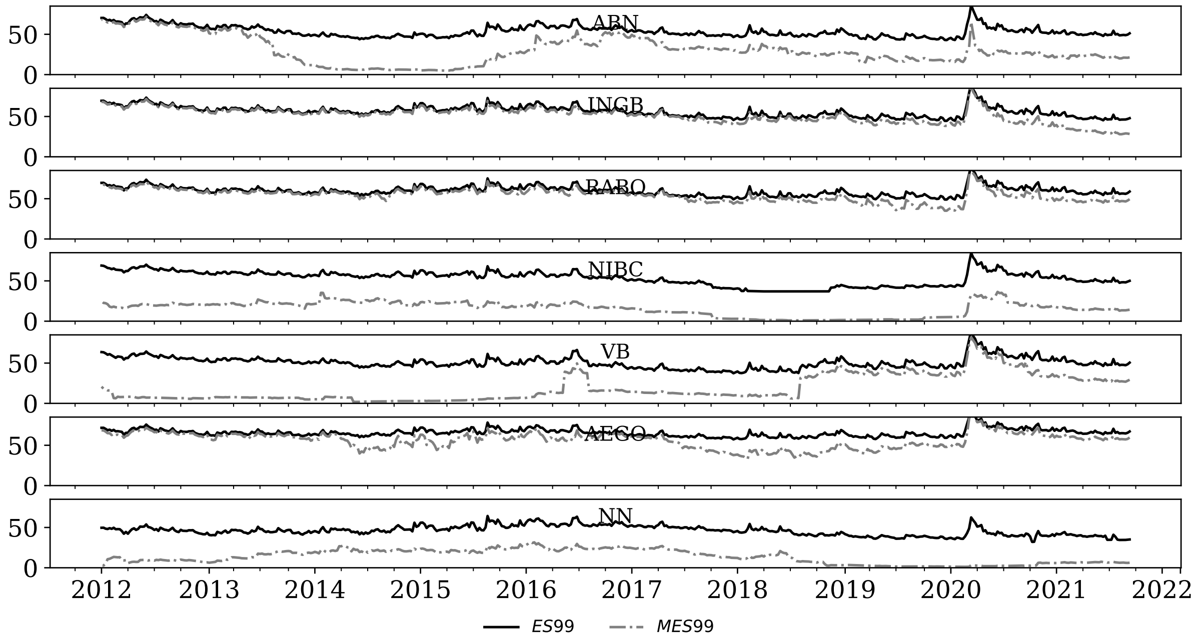
The highest systemic loss of 31.1% will occur if Rabo ends up in its tail. If that were to happen, we can see from Figure 4.7b that only 44% of the systemic losses will be due to Rabo itself. The rest will be contributed in large part by Aegon (26.2% of the risk), ING (17.9%), and ABN (about 8%).

Figure 4.8 shows rolling-window MES estimates for each institution. Note that the ES, as a measure of standalone risk, is not always moving in conjunction with the systemic contributions (MES). Periods, where the two disagree in direction, are indicative of changes in the correlation between the institution and the portfolio.

The Covid impact spike at the beginning of 2020 is a clear systemic event increasing the standalone risk of each company. For some companies, the standalone risk spikes together with the corresponding contribution and reverses a former trend of declining contributions (Rabobank, NIBC, Aegon). For NIBC, in contrast, the spike keeps the firm at an increased level of systemic contribution. For ABN and VB, the shock seems to be largely transient. An interesting case is NN. Even though it experiences an uptick in standalone risk, its contribution does not move.

Figure 4.9 shows the contributions to systemic risk over time as a share of total systemic risk. The companies are ranked on the chart by contribution as of the end of 2021. There is no change in ranking for the top three contributors over time: Rabobank,

Figure 4.8: Backtest, 99% MES



Note. This figure shows the MES vs. the ES at $q = .99\%$ for each company. The MES of the institutions sums up to the total ES of the system.

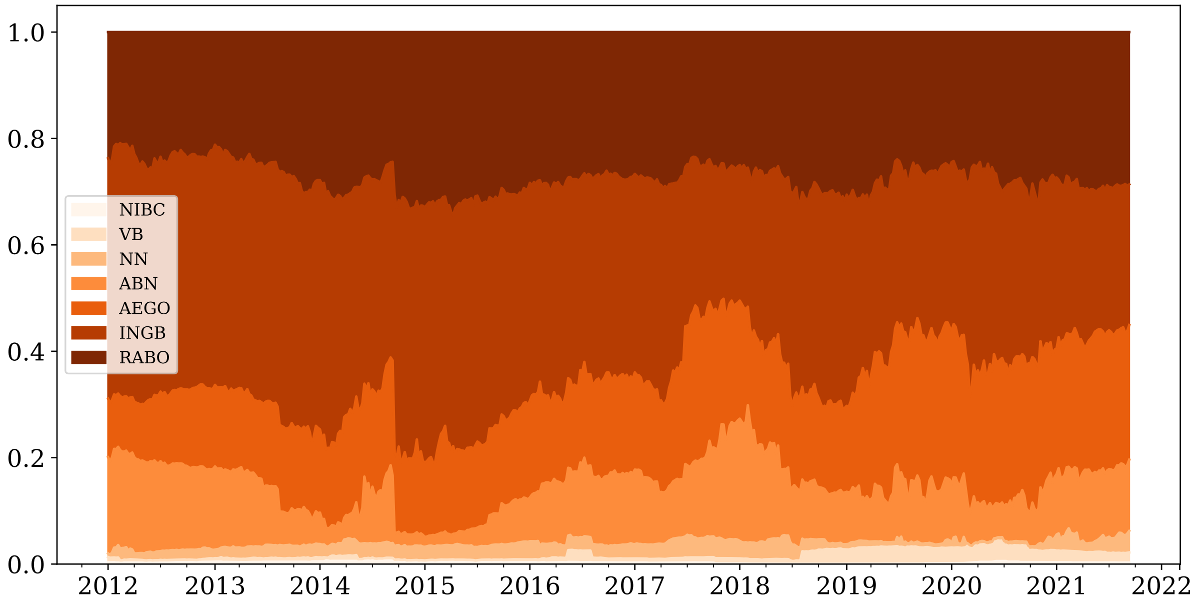
ING Bank, and Aegon. However, the relative contribution coming from ING Bank has diminished after 2017.

4.6.4 Robustness

Finally, as a robustness check, we employ two checks to verify if changing or isolating out some components of the model changes the systemic rankings we estimate. First, we look at whether several other measures of systemic risk sensitivity comply with the rankings established in Section 4.6.3. Second, we look at the sensitivity of our results relative to the parameter assumption of σ_c underlying the RR dynamics in (4.6).

As indicated in Section 4.2, there is a wide variety of measures in use that imply the systemic risk sensitivity of an institution from market data. Adrian and Brunnermeier (2016) propose the $CoVaR$ to quantify the tail-dependency between an institution and the system it is part of. It is evaluated as the worst $q\%$ losses of the system, given that an institution i is in its worst $q\%$. To align this measure with the concept underlying the MES, we invert it to get the Exposure $ECoVaR$, which now also quantifies the sensitivity of the institution's losses to a systemic tail event[17]:

Figure 4.9: 99% MES (Percentage Contribution)



This figure shows the MES vs. the ES at $q = .99\%$ respectively for each company. The MES of the institutions sums up to the total ES of the system.

$$\mathbb{P}(L_i \geq ECoVaR_i | L_{sys} \geq VaR_{sys}) = q$$

Both the MES and the $ECoVaR$ measure the institution's losses if the system ends up in the tail of its potential losses over the coming year. However, in contrast to MES, which measures the *average* loss once the system is its tail, the $ECoVaR$ zooms in deep in the tail of the potential losses of the institution, measuring the q-th quantile not only with respect to the systemic losses but also with respect to the institution's losses.

Next, we relate to another measure, which focuses only on default correlations as presented in Section 4.6.2. The idea is to compare our results to a measure that is not

[17]Note that originally Adrian and Brunnermeier (2016) define $CoVaR$ by conditioning on individual losses being equal to a quantile rather than a region of their distribution as:

$$\mathbb{P}(L_{sys} \geq CoVaR_i | L_i = VaR_i) = q$$

This allows the use of quantile regression for the estimation of the measure. On the negative side, such conditioning can give a misleading tail-risk indication when the loss distribution is fat-tailed, by not capturing the probability mass below the VaR quantile. In our case, systemic losses are strongly non-Gaussian, so we use the modified version of CoVaR, as in Huang et al. (2012), which conditions on $L_i \geq VaR_i$. See also Nolde and Zhou (2021) for the same argument, and the relation to Extreme Value Theory of the modified measure.

influenced by the assumptions on how losses are formed and how they correlate between companies, as these assumptions will inevitably affect both the MES and the $ECoVaR$ which are driven by the same loss simulations. To measure the sensitivity of individual institutions to distress in the system, Zhou (2010) defines the Vulnerability Index (VI) as the probability that institution i will be in default conditional on more than one default in the system:

$$VI_i = \mathbb{P}(\mathbb{1}_{d_i} = 1 | N_d > 1) \tag{4.20}$$

Note that Zhou (2010) relies on Extreme Value Theory to estimate the proposed measures. Also, we rely on default as an indication of distress, whereas the original measure is constructed to capture large tail movements in the equity value of the institution. Our approach differs methodologically, but based on the outlined model, the measures can be adapted.[18]

Table 4.4 summarizes the results. First, we compare the unweighted systemic risk measures to the MES rankings of Table 4.3. The rank correlations in sub-table (b) indicate strong agreement between the VI measure and the MES. The MES and the $ECoVaR$ agree only moderately. Closer inspection indicates that the disagreement is that $ECoVaR$ switches the ranks of NIBC and VB, and of ABN and ING.

Sub-table (c) compares the size-weighted measures and finds strong agreement between them. It is worth noting that only the weighted MES measure (the ranking of which coincides with the presented earlier ranking by PC to ES) sums up to total systemic risk. This is due to the additivity property (4.14). The two other risk measures do not have this property and their weighting can be considered only as a heuristic.[19]Once the measures are weighted, they show a stronger correlation among each other. It is worth noting that size itself correlates strongly to MES. It fails however to identify Rabo as

[18]The VI index is constructed by inverting an earlier measure of conditional default proposed by Segoviano and Goodhart (2009). To evaluate the impact of each institution upon the system, they measure the probability that at least one more institution becomes distressed (PAO) conditional on the distress of one particular institution: $PAO_i = \mathbb{P}(N_d > 1 | \mathbb{1}_{d_i} = 1)$. We do not explore systemic impact measures here, as an initial analysis shows that there is very little difference in rankings between the impact measures (PAO and SII) and the sensitivity measure (SII) for our sample.

Table 4.4: Systemic Rankings Comparison

	$ECoVaR$ (%)		VI (%)		$w \cdot ECoVaR$		$w \cdot VI$	
ABN	64.94	(4)	33.81	(4)	9.52	(4)	4.96	(4)
INGB	63.65	(5)	32.12	(5)	22.00	(1)	11.10	(2)
RABO	67.73	(2)	57.33	(2)	15.74	(2)	13.33	(1)
NIBC	65.39	(3)	28.06	(6)	0.48	(7)	0.21	(7)
VB	62.41	(6)	53.26	(3)	1.56	(6)	1.33	(6)
AEGO	73.95	(1)	67.00	(1)	11.49	(3)	10.41	(3)
NN	56.72	(7)	15.41	(7)	4.96	(5)	1.35	(5)

(a) Rankings

	MES	$ECoVaR$	VI
MES			
ECoVaR	0.64		
VI	0.96	0.68	

(b) Rank Correlations

	w	PC to ES	$w \cdot ECoVaR$	$w \cdot VI$
w				
PC to ES	0.93			
$w \cdot ECoVaR$	1.00	0.93		
$w \cdot VI$	0.96	0.96	0.96	

(c) Rank Correlations, Weighted Measures

Note. This set of tables shows the systemic risk rankings according to alternative measures and the rank correlations between them.

the major contributor to systemic risk. Yet, both the PC to ES and the weighted VI find Rabo to be more systemic than ING.

Next, we vary the parameter σ_c to verify to what extent the results are driven by the decision to calibrate the parameter to the VSTOXX index. Table 4.5 shows the resulting percentage difference in the 99% MES estimates when a fixed number of .15 and then to .05 is used in the model, relative to the base figures in which the value of the VSTOXX index was used (at the reference date the index has a value of .2). The parameter choice

[19] To see how the $CoVaR$ can be broken down into components that satisfy the additivity property see Puzanova and Düllmann (2013). For general discussion on the additivity of risk measures see the Euler property in e.g. Chapter 12 of Hull (2018).

Table 4.5: Percentage Change in MES when Varying σ_c

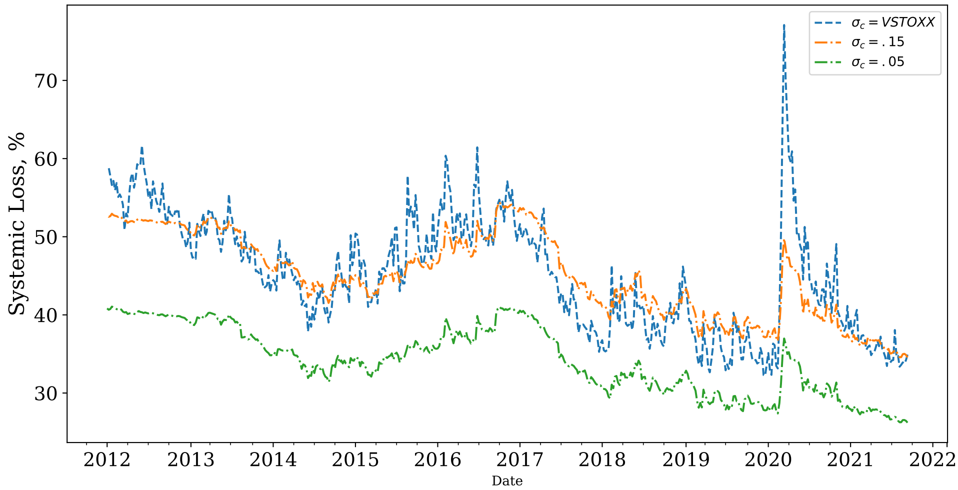
	$\sigma_c = .15$	$\sigma_c = .05$
ABN	8.22	24.31
INGB	8.22	22.54
RABO	9.50	30.04
NIBC	7.41	21.29
VB	11.36	35.07
AEGO	6.76	20.89
NN	6.25	17.27
Sys	8.28	24.90

Note. This table shows the percentage difference in MES estimates, relative to the base case from Table 4.3, when arbitrary fixed values are used to calibrate the σ_c parameter.

affects the magnitude of the MES estimates. The new MES figures however do not change the systemic ranking.

Figure 4.10 shows the ES of the system for each alternative σ_c estimate. Again, the magnitude of the tail risk values change, but the overall trends do not. Using VSTOXX also makes the estimates more sensitive to short-term variations. Having a reliable estimate of the individual institutions' asset variance will make a difference in differentiating better between their risk characteristics or in defining the magnitude of the possible losses. In absence of such data, however, using a single number matched to the implied volatility of European stocks seems an appropriate second-best alternative that does not enforce ranking changes in systemic risk across the observed universe.

Figure 4.10: Expected Shortfall of the Systemic Portfolio



Note. This figure shows ES for the system calculated based on two assumption: σ_c is scaled by the implied volatility of the VSTOXX index, and σ_c is fixed to 20%. Evaluated at .95 confidence level.

4.7 Policy Relevance

Our results are relevant for the policy framework for systemic risk. For Global Systemically Important Banks (G-SIBs), the Basel Committee on Banking Supervision and the European Banking Authority (EBA) specify in detail the methodology according to which capital surcharges are allocated to institutions that are designated as systemically important. The goal of the surcharges is to improve the resilience of the system by internalizing the systemic risk generated in the financial sector. In addition, the European Union designates banks that it regards to be systemically significant as Other Systemically Important Institutions (O-SIIs), and requires that national authorities, under EBA guidance, decide on identification procedure and on the size of the surcharges for these institutions.

There is currently a large disconnect between the academic approaches used to measure systemic risk and the regulatory approach used to set systemic capital buffers. Section 4.2 has extensively explored the academic perspective. For European regulators, the general guidance by the EBA is to focus on four criteria of systemic relevance: size, importance, complexity, and interconnectedness (EBA, 2020). Usually, a score is provided in each category and the four categories are equally weighted up to a single O-SII score. The ranked institutions are bucketed based on score ranking, and for each bucket, systemic buffer surcharges are discretionary set through a step-up ladder structure.

As we focus on institutions residing in the Netherlands, we can directly compare the systemic ranking coming out of our model to the ranking based on the O-SII surcharge rate set by the Dutch regulator. In the Netherlands, as of 2021 the following O-SII buffers apply[20]: ING Bank (2.5%), Rabo (2%), ABN (1.5%), Volksbank (1%). As Table 4.3 indicates, this ranking differs from our ranking by PC to ES, where Rabo comes before ING. Ranking by size, however, we match the O-SII results. It is difficult to generalize based on our small sample, but this could be an indication that the O-SII score is not putting enough weight on the interdependency between the institutions, and is focused more either on size or on standalone risk, where ING ranked on top.

Naturally, the sample that we have is too small to allow us to generalize any conclusions. Yet, we relate to other studies that find a difference between the policy and the academic approaches on measuring systemic risk. For example, Brogi et al. (2021) compare the G-SIB buffer rankings to systemic risk rankings calculated based on a credit portfolio approach similar to ours. They use the DIP measure provided by Huang et al. (2009, 2012) which calculates the average loss (rather than the ES) for the regulatory portfolio, where loss is again generated in default. They find significant differences in the two approaches and argue that the regulatory framework would benefit by incorporating also a risk contribution metric into generating systemic rankings.

Bianchi and Sorrentino (2021), on the other hand, explore a small sample consisting of the four Italian banks designated as systemically important and largely find consistency in the ranking based on the CoVaR measure and based on the O-SII buffer rates set by the Italian central bank. Yet, having higher frequency data allows them to link systemic risk estimates to the evolution of bank characteristics and conditions.

Overall, we can conclude that there is no downside to embedding market-based implied measures of systemic risk, as ours, into the policy process. First of all, such measures could provide a way to verify the ranking that policymakers come up with based on EBA's guidelines and using regulatory data. Any discrepancy in the rankings based on the two approaches could raise important questions, the answers of which could improve

[20]See https://www.esrb.europa.eu/national˙policy/systemically/html/index.en.html

the regulator's approach to assessing systemic risk. Or even if no discrepancy between the two appears, a market-based measure as the MES can be used to assess risks between annual policy assessments.

4.8 Conclusion

In this paper we examined the systemic linkages and the potential systemic risks arising in the Dutch financial sector. In particular, we look at seven key insurance and banking institutions. To do so we addressed a common challenge in estimating and monitoring the build-up of systemic risks: a regulator cannot resort to equity prices for institutions that are privately or state-held. In these cases, we show how high-frequency data from the CDS market can still be used to imply views on co-dependencies and joint losses. We use the Dutch financial sector as a case study for our approach.

We argue that in contrast to micro-prudential policies, an appropriate macro-prudential view should try to monitor and manage not only the risky positions of an institution on its own, but also the interdependencies between institutions and the potential for several of them to realize large losses at the same time. From that perspective, in the sample that we consider, we confirm that a risk ranking incorporating tail dependence across the institutions is different from a ranking based on standalone tail risk. From a risk management point of view, it is clear that a focus on the former is more important if the goal is to curb risk in the total portfolio. Yet, we find in our sample that the latter is closer to the current ranking based on the O-SII capital surcharges for systemic risk.

In the process, we presented a model, which builds upon the existing academic literature that addresses systemic risk from a structured credit angle (Huang et al., 2012; Puzanova and Düllmann, 2013). The financial institutions in the system are seen as part of a defaultable loan portfolio. Systemic losses occur in the case of default of one or several institutions. The average tail losses of the portfolio (the ES measure) speak for the magnitude of the systemic risk. The average losses of each institution, given that the system is in its tail, speak for the sensitivity of each institution to systemic risk. The share of the portfolio tail risk that can be attributed to each institution speaks for the contribution of the institution to systemic losses. We extend the existing approaches by also incorporating dependency in the size of the losses, and not purely in the default probabilities.

Our research speaks directly to the policy debate around the risk rankings used to set additional buffer charges for systemic risk for banks. We find certain differences in the top three institutions ranked as most systemic by our portfolio-based approach and the regulatory approach. The sample that we consider is too small to draw general conclusions but may indicate a disconnect between how regulators measure systemic importance, and what market co-movements in the price of default protection imply. A natural extension of the current study would be to expand the universe of institutions that are considered and to observe if those rankings systematically differ across European countries. The O-SII buffer rates in Europe are mandated separately by each national regulator, each following its own implementation of the EBA guidelines.

It needs to be acknowledged that there is currently little theoretical backing on determining the size of the capital buffers that institutions need once they are designated as systemic. The policy approach has been to recommend a two-step heuristic, where in the first step institutions are ranked based on a set of criteria associated with systemic importance, and in the second they are bucketed together and surcharges are set in a step-up manner to each bucket. This holds both of O-SIIs and for G-SIIs. Previous studies have found that the approach is very sensitive both to the ranking and bucketing methodologies used (Brogi et al., 2021). In the methodology that we propose, it is natural to link the size of the capital surcharges directly to the possible systemic contributions, measured by the weighted MES. Further research is needed to determine what mapping between the two would be socially optimal.

A larger sample would also allow the exploration of additional features in the systemic risk model. In fact, the currently proposed portfolio approach could be considered a basic architecture, which is extendable to incorporate specific observed stylized features of asset prices or of the structure of the examined financial network. Since tail correlations between the institutions are a key driver of systemic contributions, it is worthwhile exploring non-linear structures of these dependencies. The ability to model large multi-dimensional dependencies is key. Oh and Patton (2018) for example suggests the use of a factor Copula approach. Wang (2021) suggest a deep learning approach. Alternatively, network models could be used to mimic the often observed core-periphery structure of the financial sector (Bräuning and Koopman, 2016; Andrieş et al., 2022). Institutions that constitute the core of the network could be dominant drivers of systemic risk (Glasserman and Young, 2016; Jackson and Pernoud, 2021).

To sum up, estimating systemic risk contributions properly is essential for the efficient regulation of the financial system. The additional capital surcharges are a cost that needs to go to the institutions generating the systemic externality, so identifying these institutions is crucial. More research into the methods used to quantify and attribute systemic risk is thus important.

4.A The Merton model of firm value

We present the baseline model of Merton (1974). The value of assets is an exogenous process following a Geometric Brownian Motion

$$dV_t = \mu_v V_t dt + \sigma_v V_t dW_t \tag{4.21}$$

The company's debt is a zero-coupon bond with maturity T and face value of D and default can occur only at the moment when it matures. If the value of its assets are below the face value of its debt, the owners of the company will prefer to succumb to default. If the value is above the value of debt, the owners retain any residual value.

$$E_T = \max[0, V_T - D]$$

In that case, the value of equity at maturity can be seen as a long European call option on the firm's assets, and before maturity can be evaluated through the Black-Scholes (BS) equation as:

$$E(t, V_t, \sigma_v) = V_t N(d_1) - D^* N(d_2) \tag{4.22}$$

where $D^* = De^{-r\Delta t}$ is the value of debt, discounted at the risk-free rate r, $\Delta t = T - t$ is the time until debt maturity, $N(.)$ stands for the standard normal distribution, and d_1 and d_2 are given as follows: [21]

$$d_1 = -\frac{\ln\left(\frac{D^*}{V_t}\right)}{\sigma_v\sqrt{\Delta t}} + \frac{1}{2}\sigma_v\sqrt{\Delta t} \tag{4.23}$$

$$d_2 = d_1 - \sigma_v\sqrt{\Delta t} \tag{4.24}$$

In the Merton framework under the risk-neutral measure d_2 corresponds to Distance to Default (DD), a measure often used to assess the credit risk of a firm. Loosely speaking it measures the number of standard deviations of the asset value of the firm to the default barrier point (Chan-Lau and Gravelle, 2005). The risk neutral survival probability in the Merton model can be shown to be $P(V_t > D) = N(d_2)$, and $N(-d_2)$ is the default probability, where d_2 is then the risk-neutral Distance-to-Default measure DD.

Both the asset value and the asset volatility are unknown. As $E = E(t, V_t, \sigma_t)$ is a function of the stochastic underlying asset value, applying Ito's rule we can write the default probability as:

$$dE = \left(\frac{\partial E}{\partial t} + \mu_v V \frac{\partial E}{\partial V_t} + \frac{1}{2}\sigma_v^2 V^2 \frac{\partial^2 E^2}{\partial V_t^2}\right) dt + \sigma_v V_t \frac{\partial E}{\partial V_t} dW_t$$

The standard approach in calibrating the model, relying on Ronn and Verma (1986), notes that stock prices E_t are themselves observable on the market and follow a Geometric Brownian motion of the type

$$dE_t = \mu_E E_t dt + \sigma_E E_t dW_t$$

Matching coefficients in the diffusion term with (4.22) we get

$$\sigma_E E_t = \sigma_v V_t \frac{\partial E}{\partial V_t} \tag{4.25}$$

We can then solve the system of two equations and two variables defined by (4.22) and (4.25). In Merton's setting, we get that $\frac{\partial E}{\partial V_t} = N(d_1)$, where the derivative is also known as the option delta in option pricing theory.

At any time, the value of assets can be decomposed by sources of financing into debt and equity, so we can write the current market value of its debt as

$$B_t = V_t - E_t \tag{4.26}$$

which, using (4.22), can also be written as

$$B(t, V_t, \sigma_v) = V_t N(-d_1) + D^* N(d_2) \tag{4.27}$$

Equivalently, at maturity bondholders either get back the face value of debt or if the company defaults, they get the residual asset value, such that:

$$B_T = \min[D, V_T] = D - \min[0, D - V_T]$$

As a result, the value of liabilities corresponds to a portfolio of a short put with strike D and a long risk-free bond with the same face value. Valuing liabilities before maturity can again be done through the BS formulas by evaluating: $B_t = De^{-r(T-t)} - P(V_t)$ where $P(.)$ is the corresponding value of a European put option with a strike D written on the asset's underlying.

Using the put-call parity, with $E(.)$ and $P(.)$ the values for a call and a put written on the company assets as an underlying, we have $E(V_t) - P(V_t) = V_t - D\exp\{-r(T-t)\}$ which implies

$$B_t = V_t - E(V_t)$$

At the same time, denoting y_t as the yield on the corporate bond, we have

$$B_t = D^* e^{-(y_t - r)\Delta t}$$

which implies a corporate bond spread:

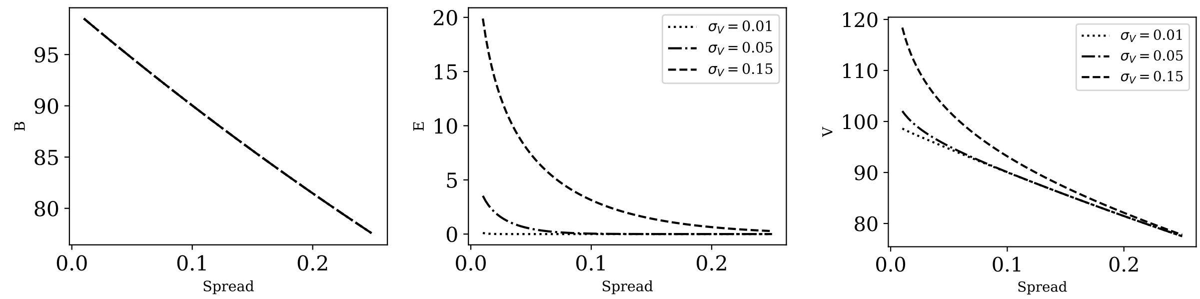
$$s(t, V_t, \sigma_v) = y_t - r = -\frac{1}{\Delta t} \ln \frac{B_t}{D^*} \tag{4.28}$$

where B_t is given by (4.27). Note then that:

$$\frac{\partial s_t}{\partial V_t} = -\frac{1}{\Delta t} \frac{N(-d_1)}{B_t} \tag{4.29}$$

where we make use of the fact that $\frac{\partial B_t}{\partial V_t} = \frac{\partial (V_t - E_t)}{\partial V_t} = 1 - N(d_1) = N(-d_1)$.

Figure 4.11: Merton Model



(a) (b) (c)

This figure shows the results of using the spread level to imply through the Merton model (a) the firm's liabilities (b) the firm's Equity value and (c) the asset value of the company. The company's debt is fixed at 100.

4.B Latent Factor Model Estimation

We apply the following algorithm based on Andersen and Basu (2003) to estimate the latent factor model from time-series data of the institutions' CDS prices.

Assume that Σ is an $n \times n$ matrix containing the target asset correlations between the key institutions. Assume the following factor model

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{M} + \boldsymbol{Z}$$

where $\boldsymbol{U}$ is an $n \times 1$ vector of standardized asset returns for the n institutions, $\boldsymbol{A}$ is an $n \times f$ common factor loadings matrix, $\boldsymbol{M}$ is an $f \times 1$ vector of common factors and $\boldsymbol{Z}$ is a $n \times 1$ vector of idiosyncratic factors. All factors are independent of each other with zero expectation and unit variance.

The problem is one of solving for $\boldsymbol{A}$ by minimizing the least squared difference of the model correlation matrix to the target one, such that:

$$\min_{\boldsymbol{A}} \left\{ \left(\boldsymbol{\Sigma} - \boldsymbol{A}\boldsymbol{A}' - \boldsymbol{F}\right) \left(\Sigma - \boldsymbol{A}\boldsymbol{A}' - \boldsymbol{F}\right)' \right\}$$

where $\boldsymbol{F}$ is a diagonal matrix such that $diag(\boldsymbol{F}) = 1 - diag(\boldsymbol{A}\boldsymbol{A}')$.

The numerical solution algorithm then is

1. Guess $\boldsymbol{F}^0$

2. Perform PCA on $\boldsymbol{\Sigma}-\boldsymbol{F}^i$ and compute $\boldsymbol{A}^i = \boldsymbol{E}^i\sqrt{\boldsymbol{\Lambda}}^i$, where i is an iterations counter, $\boldsymbol{E}$ is a matrix of the normalized column eigenvectors of $\boldsymbol{\Sigma}-\boldsymbol{F}$, $\sqrt{\boldsymbol{\Lambda}}$ is Cholesky decomposition of the diagonal matrix containing the f largest eigenvalues of $\boldsymbol{\Sigma}-\boldsymbol{F}$.

3. Calculate $\boldsymbol{F}^{i+1}$

4. Continue with Step 2, until $\boldsymbol{F}^{i+1}$ is sufficiently close to $\boldsymbol{F}^i$.

4.C Charts and Graphs

Figure 4.12: CDS Prices (bps)

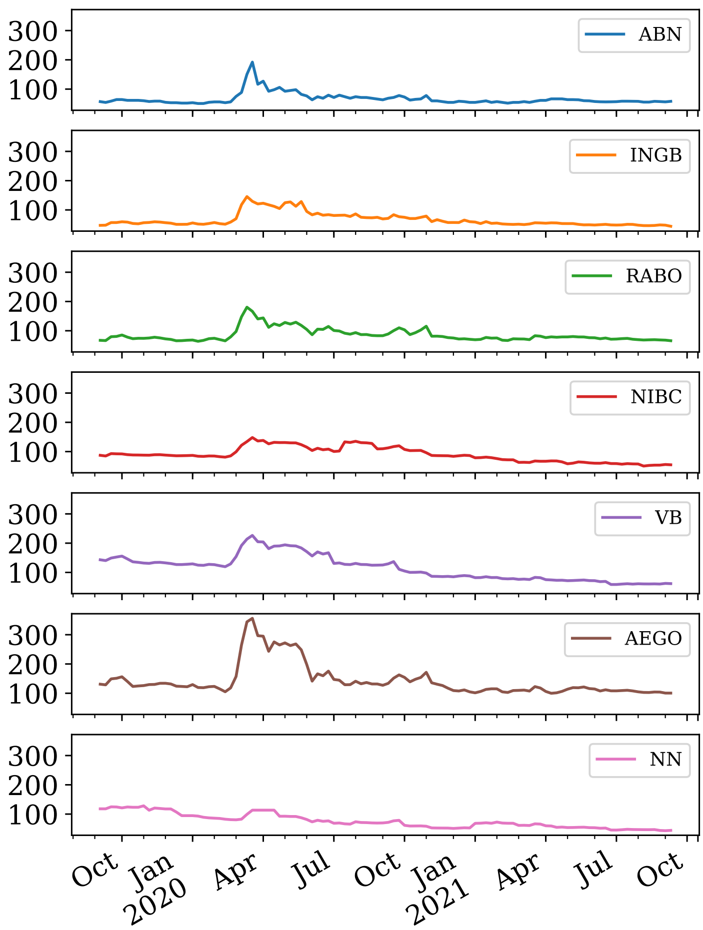


Figure 4.13: CDS Prices Log Changes

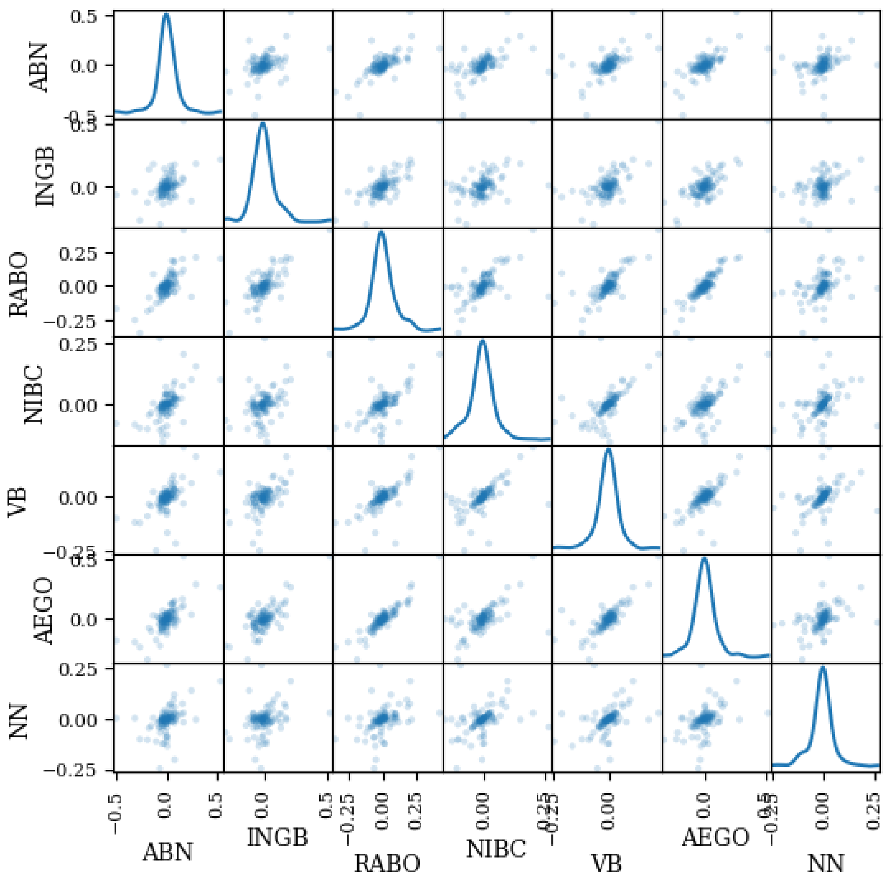


Figure 4.14: VSTOXX

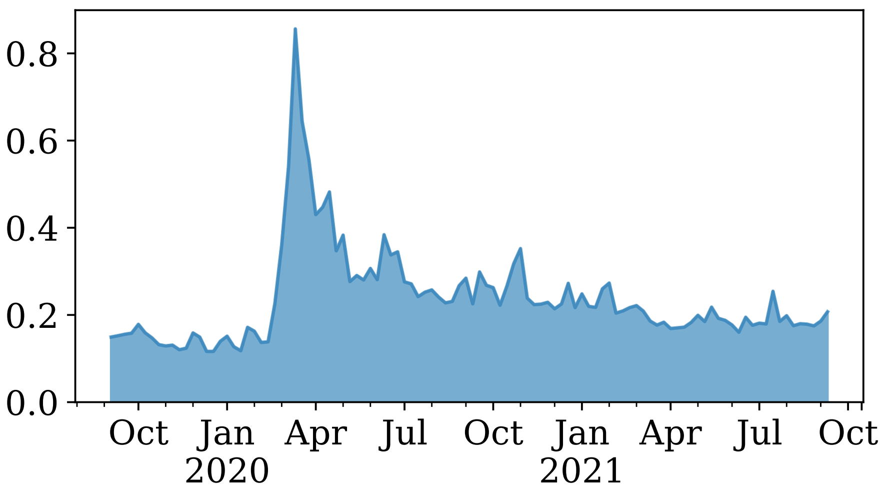


Figure 4.15: Liability Weights

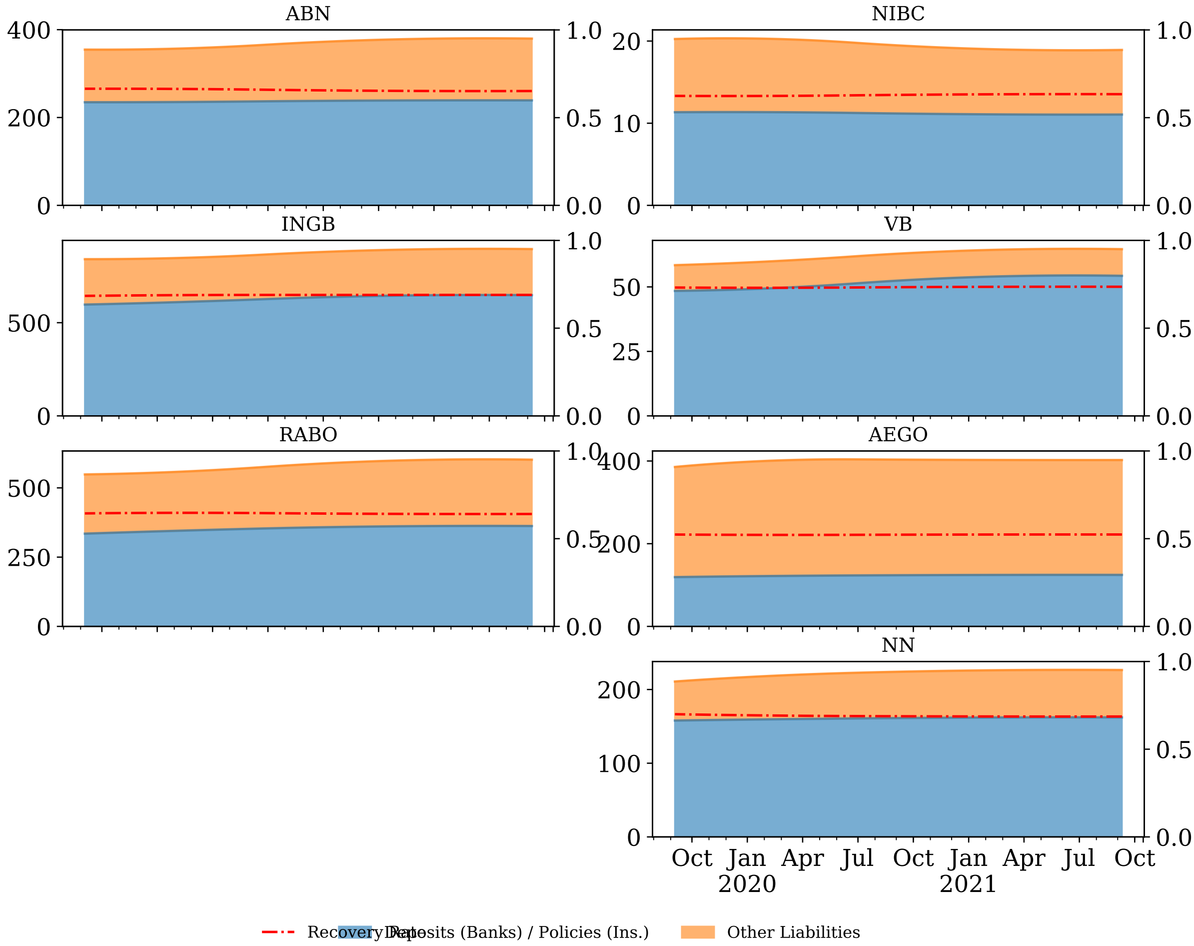


Figure 4.16: Simulations, Scatter Matrix

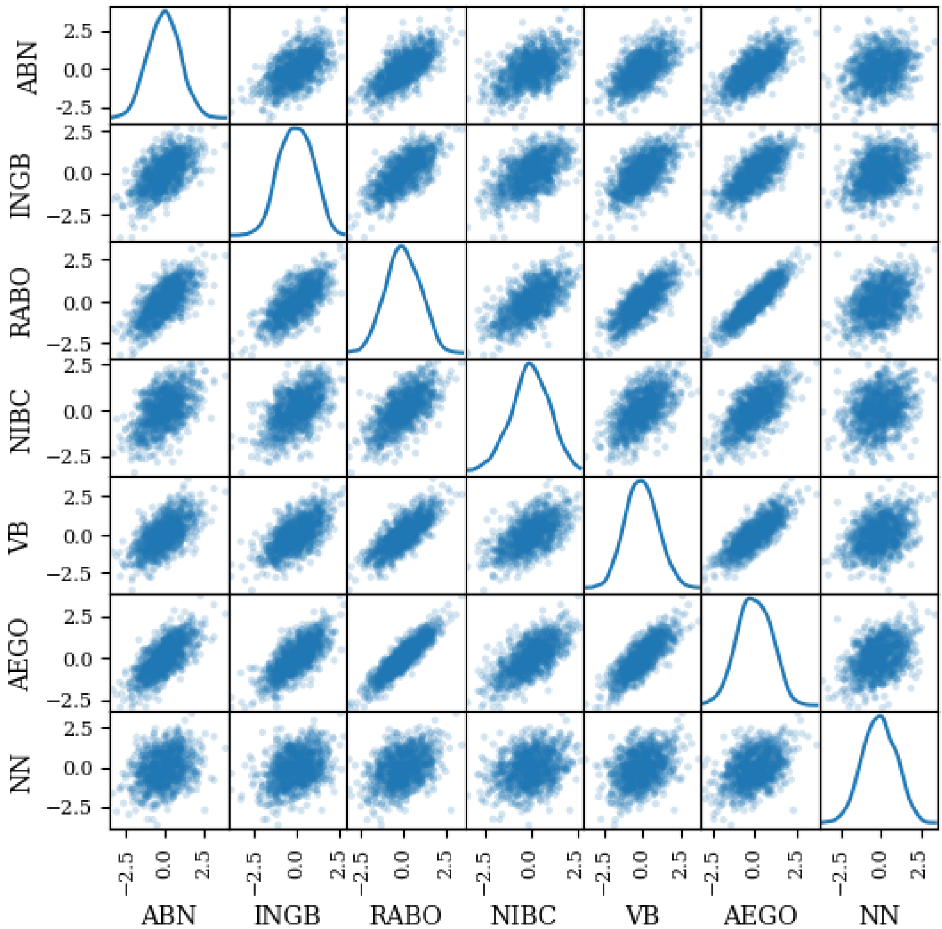
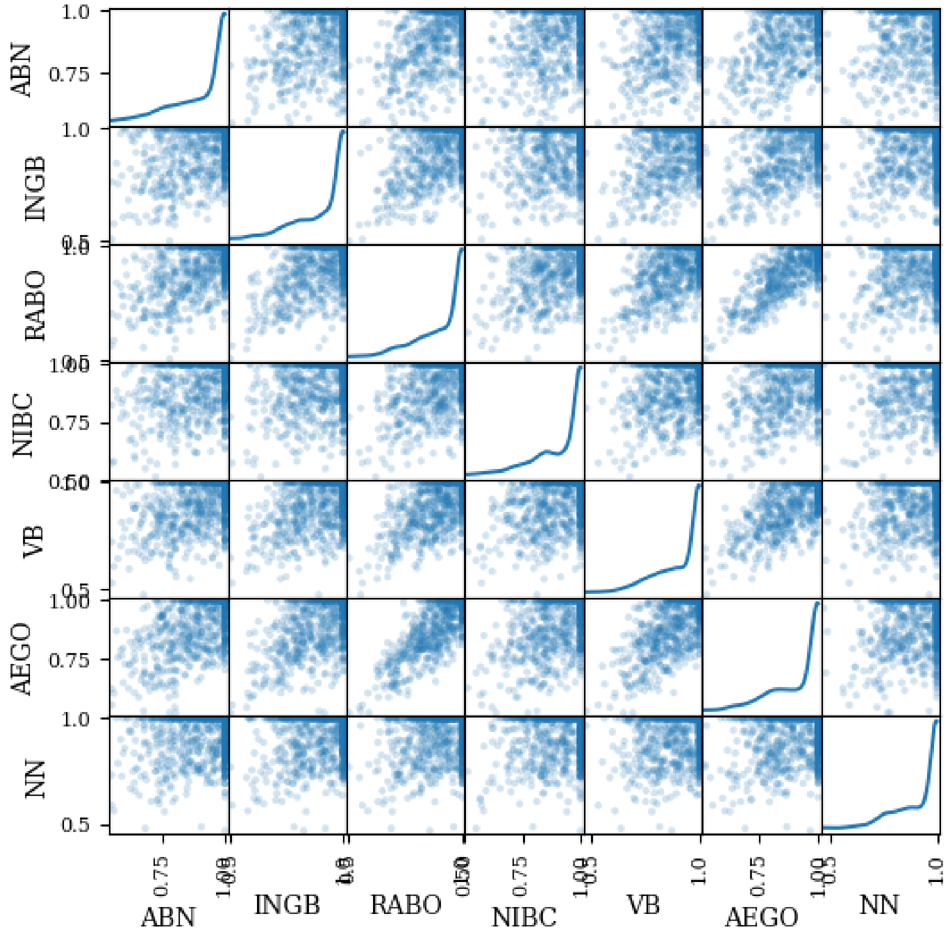


Figure 4.17: Recovery Rate Simulations



Bibliography

Acharya, V., Drechsler, I., and Schnabl, P. 2014. A pyrrhic victory? Bank bailouts and sovereign credit risk. *Journal of Finance*, 69(6):2689–2739.

Acharya, V. V. and Johnson, T. 2005. Insider trading in credit derivatives. CEPR Discussion Papers 5180, C.E.P.R. Discussion Papers.

Acharya, V. V. and Pedersen, L. H. 2005. Asset pricing with liquidity risk. *Journal of Financial Economics*, 77(2):375–410.

Acharya, V. V., Bharath, S. T., and Srinivasan, A. 2007. Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of financial economics*, 85 (3):787–821.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. 2017. Measuring systemic risk. *Review of Financial Studies*, 30(1):2–47.

Adrian, T. and Brunnermeier, M. K. 2016. CoVaR. *American Economic Review*, 106(7): 1705–1741.

Albuquerque, R. A., Cassel, J., Phalippou, L., and Schroth, E. J. 2018. Liquidity Provision in the Secondary Market for Private Equity Fund Stakes. Technical report.

Aldasoro, I. and Ehlers, T. 2018. The credit default swap market: What a difference a decade makes. *BIS Quarterly Review*.

Altman, E., Resti, A., and Sironi, A. 2004. Default recovery rates in credit risk modelling: A review of the literature and empirical evidence. *Economic Notes*, 33(2):183–208.

Amihud, Y. and Mendelson, H. 1986. Liquidity and stock returns. *Financial Analysts Journal, vol. 42, no. 3, 1986, pp. 43–48.*

Amihud, Y. and Mendelson, H. 1991. Liquidity, stock returns and financial policy. *Financial Analysts Journal*, 47(6):56–66.

Amihud, Y. and Mendelson, H. 2015. The Pricing of Illiquidity as a Characteristic and as Risk. *Multinational Finance Journal*, 19(3):149–168.

Andersen, L. and Sidenius, J. 2005. Extensions to the gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk*, 1.

Andersen, S.-J., Leif and Basu, S. 2003. All your hedges in one basket. *Risk*, 16:67–72.

Andonov, A., Eichholtz, P., and Kok, N. 2015. Intermediated investment management in private markets: Evidence from pension fund investments in real estate. *Journal of Financial Markets*, 22:73–103.

Andonov, A., Kräussl, R., and Rauh, J. 2021. Institutional investors and infrastructure investing. *The Review of Financial Studies*, 34(8):3880–3934.

Andrieş, A. M., Ongena, S., Sprincean, N., and Tunaru, R. 2022. Risk spillovers and interconnectedness between systemically important institutions. *Journal of Financial Stability*, 58:100963. ISSN 1572-3089.

Ang, A. 2011. Illiquid assets. Technical Report 28, CFA Institute Conference Proceedings Quarterly.

Ang, A. 2014, *Asset Management*. Oxford University Press.

Ang, A., Papanikolaou, D., and Westerfield, M. M. 2014. Portfolio choice with illiquid assets. *Management Science*, 60(11):2737–2761.

Annaert, J., De Ceuster, M., Van Roy, P., and Vespro, C. 2013. What determines euro area bank CDS spreads? *Journal of International Money and Finance*, 32(C):444–461.

Arora, N., Gandhi, P., and Longstaff, F. A. 2012. Counterparty credit risk and the credit default swap market. *Journal of Financial Economics*, 103(2):280–293. ISSN 0304-405X.

Artzner, P. 1999. Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal*, 3(2):11–25.

Augustin, P. and Schnitzler, J. 2021. Disentangling types of liquidity and testing limits-to-arbitrage theories in the CDS–bond basis. *European Financial Management*, 27(1): 120–146.

Aven, T. 2013. On the meaning and use of the risk appetite concept. *Risk Analysis*, 33 (3):462–468.

Avino, D. E., Conlon, T., and Cotter, J. 2019. Credit default swaps as indicators of bank financial distress. *Journal of International Money and Finance*, 94(C):132–139.

Back, K. E. 2010, *Asset Pricing and Portfolio Choice Theory.* Oxford University Press.

Bai, J. and Collin-Dufresne, P. 2019. The CDS-bond basis. *Financial Management*, 48 (2):417–439.

Ball, L. and Mankiw, N. G. 2007. Intergenerational Risk Sharing in the Spirit of Arrow, Debreu, and Rawls, with Applications to Social Security Design. *Journal of Political Economy*, 115(4):523–547.

Baur, D. G. and Schulze, N. 2009. Financial market stability–a test. *Journal of International Financial Markets, Institutions and Money*, 19(3):506–519.

Beetsma, R. and Romp, W. *Intergenerational Risk Sharing*, chapter 6, pages 311–380. Handbook of the Economics of Population Aging. Elsevier, 2016.

Bharath, S. T. and Shumway, T. 2008. Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21(3):1339–1369.

Bianchi, M. L. and Sorrentino, A. M. 2021. Exploring the systemic risk of domestic banks with δcovar and elastic-net. *Journal of Financial Services Research*, pages 1–15.

Bichuch, M. and Guasoni, P. 2018. Investing with liquid and illiquid assets. *Mathematical Finance*, 28(1):119–152.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559. ISSN 0304-405X.

Boelaars, I. and Mehlkopf, R. 2018. Optimal risk-sharing in pension funds when stock and labor markets are co-integrated. DNB Working Papers 595, Netherlands Central Bank, Research Department.

Bollen, N. P. B. and Sensoy, B. A. 2015. How Much for a Haircut? Illiquidity, Secondary Markets, and the Value of Private Equity. Working Paper Series 2015-08, Ohio State University, Charles A. Dice Center for Research in Financial Economics.

Bongaerts, D., De Jong, F., and Driessen, J. 2012. An asset pricing approach to liquidity effects in corporate bond marketsets. Technical report, Tilburg University.

Bovenberg, L. and Mehlkopf, R. 2014. Optimal design of funded pension schemes. *Annual Review of Economics*, 6:445–474.

Boyle, P. P. and Lin, X. 1997. Optimal portfolio selection with transaction costs. *North American Actuarial Journal*, 1(2):27–39.

Brandimarte, P. 2006, *Numerical Methods in Finance and Economics, Second Edition.* John Wiley and Sons Inc.

Brauneis, A., Mestel, R., Riordan, R., and Theissen, E. 2021. How to measure the liquidity of cryptocurrency markets? *Journal of Banking & Finance*, 124:106041. ISSN 0378-4266.

Brogi, M., Lagasio, V., and Riccetti, L. 2021. Systemic risk measurement: bucketing global systemically important banks. *Annals of Finance*, 17(3):319–351.

Brownlees, C. and Engle, R. F. 2017. SRISK: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies*, 30(1):48–79.

Brunnermeier, M. K. and Pedersen, L. H. 2009. Market liquidity and funding liquidity. *The Review of Financial Studies*, 22(6):2201–2238.

Bräuning, F. and Koopman, S. J. 2016. The dynamic factor network model with an application to global credit risk. Working Papers 16-13, Federal Reserve Bank of Boston.

Buccola, S. T. 1982. Portfolio Selection Under Exponential And Quadratic Utility. *Western Journal of Agricultural Economics*, 7(1):1–10.

Buss, A., Uppal, R., and Vilkov, G. 2015. Asset prices in general equilibrium with recursive utility and illiquidity induced by transactions costs. Technical report.

Cai, Y. and Judd, K. L. 2014. Advances in numerical dynamic programming and new applications. *Handbook of computational economics.*

Cai, Y., Judd, K. L., and Xu, R. 2013. Numerical Solution of Dynamic Portfolio Optimization with Transaction Costs. *NBER Working Paper No. w18709.*

Campbell, J. Y. and Viceira, L. M. 2002, *Strategic Asset Allocation, Chapter 5.* Oxford University Press, New York.

Campbell, J. Y., Chacko, G., Rodriguez, J., and Viceira, L. M. 2004. Strategic asset allocation in a continuous-time var model. *Journal of Economic Dynamics and Control*, 28(11):2195–2214.

Campbell, J. and Nosbusch, Y. 2007. Intergenerational risksharing and equilibrium asset prices. *Journal of Monetary Economics*, 54(8):2251–2268.

Carr, P. and Wu, L. 2011. A simple robust link between american puts and credit protection. *Review of Financial Studies*, 24(2):473–505.

Cerný, A. 2009, *Mathematical Techniques in Finance: Tools for Incomplete Markets.* Princeton University Press, rev - revised, 2 edition.

Chan-Lau, J. A. and Gravelle, T. 2005. The end: A new indicator of financial and nonfinancial corporate sector vulnerability. IMF Working Papers 2005/231, International Monetary Fund.

Chan-Lau, J. A. and Sy, A. N. R. 2007. Distance-to-default in banking: A bridge too far? *Journal of Banking Regulation*, 9. ISSN 00222879, 15384616.

Chatziantoniou, I., Filis, G., and Floros, C. 2017. Asset prices regime-switching and the role of inflation targeting monetary policy. *Global Finance Journal*, 32:97–112.

Chen, D. H., Beetsma, R. M., Ponds, E. H., and Romp, W. E. 2016. Intergenerational risk-sharing through funded pensions and public debt. *Journal of Pension Economics & Finance*, 15(2):127–159.

Christensen, J. 2006. Joint estimation of default and recovery risk: A simulation study. Technical report.

Ciurila, N. and Romp, W. E. 2015. The political arrangement of pay-as-you-go pension systems in the presence of financial and demographic shocks.

Cochrane, J. H. 2022. Portfolios for long-term investors. *Review of Finance*, 26(1):1–42.

Constantinides, G. M. 1986. Capital market equilibrium with transaction costs. *Journal of Political Economy*, 94(4):842–62.

Cox, J. C. and Huang, C.-F. 1989. Optimum consumption and portfolio policies when asset prices follow a di§usion process.

Crawford, R. and O'Dea, C. 2020. Household portfolios and financial preparedness for retirement. *Quantitative Economics*, 11(2):637–670.

Crosbie, P. and Bohn, J. 2002. Modeling default risk. Technical report, KVN.

Cui, J., De Jong, F., and Ponds, E. 2011. Intergenerational risk sharing within funded pension schemes. *Journal of Pension Economics and Finance*, 10(1):1–29.

Culp, C. L., van der Merwe, A., and Stärkle, B. J. Single-name credit default swaps. In *Credit Default Swaps*, Palgrave Studies in Risk and Insurance, chapter 0, pages 219–248. Palgrave Macmillan, 2018.

Dai, M., Jin, H., and Liu, H. 2011. Illiquidity, position limits, and optimal investment for mutual funds. *Journal of Economic Theory*, 146(4):1598–1630. ISSN 0022-0531.

D'Amato, M. and Galasso, V. 2010. Political intergenerational risk sharing. *Journal of Public Economics*, 94(9-10):628–637.

de Haan, J., Jin, Z., and Zhou, C. 2019. Micro-prudential regulation and banks' systemic risk. DNB Working Papers 656, Netherlands Central Bank, Research Department.

Diamond, D. W. and Rajan, R. G. 2011. Fear of fire sales, illiquidity seeking, and credit freezes. *The Quarterly Journal of Economics*, 126(2):557–591.

Diamond, P. A. 1982. Aggregate Demand Management in Search Equilibrium. *Journal of Political Economy*, 90(5):881–94.

DNB. Perspective on the structure of the Dutch banking sector. In *Perspective on the structure of the Dutch banking sector: efficiency and stability through competition and diversity*, volume 41 of *IFC Bulletins chapters*. DNB, 2015.

DNB. 2019. Size of the banking sector: Implications for financial stability. *Tech Report*.

Duan, J. 1994. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4(2):155–167.

Duan, J. 2000. Correction: Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 10(4):461–462.

Duffie, D. 2001, *Dynamic Asset Pricing Theory, 3d ed.* Princeton University Press, Princeton, N.J.

Duffie, D. 1999. Credit swap valuation. *Financial Analysts Journal*, 55.

Duffie, D. and Singleton, K. J. 1999. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12(4):687–720.

Duffie, D., Garleanu, N., and Pedersen, L. H. 2004. Over-the-Counter Markets. NBER Working Papers 10816, National Bureau of Economic Research, Inc.

Duffie, D., Garleanu, N., and Pedersen, L. H. 2005. Over-the-Counter Markets. *Econometrica*, 73(6):1815–1847.

EBA. 2020. EBA report on the appropriate methodology to calibrate O-SII buffer rates. Policy report, European Banking Authority.

Engle, R. 2018. Systemic risk 10 years later. *Annual Review of Financial Economics*, 10 (1):125–152.

Franzoni, F., Nowak, E., and Phalippou, L. 2012. Private equity performance and liquidity risk. *Journal of Finance*, 67(6):2341–2373.

Frye, J. 2000. Collateral damage: A source of systematic credit risk. *Risk*, 13:91–94.

Gai, P. and Vause, N. 2006. Measuring investors' risk appetite. *International Journal of Central Banking*, 2(1).

Gennotte, G. and Jung, A. 1994. Investment Strategies under Transaction Costs: The Finite Horizon Case. *Management Science*, 40(3):385–404.

Glasserman, P. and Young, H. P. 2016. Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831.

Gollier, C. 2008. Intergenerational risk-sharing and risk-taking of a pension fund. *Journal of Public Economics*, 92(5-6):1463–1485.

Gordon, R. H. and Varian, H. R. 1988. Intergenerational risk sharing. *Journal of Public Economics*, 37(2):185–202.

Gottardi, P. and Kubler, F. 2011. Social security and risk sharing. *Journal of Economic Theory*, 146(3):1078–1106.

Goyenkoa, H. C. W., Ruslan Y. and Trzcinkab, C. A. 2009. Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92(2):153 – 181. ISSN 0304-405X.

Gropp, R., Vesala, J., and Vulpes, G. 2006. Equity and bond market signals as leading indicators of bank fragility. *Journal of Money, Credit and Banking*, 38(2):399–428. ISSN 00222879, 15384616.

Gârleanu, N. 2009. Portfolio choice and pricing in illiquid markets. *Journal of Economic Theory*, 144(2):532–564.

Hattori, M., Schrimpf, A., and Sushko, V. 2016. The response of tail risk perceptions to unconventional monetary policy. *American Economic Journal: Macroeconomics*, 8(2): 111–36.

Hemert, O. v. 2005. Optimal intergenerational risk sharing. LSE Research Online Documents on Economics 24660, London School of Economics and Political Science, LSE Library.

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

Huang, X., Zhou, H., and Zhu, H. 2012. Systemic risk contributions. *Journal of Financial Services Research*, 42(1):55–83.

Hull, J. and White, A. 2004. Valuation of a cdo and an nth to default CDS without monte carlo simulation. *Journal of Derivatives - J DERIV*, 12:8–23.

Hull, J. C. 2018, *Risk Management and Financial Institutions*. John Wiley & Sons. ISBN 978-1-119-44811-2.

IJtsma, P. and Spierdijk, L. 2017. Systemic risk with endogenous loss given default. *Journal of Empirical Finance*, 44:145–157.

Illing, M. and Aaron, M. 2005. A brief survey of risk-appetite indexes. *Tech Report, Bank of Canada*.

Jackson, M. O. and Pernoud, A. 2021. Systemic risk in financial networks: A survey. *Annual Review of Economics*, 13(1):171–202.

Jansen, K. A. and Werker, B. J. M. 2020. The Liquidity Premium Across Asset Classes. Working paper.

Jayaram, K. and Gadanecz, B. Macroprudential policy frameworks, instruments and indicators: a review. In for International Settlements, B., editor, *Combining micro and macro data for financial stability analysis*, volume 41 of *IFC Bulletins chapters*. BIS, 2016.

Jessen, C. and Lando, D. 2015. Robustness of distance-to-default. *Journal of Banking & Finance*, 50:493–505. ISSN 0378-4266.

JP Morgan. 2020. JP Morgan Asset Management: Long-Term Capital Market Assumptions 2020. *Research Report.*

JP Morgan. 2022. JP Morgan Asset Management: Long-Term Capital Market Assumptions 2018. *Research Report.*

Judd, K. L. 1998, *Numerical Methods in Economics.* MIT Press.

Kamga, K. and Wilde, C. Liquidity premium in CDS markets. Working paper.

Kapadia, N. and Pu, X. 2012. Limited arbitrage between equity and credit markets. *Journal of Financial Economics*, 105(3):542–564. ISSN 0304-405X.

Kaplan, G. and Violante, G. L. 2014. A Model of the Consumption Response to Fiscal Stimulus Payments. *Econometrica*, 82(4):1199–1239.

Karatzas, I. and Shreve, S. E. 1998, *Methods of Mathematical Finance.* Springer-Verlag.

Kaserer, C. and Klein, C. 2019. Systemic risk in financial markets: How systemically important are insurers? *Journal of Risk & Insurance*, 86(3):729–759.

Kenc, T., Cevik, E. I., and Dibooglu, S. 2021. Bank default indicators with volatility clustering. *Annals of Finance*, 17(1):127–151.

Kleymenova, A., Talmor, E., and Vasvari, F. P. 2012. Liquidity in the secondaries private equity market. *working paper, London Business School.*

Korteweg, A. G. and Westerfield, M. M. 2022. Asset allocation with private equity. *Available at SSRN 4017858.*

Krueger, D. and Kubler, F. 2006. Pareto-Improving Social Security Reform when Financial Markets are Incomplete!? *American Economic Review*, 96(3):737–755.

Lancia, F., Russo, F., and Worrall, T. S. 2020. Optimal sustainable intergenerational insurance. CERP discussion paper.

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

Leland, H. E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance*, 49(4):1213–1252.

Levy, H. and Markowtiz, H. M. 1979. Approximating Expected Utility by a Function of Mean and Variance. *American Economic Review*, 69(3):308–317.

Liu, J., Longstaff, F. A., and Pan, J. 2003. Dynamic asset allocation with event risk. *Journal of Finance*, 58(1):231–259.

Longstaff, F. A. 2001. Optimal Portfolio Choice and the Valuation of Illiquid Securities. *Review of Financial Studies*, 14(2):407–31.

Loon, Y. C. and Zhong, Z. K. 2014. The impact of central clearing on counterparty risk, liquidity, and trading: Evidence from the credit default swap market. *Journal of Financial Economics*, 112(1):91–115.

Loon, Y. C. and Zhong, Z. K. 2016. Does Dodd-Frank affect OTC transaction costs and liquidity? Evidence from real-time CDS trade reports. *Journal of Financial Economics*, 119(3):645–672.

Lou, X. and Sadka, R. 2005. Liquidity and asset prices. *Foundations and Trends in Finance, Vol. 1, No 4 (2005) 269–364*.

Lusardi, A. and Mitchell, O. 2007. Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics*, 54 (1):205–224.

Magill, M. J. P. and Constantinides, G. M. 1976. Portfolio selection with transactions costs. *Journal of Economic Theory*, 13(2):245–263.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance*, 7(1):77–91. ISSN 00221082, 15406261.

Merton, R. C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):867–887.

Merton, R. C. 1971. Optimal consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413.

Merton, R. C. 1973. An intertemporal capital asset pricing model. *Econometrica*, 41: 867–887.

Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2):449–470.

Merton, R. C. 1981. On the Role of Social Security as a Means for Efficient Risk-Bearing in an Economy Where Human Capital Is Not Tradeable. NBER Working Papers 0743, National Bureau of Economic Research, Inc.

Miklós, K. and Ádám, S. 2002. Portfolio Choice with Illiquid Assets. Rajk László Szakkollégium Working Papers 6, Rajk László College.

Miranda, M. J. and Fackler, P. L. 2002, *Applied Computational Economics and Finance*. The MIT Press.

Moore, K. S. and Young, V. R. 2006. Optimal insurance in a continuous-time model. *Insurance: Mathematics and Economics*, 39(1):47 – 68. ISSN 0167-6687.

Moore, K. and Zhou, C. 2012. Identifying systemically important financial institutions: size and other determinants. DNB Working Papers 347, Netherlands Central Bank, Research Department.

Moratis, G. and Sakellaris, P. 2021. Measuring the systemic importance of banks. *Journal of Financial Stability*, 54:100878. ISSN 1572-3089.

Munk, C. 2013, *Financial Asset Pricing Theory*. Oxford University Press.

Munk, C. 2020. A mean-variance benchmark for household portfolios over the life cycle. *Journal of Banking & Finance*, 116(C).

Nadauld, T. D., Sensoy, B. A., Vorkink, K., and Weisbach, M. S. 2019. The liquidity cost of private equity investments: Evidence from secondary market transactions. *Journal of Financial Economics*, 132(3):158–181.

Nagel, S. and Purnanandam, A. 2019. Banks' Risk Dynamics and Distance to Default. *The Review of Financial Studies*, 33(6):2421–2467. ISSN 0893-9454.

Nakajima, M. and Telyukova, I. 2020. Home equity in retirement. *International Economic Review*, 61(2):573–616.

Nolde, N. and Zhou, C. 2021. Extreme value analysis for financial risk management. *Annual Review of Statistics and Its Application*, 8(1):217–240.

OECD. 2019. Annual Survey of Large Pension Funds and Public Pension Reserve Funds. Technical report.

Oh, D. H. and Patton, A. J. 2018. Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics*, 36(2):181–195.

Olijslagers, S., Petersen, A., de Vette, N., and van Wijnbergen, S. 2019. What option prices tell us about the ECB's unconventional monetary policies.

Paddrik, M. and Tompaidis, S. 2019. Market-making costs and liquidity: Evidence from CDS markets. Working Papers 19-01, Office of Financial Research, US Department of the Treasury.

Pan, J. and Singleton, K. J. 2008. Default and recovery implicit in the term structure of sovereign CDS spreads. *Journal of Finance*, 63(5):2345–2384.

Pastor, L. and Stambaugh, R. F. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, 111(3):642–685.

PensionsEurope. 2018. Pension Funds Statistics and Trends. Technical report.

Puzanova, N. and Düllmann, K. 2013. Systemic risk contributions: A credit portfolio approach. *Journal of Banking & Finance*, 37(4):1243–1257.

Qiu, J. and Yu, F. 2012. Endogenous liquidity in credit derivatives. *Journal of Financial Economics*, 103(3):611–631. ISSN 0304-405X.

Rachev, S. T., Stoyanov, S., and Fabozzi, F. J. 2008, *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization.* John Wiley & Sons, Inc. ISBN 978-0-470-05316-4.

Ramadorai, T. 2012. The secondary market for hedge funds and the closed hedge fund premium. *Journal of Finance*, 67(2):479–512.

Ronn, E. I. and Verma, A. K. 1986. Pricing risk-adjusted deposit insurance: An option-based model. *Journal of Finance*, 41(4):871–895. ISSN 00221082, 15406261.

Ross, S. M. 2007, *Introduction to Probability Models, Ninth Edition.* Elsevier Inc.

Ross, S. A. 1999. Adding Risks: Samuelson's Fallacy of Large Numbers Revisited. *Journal of Financial and Quantitative Analysis*, 34(3):323–339.

Rust, J. 1996. Numerical dynamic programming in economics. *Handbook of computational economics*, 1:619–729.

Samuelson, P. 1963. Risk and uncertainty: A fallacy of large numbers. *Scientia*, 57(98): 108.

Schumacher, J. M. 2021. A note on Gollier's model for a collective pension scheme. *Journal of Pension Economics & Finance*, 20(2):187–211.

Segoviano, M. A. and Goodhart, C. 2009. Banking stability measures. *IMF Working paper WP/09/4.*

Shao, A., Chen, H., and Sherris, M. 2019. To borrow or insure? Long term care costs and the impact of housing. *Insurance: Mathematics and Economics*, 85:15–34.

Sharpe, W. F. 2007. Expected Utility Asset Allocation. *Financial Analysts Journal*, 63 (5).

Shiller, R. J. 1999. Social security and institutions for intergenerational, intragenerational, and international risk-sharing. *Carnegie-Rochester Conference Series on Public Policy*, 50(1):165–204.

Shleifer, A. and Vishny, R. W. 1992. Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance*, 47(4):1343–1366.

Shreve, S. E. 2004, *Stochastic Calculus for Finance II, Continuous-Time Models*. New York: Springer-Verlag.

Sousa, R. 2010. Housing wealth, financial wealth, money demand and policy rule: Evidence from the euro area. *North American Journal of Economics and Finance*, 21(1): 88–105.

Sundaresan, S. 2013. A review of Merton's model of the firm's capital structure with its wide applications. *Annual Review of Financial Economics*, 5(1):21–41.

Tarashev, N. A. and Zhu, H. 2006. The pricing of portfolio credit risk. BIS Working Papers 214, Bank for International Settlements.

Vasicek, O. 1987. Probability of loss on loan portfolio. Technical report.

Vayanos, D. and Wang, J. 2012. Market Liquidity - Theory and Empirical Evidence. FMG Discussion Papers dp709, Financial Markets Group.

Wang, W. 2021. Modelling systemic risk using neural network quantile regression. *Empirical Economics*. ISSN 0377-7332.

Wu, L. 2003. Jumps and dynamic asset allocation. *Review of Quantitative Finance and Accounting*, 20(3):207–243.

Yakov Amihud, H. M. and Pedersen, L. H. 2005. Liquidity and asset prices. *Foundations and Trends in Finance, Vol. 1, No 4 (2005) 269–364*.

Zabel, E. 1973. Consumer Choice, Portfolio Decisions, and Transaction Costs. *Econometrica*, 41(2):321–35.

Zhang, B. Y., Zhou, H., and Zhu, H. 2009. Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *Review of Financial Studies*, 22(12):5099–5131.

Zhou, C. 2010. Are banks too big to fail? Measuring systemic importance of financial institutions. *International Journal of Central Banking*, 6(34):205–250.

Summary in English

In this thesis, we consider three non-trivial problems of risk allocation and apply approaches from theoretical finance and risk management to address several policy debates from a macro-finance point of view. The three topics considered are diverse, but there is a common theme that runs through each chapter. In each, we focus on the resolution of barriers to a first-best risk allocation rule. The barriers can be in the form of market illiquidity, current generations not being able to participate in the shocks affecting future generations, or risk spillovers between systemic institutions.

The first essay in Chapter 2 addresses a classical finance problem of allocating risks efficiently in an investment portfolio. In our set-up,illiquidity in one of the assets exists as uncertainty in the immediate availability of a market where price risk can be traded. As a result, illiquidity acts as an additional non-hedgeable risk component, affecting the portfolio choice decision. We show how illiquidity endogenizes the risk aversion of investors, making the portfolio choice and consumption decisions functions of the share of illiquid wealth.

The second essay in Chapter 3 puts the allocating problem into a policy-relevant setting by considering how illiquidity in the form of uncertain trading costs affects the ability of different generations to share financial risk with each other. We use a stylized two-period overlapping generations framework, where each generation makes a portfolio allocation decision for retirement. In this context, designing optimal social security institutions can also be seen as a risk management problem. A policymaker balances, on one hand, the benefits of a wider risk-bearing pool by integrating the young into the financial shock affecting the elderly, and on the other, the costs of potentially destabilizing the young's retirement savings.

The third essay in Chapter 4 shows how monitoring and evaluating systemic risk can be done through a risk management lens. The supervisor implicitly owns a portfolio of defaultable loans corresponding to the liabilities of financial institutions, and needs to manage the tail risk of the portfolio. We, thus, propose a credit portfolio approach for evaluating systemic risk and attributing it across institutions by constructing a model that can be estimated from high-frequency CDS data.

Summary in Dutch

In dit proefschrift beschouwen we drie niet-triviale problemen van risicoallocatie en passen we benaderingen uit finance en risicobeheer toe om verschillende beleidsdebatten vanuit macrofinancieel oogpunt te analyseren. De drie behandelde onderwerpen zijn divers, maar er is een gemeenschappelijk thema dat alle hoofdstukken verbindt. In elk hoofdstuk richten we ons op het wegnemen van belemmeringen om dichter bij een first-best risico-allocatieregel te komen. Voorbeelden van belemmeringen zijn: illiquiditeit, het feit dat huidige generaties niet kunnen deelnemen aan de schokken die toekomstige generaties zullen treffen, of spillovers van risico tussen systeeminstellingen.

Het eerste essay in hoofdstuk (2) behandelt een klassiek financieel probleem van het efficiënt toewijzen van risico's in een beleggingsportefeuille. In onze opzet modelleren we illiquiditeit in een van de activa als onzekerheid in de beschikbaarheid van een markt waar prijsrisico kan worden verhandeld. Als gevolg hiervan fungeert illiquiditeit als een extra onverzekerbare risicocomponent, die van invloed is op de portefeuillekeuze. We laten zien hoe illiquiditeit de risicoaversie van beleggers endogeniseert, waardoor in wezen de portefeuillekeuze en consumptiebeslissingen afhankelijk worden van het aandeel illiquide activa.

Het tweede essay in hoofdstuk (3) plaatst het allocatieprobleem in een beleidsrelevante omgeving door na te gaan hoe illiquiditeit in de vorm van onzekere handelskosten van invloed is op het vermogen van verschillende generaties om financiële risico's met elkaar te delen. We gebruiken een gestileerd raamwerk voor overlappende generaties met twee perioden, waarbij elke generatie een beslissing neemt over de portefeuilletoewijzing voor pensionering. In deze context kan het ontwerpen van een optimaal socialezekerheidssysteem ook worden gezien als een risicobeheersingsprobleem. Een beleidsmaker balanceert enerzijds de voordelen van een grotere risicodragende pool door de jongeren te integreren

in de financiële schok die de huidige bejaarden treft, en anderzijds de kosten van het potentieel destabiliseren van het pensioensparen van de jongeren in de loop van de tijd, waarbij extra risico wordt toegevoegd aan het vermogen van de jongeren.

Het derde essay in hoofdstuk (4) laat zien hoe het monitoren en evalueren van systeemrisico's kan worden gedaan door het probleem te bekijken door een risicobeheerlens. De toezichthouder bezit impliciet een portefeuille van in gebreke blijvende leningen die overeenkomt met de verplichtingen van financiële instellingen, en moet het staartrisico van de portefeuille beheren. Daarom stellen we een kredietportefeuillebenadering voor om systeemrisico's te evalueren en toe te wijzen aan instellingen door een model te construeren dat kan worden geschat op basis van hoogfrequente CDS-data.