The London School of Economics and Political Science

Topics in Portfolio Optimisation and Systemic Risk

Mathieu Steve Dubois

A thesis submitted to the Department of Mathematics of the London School of Economics and Political Science for the degree of

Doctor of Philosophy

London, October 2015

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 126 pages.

Statement of conjoint work:

I confirm that first part of the thesis was first published as a co-authored article with my supervisor, Dr. Luitgard Veraart, in SIAM Journal on Financial Mathematics, Vol. 6, pp. 201-241, 2015; published by the Society for Industrial and Applied Mathematics (SIAM).

Abstract

This thesis is concerned with different sources of risk occurring in financial markets. We follow a bottom-up approach by carrying out an analysis from the perspective of a single investor to the whole banking system.

We first consider an investor who faces parameter uncertainty in a continuous-time financial market. We model the investor's preference by a power utility function leading to constant relative risk aversion. We show that the loss in expected utility is large when using a simple plug-in strategy for unknown parameters. We also provide theoretical results that show the trade-off between holding a well-diversified portfolio and a portfolio that is robust against estimation errors. To reduce the effect of estimation, we constrain the weights of the risky assets with a norm leading to a sparse portfolio. We provide analytical results that show how the sparsity of the constrained portfolio depends on the coefficient of relative risk aversion. Based on a simulation study, we demonstrate the existence and the uniqueness of an optimal bound on the norm for each level of relative risk aversion.

Next, we consider the interbank lending market as a network in which the nodes represent banks and the directed edges represent interbank liabilities. The interbank network is characterised by the matrix of liabilities whose entries are not directly observable, as bank balance sheets provide only total exposures to the interbank market. Using a Bayesian approach, we assume that each entry follows a Gamma distributed prior. We then construct a Gibbs sampler of the conditional joint distribution of interbank liabilities given total interbank liabilities and total interbank assets. We illustrate our methodology with a stress test on two networks of eleven and seventy-six banks. We identify under which circumstances the choice of the prior influences the stability and the structure of the network.

Acknowledgements

First and foremost I would like to thank Luitgard Veraart for the supervision of this thesis. Her guidance and constant support have been invaluable. I am also grateful to Jörn Sass and Mihail Zervos for readily accepting to act as my examiners.

I would like to thank the faculty members of the Department of Mathematics and the Systemic Risk Center. They provided a pleasant and stimulating atmosphere to carry out academic research. Special thanks go to the participants of the Financial Mathematics Reading Group, in particular Christoph Czychowsky and Jose Pasos, for their thought-provoking questions that helped to grow abstract and nascent ideas into tangible concepts. I am also grateful to Albina Danilova, Pavel Gapeev, Katja Neugebauer, and Andreas Uthemann for generously sharing their experience of the inner workings of the academic world.

I would also like to express my gratitude to my scholarship donors Liz and Peter Jones. Thanks to their warm hospitality, I have got the chance to discover English culture. Finally, I would like to thank my parents, Françoise and Michel, and my partner, Polina. Their support and optimism have been a driving force throughout my graduate studies.

Financial support by the Jones/Markham scholarship and the Economic and Social Research Council [grant number ES/K002309/1] is gratefully acknowledged.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One should always divide his wealth into three parts: a third in land, a third in merchandise, and a third ready to hand.

- Babylonian Talmud: Tractate Baba Mezi'a, 3rd to 5th century A.D.

As the quote above suggests, the problem of wealth allocation through diversification is a long standing concern. The first methodological framework for a mathematical formulation of diversification is developed in Markowitz (1952). In his seminal article, he assumes that investors care about mean returns and consider variance as "an undesirable thing", (Markowitz, 1952, p. 77). He then shows that the risk specific to each asset, also called idiosyncratic risk, can be eliminated through diversification by taking advantage of the imperfect correlation between assets. Finally, he identifies the portfolio with maximal mean return for a given level of variance. By allowing the variance to vary, the set of optimal portfolios (the efficient frontier) is obtained. Every portfolio below the efficient frontier is suboptimal and can be further diversified, while the region above the frontier is simply unattainable with the given universe of risky assets.

The mean-variance approach has had a profound impact in financial economics and research in modern portfolio theory has focused on understanding its limitations and improving on it. This framework implicitly assumes that the financial market is granular so that the interactions between actors are not relevant. The systemic risk literature takes a complementary approach by studying when and how the stability of the financial market is affected by the structure of exposures among banks. Therefore, from a

systemic perspective, the diversification of exposures is a determinant factor of stability.

This thesis investigates different sources of risk faced by individual investors and financial institutions. In the first part of the thesis, we study the performance of a standard dynamic model-based investment strategy for investors facing parameter uncertainty. To control the effect of estimation, we constrain the proportion of wealth invested in risky assets with a suitable norm leading to a sparse portfolio. We present novel analytical results that show how the sparsity of the constrained portfolio depends on the investor's risk aversion. Our work contributes to the portfolio theory literature by identifying an optimal degree of diversification with respect to the estimation procedure and the investor's preferences. We build a sparse dynamic strategy that is robust against estimation errors for each risk averse investor. We then show that the optimally constrained strategy performs well out-of-sample.

In the second part of the thesis, we model the interbank lending market as a network in which nodes represent banks and directed edges represent interbank liabilities. The network is then characterised by the matrix of liabilities whose entries are not directly observable, as a bank balance sheet provides only its total of interbank liabilities (sum of rows) and its total interbank assets (sum of columns).

Using a Bayesian approach, we assume that each entry follows a specific prior distribution. We then develop a novel numerical method to sample from the conditional joint distribution of interbank liabilities given total interbank liabilities and total interbank assets. Our methodology allows us to carry out stress tests. Our results establish under which circumstances the choice of the prior distribution has an influence both on the structure and the stability of the interbank market.

In the rest of this chapter, we review the literature on estimation risk in portfolio theory and systemic risk. Through this historical perspective, we aim to convince the reader that diversification is a powerful, yet delicate, tool to manage both sources of risk.

1.1 Portfolio Choice Theory and Parameter Uncertainty

The implementation of a portfolio, based on a model of risk and return, requires the estimation of expected returns, variance and correlation between assets, as these quantities are not directly observed. Parameter uncertainty is a major source of risk since strategies based on naive estimation methods are known to perform poorly. In portfolio theory, there are two different econometric approaches to treat the problem of parameter uncertainty, namely decision theory and plug-in estimation. In the decision theory approach, each parameter has a prior distribution and uncertainty about these parameters is included in the objective function of the optimisation problem; see Brandt (2010) for a review of the literature on decision theory applied to portfolio choice problems. Therefore, the portfolio rule is optimal with respect to the prior beliefs of the investor.

In this section, we review the literature on plug-in estimation as the method developed in the first part of the thesis is based on a plug-in estimator. In the plug-in estimation approach, the estimators of the parameters, obtained through frequentist or Bayesian inference, are simply plugged in the expression of the optimal portfolio weights. The poor performance of plug-in mean-variance efficient strategies has been largely documented; see, e.g., Jobson and Korkie (1980), Michaud (1989) and Best and Grauer (1991). In particular, the mean-variance portfolio tends to magnify the effect of estimation errors by allocating extreme weights to assets whose parameters are the least accurately estimated. Moreover, the problem worsens as the number of risky assets increases. As a result, investors should hold positions that are untenable under real conditions. Note that a dynamic continuous-time framework raises the same difficulties since part of the portfolio corresponds to the mean-variance efficient term; see Merton (1971).

Three methods have been developed to control the effect of estimation error. They all aim at avoiding the presence of extreme weights in the portfolio to maintain a more uniform repartition of wealth across assets. Following the shrinkage estimation procedure of James and Stein (1961), the first method consists in shrinking the mean to a common value; see Jobson and Korkie (1981) and Jorion (1986). Shrinkage can also be applied to the covariance matrix, for example towards the identity matrix as in

Ledoit and Wolf (2004b), or directly on the portfolio weights, for example towards an equally-weighted portfolio.

The second method imposes a factor structure on the covaration of asset returns to reduce the number of elements to be estimated in the covariance matrix; see the review of Fan et al. (2012). The factors can be identified through an equilibrium model such the Capital Asset Pricing Model of Sharpe (1964), the Intertemporal Capital Asset Pricing Model of Merton (1973), on firm characteristics as in Fama and French (1993), or through a factor analysis; see, e.g., Roll and Ross (1980). In a minimum-variance framework, Chan et al. (1999) show that estimation of the covariance matrix with factors models improves the performance of plug-in strategies. However, they demonstrate that no factor model stands out significantly. Therefore, the identification, the selection, as well as the interpretation, of the factors to be used is still an open debate.

The third method consists in adding constraints to the optimisation problem; see, e.g., Frost and Savarino (1988) and Chopra (1993). Although practitioners are attracted by naive "talmudic" diversification and conservative constraints, such as the no-short sale constraint, Green and Hollifield (1992) show that optimally diversified portfolios may include extreme weights to reduce systematic risk. As such, controlling estimation risk by adding constraints is not necessarily justified from a theoretical perspective. Jagannathan and Ma (2003) resolve this conflict, between theory and traditional asset management practices, by showing that the no-short sale constraint is a shrinkage procedure. Indeed, imposing a no-short sale constraint is tantamount to shrink towards zero large elements of the covariance matrix estimator. Consequently, the associated portfolio is sparse. While excluding assets from the investment rule inherently reduces the effect of estimation error, the exposure to idiosyncratic risk is increased. In particular, DeMiguel et al. (2009a) show that no-short sale constrained portfolios, as well as none of the approaches mentioned above, are able to beat consistently an equally-weighted portfolio. As such, a series of articles relax the no-short sale constraint to the L_1-constraint, which controls the percentage of short positions held in the portfolio; see DeMiguel et al. (2009a), Brodie et al. (2009) and Fan et al. (2012). This constraint also induces sparsity in the portfolio, hence reducing estimation risk. The resulting portfolio outperforms out-of-sample the no-short sale constrained and the equally-weighted

portfolios in terms of Sharpe ratio.

Because the estimation error in the sample mean is large when based on historical returns, mean-variance portfolios usually perform worse out-of-sample than minimun-variance portfolios. Hence, the main focus of the literature, that we have reviewed, is on the estimation of the covariance matrix. However, the mean-variance criterion applies only if returns are normal or the investor has a quadratic utility. Otherwise, for general utility functions in a market with non-normal returns, the portfolio is only an approximation of the true optimal one; see Levy and Markowitz (1979). Therefore, the development of methods that take into account the estimation of mean returns, and deliver a stable out-of-sample performance, still challenges researchers in applied portfolio theory.

1.2 Systemic Risk in Financial Networks

Although there is not a full agreement on the definition of systemic risk in the academic community, we define it as the risk of an external shock spreading to the entire financial network through different channels of endogenous risk[1]. We list here three main channels of systemic risk; for a complete survey, see De Bandt et al. (2009) and Benoit et al. (2015). The first channel is that financial institutions are exposed to similar risk by investing in highly correlated assets; see Haldane (2009). As there is not enough diversification in portfolios at a systemic level, banks fail together. This behaviour can be nonetheless optimal from the banks perspective. For example, Acharya and Yorulmazer (2008) show that banks maximize their profit by failing together as it ensures a common bail-out from the government. Similarly Farhi and Tirole (2012) argue that bail-outs are optimal when the entire system fails, since any bail-out has a fixed cost for the government.

The second channel is liquidity contagion. Liquidity crises are driven by an amplification mechanism which can be decomposed into a loss spiral and a margin spiral; see Brunnermeier and Oehmke (2013). The loss spiral arises as levered financial institutions are very sensitive to a fall in value of their total assets so that they are forced to liquidate their assets at the same time to reach their respective regulatory leverage ratios. This increased

[1]This definition applies to 205 papers, over the past 35 years, surveyed in Benoit et al. (2015)

pressure leads to fire sales, i.e. assets are traded with a discount. Then a margin spiral arises as fire sales are a signal for higher volatility in which case creditors require higher margins amplifying the volatility. Hence, financial institutions are trapped into two contagious spirals reinforcing each other.

The third channel is default contagion, which can be thought of as a domino effect. In the case of an external shock, the default of one bank leads to a reduction of expected payments to other banks. If such banks are too exposed to the interbank market, or their assets are also significantly reduced by the external shock, they will not be able to cover their losses and they will also default. The spread of balance sheet contagion depends fundamentally on the diversification of banks' bilateral exposures. Under different modeling assumptions, Allen and Gale (2000) and Freixas et al. (2000) show that a complete network in which all banks are connected is less prone to contagion than a network in which banks are minimally connected in a circular credit chain fashion. Moreover, in Freixas et al. (2000), a large enough number of banks ensures the stability of a complete network, while the number of banks has no effect on the stability of a circular network.

Based on the Erdős and Rényi (1959) random graph model, in which banks are exposed to each other with a fixed probability, Nier et al. (2007) use simulations to highlight patterns affecting the stability of the network. In particular, their results demonstrate that there is a non monotonic M-shaped relation between the number of defaults and the number of interbank connections. They also identify additional parameters such as the capitalisation of banks, the size of exposures to the lending interbank network as well their concentration as determinant factors for the stability of the interbank lending market. Note that the three articles cited above assume that the default contagion is triggered by an exogeneous idiosyncratic shock on a single bank. This assumption is not likely to provide a full picture of the network stability in times of crises as several institutions are affected simultaneously. In a recent article, Acemoglu et al. (2015) actually show that the capacity of a complete network to absorb defaults depends on the size and the number of shocks applied to the system.

Based on data of national banking systems, a significant number of empirical papers has been published; see, e.g., Furfine (2003) for the United States, Wells (2004) for Germany, Elsinger et al. (2006) for the UK, Van and Liedorp (2006) for the Netherlands, Degryse and Nguyen (2007) for

Belgium. They analyse the stability of the respective networks through different methods of stress test and these papers establish that contagion due interbank lending is limited. Nevertheless, Upper (2011) argues that no clear cut conclusions can be made from these studies given the difference of network structure across countries as well as the variety of simulation methods. Moreover, he identifies two main methodological drawbacks on which most of the studies rely. Similarly to the early theoretical literature, only balance sheet contagion triggered by a single default is analysed. Next, because of the censorship of individual banks' exposures in most countries, the stress tests require an estimation of the actual network. The construction of the network relies on minimising the Kullback-Leibler divergence with respect to a prior network in which liabilities are evenly distributed. The network obtained through this method is complete in the sense of Allen and Gale (2000) and it is a feature at odds with fully observable networks; see Upper and Worms (2004), Cocco et al. (2009), and Cont et al. (2013). Indeed, observed networks are usually sparse and have a core periphery structure.

To answer the need of an estimation method providing a more realistic network structure, Gandy and Veraart (2015) develop a Bayesian approach in which each exposure follows an exponential prior conditionally on a prior probability of existence. Using a Markov Chain Monte Carlo method, they draw samples from the joint distribution of the individual liabilities conditionally on the information provided by balance sheets. Their model allows for full heterogeneity in the parameters and also for tiered network structures.

Part I

Optimal Diversification in Presence of Parameter Uncertainty

Chapter 2

The Merton Portfolio with Unknown Drift

> *The first myth is that this research is only about how to "beat the market".*

- Ioannis Karatzas and Steven E. Shreve, *Methods of Mathematical Finance*

2.1 Introduction

We consider a financial market consisting of one risk-free asset and a large number of risky assets following a multi-dimensional geometric Brownian motion. In this market, we assume that there is an investor with a power utility function seeking to maximise the expected utility of her final wealth. If all parameters are known, Merton (1971) shows that the optimal fraction of wealth allocated in each risky asset is characterised by the drift and the volatility matrix of assets returns[1].

These two quantities are not directly observable, and they are typically replaced by estimates computed from historical data[2]. For instance, we can plug simple estimators, such as the sample or the maximum likelihood estimators, into the analytical expression of the optimal portfolio weights.

[1] When all parameters are known, Merton (1971) shows that the utility maximisation problem is reduced to a mean-variance problem.

[2] It is also possible to use estimates of the covariance matrix, or equivalently the volatility matrix, based on forward-looking information; see the recent work of Kempf et al. (2015) for a successful application in portfolio selection.

However, the resulting plug-in strategy is likely to differ considerably from the true optimal strategy. The main source of error comes from the estimation of the drift. Indeed, the accuracy of the estimator of the drift does not depend on the frequency of observations but on the length of the observation interval. To obtain a reasonable precision for estimators of the drift, one needs to use a very long period of observation; see Merton (1980).

Moreover, when the number of risky assets is large, the problem becomes even more prominent as estimation errors accumulate across the positions in the risky assets. Based on Merton (1971) and restricted to a logarithmic utility function, Gandy and Veraart (2013) show that the expected utility associated with the plug-in strategy can degenerate to $-\infty$ as the number of assets increases[3] .

In this chapter, we extend the approach of Gandy and Veraart (2013) to general power utility functions. By taking into account the coefficient of relative risk aversion (RRA) explicitly, we are able to pin down the influence of both risk aversion and estimation risk on the expected utility.

Next, we impose an L_1-constraint on the portfolio weights. This constraint induces sparsity in the portfolio, i.e., most of the weights are set to zero, and it naturally reduces the accumulation of estimation error. Our objective is to select the bound of the L_1-constraint which gives the optimal degree of sparsity in the portfolio.

Holding a sparse portfolio is known to be an efficient way to reduce exposure to estimation risk[4]. In a minimum-variance framework with a large number of risky assets (500 stocks), Jagannathan and Ma (2003) show that the no-short sale constraint is likely to improve the performance of the plug-in strategy based on the sample covariance matrix. In this case, the estimation error is large, and constraining the amount of short positions can help because the associated plug-in strategy is sparse. However, they also demonstrate that the portfolio has too many weights set to zero. As a

[3]In the static framework it is well known that plug-in mean-variance efficient portfolios perform poorly out-of-sample. In particular, no standard plug-in strategy, outperforms consistently the equally-weighted portfolio benchmark in terms of Sharpe ratio, certainty equivalence, and turnover; see DeMiguel et al. (2009b) and references therein.

[4]Sparsity could also be induced by considering an optimal subset of available assets, i.e., by solving an L_0-norm problem. Even with a quadratic objective, this is a NP-hard problem; see Natarajan (1995) and Bach et al. (2010). In the portfolio selection literature, simple convex constraints such as L_1-constraints inducing sparsity are favored because of their computational tractability.

result, the poor performance of only a few assets can dramatically influence the performance of the portfolio.

To overcome this problem DeMiguel et al. (2009a), Brodie et al. (2009) and Fan et al. (2012) generalise the no-short sale constraint to the less restrictive L_1-constraint. The L_1-constraint is more flexible because it imposes an upper bound on the proportion of short positions. Thus, the set of admissible portfolios is augmented by relaxing the constraint while keeping a reasonable sparsity. With a suitable bound the constrained plug-in strategy has a smaller out-of-sample variance than benchmark portfolios such as the no-short sale minimum variance portfolio and the equally-weighted portfolio; see DeMiguel et al. (2009a), Brodie et al. (2009) and Fan et al. (2012). Moreover, it also outperforms strategies based on James-Stein estimators in the static framework of DeMiguel et al. (2009a) and in the dynamic framework of Gandy and Veraart (2013).

The identification of a good bound for the L_1-constraint is decisive for the performance of the constrained plug-in strategy. The empirical results of Fan et al. (2012) demonstrate that the out-of-sample variance of the constrained minimum-variance plug-in strategy is convex in the bound of the L_1-constraint. In particular, the variance can be reduced by half by moving from the no-short sale constraint to the optimal L_1-constraint. Relaxing the constraint further increases the variance up to twenty-five percent. Therefore, a relatively small interval has to be identified for the bound. Alternatively, DeMiguel et al. (2009a) suggest to select the bound, which minimises the variance, using the cross-validation method. None of these papers characterise the structure of the constrained strategy nor do they justify explicitly the existence of an optimal bound. We address these problems as outlined below.

In Section 2.2, we introduce the general setup and recall the optimal investment strategies when parameters are assumed to be observable. We then drop the observability assumption of the drift in Section 2.3. We estimate the drift vector with the maximum likelihood estimator (MLE) and we assume the volatility matrix to be known. This assumption is justified by assuming that prices are observed continuously and hence the volatility is directly observable from the quadratic variation of the logarithmic asset price but the drift is not.

In Section 2.3, Theorem 2.3.2, we show that the expected utility of the

plug-in strategy is equal to the theoretical expected utility of the optimal strategy with known parameters times a loss factor. For a fixed investment horizon, this loss factor is increasing in the number of risky assets. In particular, the expected utility can degenerate as the number of risky assets increases. When the drift is estimated, the diversification of the plug-in strategy clearly hurts.

In Section 3.1, we introduce the L_1-norm as a way to constrain investment weights. We demonstrate that the sparsity of the optimal L_1-constrained strategy depends to a large extent on the coefficient of RRA. To understand the relation between the structure of the constrained strategy and the coefficient of RRA, we provide the analytical solution of the optimal L_1-constrained portfolio for independent risky assets in Theorem 3.2.2. In this case only the assets with the largest absolute excess returns are selected. The L_1-constrained portfolio rule consists in shrinking the excess returns towards zero by an intensity which is the same for all assets. If the absolute excess return of an asset is smaller than this constant, we do not invest in it. The number of assets invested in and the shrinkage intensity depend both on the bound of the L_1-constraint and the level of risk aversion. The L_1-constrained strategy becomes less sparse as the coefficient of RRA increases, both for the true and the estimated drift. In terms of diversification, increasing the coefficient of RRA is similar to relaxing the constraint.

When facing parameter uncertainty, we show, in Proposition 3.3.2, that imposing an L_1-constraint rules-out degeneracy of the expected utility of the plug-in strategy. Indeed, even if the number of assets goes to infinity, the L_1-constrained portfolio remains sparse, which prevents accumulation of estimation error.

With a fixed number of assets, the key point is to analyse the trade-off between the loss due to the lack of diversification, introduced by the L_1-constraint, and the loss due to estimation error. As we relax the constraint, the loss due to under-diversification decreases, while the loss due to estimation increases. These two losses move in opposite directions. Depending on the structure of the drift, the volatility matrix, and the method of estimation, there can be an L_1-bound which minimises the total loss of the constrained plug-in strategy for each level of risk aversion.

For a general volatility matrix, we do not have closed form results for

optimal L_1-constrained strategies. Therefore we use in Section 3.4 a simulation study to investigate the properties and the performance of the L_1-constrained portfolio, when assets are correlated. Similarly to the independent case, the L_1-constrained strategy becomes less sparse as the coefficient of RRA increases. We present numerical examples which show that the L_1-bound can be chosen in an optimal way to minimise the loss due to estimation. This optimal choice depends crucially on the level of risk aversion.

Finally, in Chapter 4, we consider a universe of stocks based on the S&P 500 from 2006 to 2011 and we investigate the out-of-sample performance of the unconstrained and the constrained plug-in strategies. We trade daily over one-month-long intervals to test the strategies. We calibrate the optimal bound of the constrained strategy based on two different numerical methods. Our results confirm that the unconstrained strategy has very unstable returns. We also demonstrate that imposing the appropriate L_1-constraint improves greatly the performance of the plug-in strategy. While, on average, the constrained strategy has a higher variance than the equally-weighted portfolio, it delivers a utility of terminal wealth which is in the same range. Hence, the L_1-constrained plug-in strategy has a comparable performance to the equally-weighted portfolio, even when the drift is estimated.

2.2 Model Setup

We consider a financial market where trading takes place continuously over a finite time interval $[0, T]$ for $0 < T < \infty$. The market consists of one risk-free asset with time-t price $S_0(t)$ and d risky assets with time-t price $S_i(t)$ for $i = 1, \ldots, d$. Their dynamics are given by

$$\begin{aligned} dS_0(t) &= S_0(t)\, r dt, & S_0(0) &= 1, \\ dS_i(t) &= S_i(t)\left(\mu_i dt + \sum_{j=1}^{d} \sigma_{ij} dW_j(t)\right), & S_i(0) &> 0, \text{ for } i = 1, \ldots, d, \end{aligned}$$

where $r \geq 0$ is the constant interest rate, $\mu = (\mu_1, \ldots, \mu_d)^\top$ is the constant drift and $\sigma = (\sigma_{ij})_{1 \leq i,j \leq d}$ is the constant $d \times d$ volatility matrix. We assume that σ is of full rank. Furthermore, $W = (W_1, \ldots, W_d)^\top$ is a standard d-dimensional Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We denote by $X^{\pi}(t)$ the investor's wealth at time t when using strategy π, which is given by

$$dX^{\pi}(t) = \sum_{i=0}^{d} \pi_i(t) X^{\pi}(t) \frac{dS_i(t)}{S_i(t)},$$

for a constant initial wealth $X^{\pi}(0) = X(0) > 0$. Here, $\pi_i(t)$ denotes the fraction of wealth invested in the ith asset at time t. Hence, $\sum_{i=0}^{d} \pi_i(t) = 1$ for all t. Using $\pi_0(t) = 1 - \sum_{i=1}^{d} \pi_i(t)$ and setting $\pi(t) = (\pi_1(t), \ldots, \pi_d(t))^{\top}$, we obtain

$$\frac{dX^{\pi}(t)}{X^{\pi}(t)} = \left(r + \pi(t)^{\top}(\mu - r1)\right) dt + \pi^{\top}(t) \sigma dW(t), \tag{2.2.1}$$

where $1 = (1, \ldots, 1)^{\top} \in \mathbb{R}^d$. For strategies $(\pi(t))_{t \geq 0}$ that are adapted to the filtration $(\mathcal{F}(t))_{t \geq 0}$ with $\mathcal{F}(t) = \sigma(W(s), s \leq t)$ and that are sufficiently integrable, the solution of (2.2.1) is

$$X^{\pi}(T) = X(0) \exp\left(\int_0^T \left(r + \pi(t)^{\top}(\mu - r1) - \frac{1}{2}\pi(t)^{\top} \Sigma \pi(t)\right) dt + \int_0^T \pi(t)^{\top} \sigma dW(t)\right),$$

where $\Sigma = \sigma\sigma^{\top}$. Note that Σ is positive definite.

2.2.1 The Investor's Objective and the Classical Solution

We consider an investor with constant relative risk aversion (CRRA) utility function of the type

$$U_{\gamma}(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma} & \text{for } \gamma > 1, \\ \log(x) & \text{for } \gamma = 1, \end{cases}$$

for $x > 0$, where γ is the coefficient of relative risk aversion (RRA).

The investor's goal is to maximise

$$V_{\gamma}(\pi | \mu, \sigma) := \mathbb{E}[U_{\gamma}(X^{\pi}(T))] \tag{2.2.2}$$

over strategies π which are sufficiently integrable so that $V_{\gamma}(\pi | \mu, \sigma)$ is well-defined. We call such strategies admissible. We use the notation $V_{\gamma}(\pi | \mu, \sigma)$

to emphasize that the objective function depends on γ, μ and σ.

Merton (1971) shows that the optimal strategy, denoted by π^* and called the Merton ratio, consists in holding a constant proportion in each asset:

$$\pi^*(t) = \frac{1}{\gamma}\Sigma^{-1}(\mu - r1) \quad \forall t \in [0,T].$$

Therefore, the Merton ratio also maximises the mean-variance term

$$M_\gamma(\pi) = \pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi.$$

Finally, the corresponding expected utility is given by

(2.2.3)
$$V_\gamma(\pi^*|\mu,\sigma) = \begin{cases} K_\gamma \exp\left(\frac{(1-\gamma)T}{2\gamma}(\mu - r1)^\top \Sigma^{-1}(\mu - r1)\right), & \gamma > 1, \\ K_\gamma + \frac{T}{2}(\mu - r1)^\top \Sigma^{-1}(\mu - r), & \gamma = 1, \end{cases}$$

with

(2.2.4)
$$K_\gamma = \begin{cases} \frac{X(0)^{1-\gamma}}{1-\gamma}\exp((1-\gamma)rT), & \gamma > 1, \\ \log(X(0)) + rT, & \gamma = 1. \end{cases}$$

Regardless of the magnitude of the initial wealth $X(0)$, K_γ is always strictly negative for $\gamma > 1$.

2.2.2 The Effect of Diversification for Known Parameters

When the true parameters are known, the investor optimally diversifies her portfolio by investing the corresponding Merton ratio in each stock. As more stocks become available, one simply maximises over a larger set of strategies and the expected utility increases at a rate which, by (2.2.3), depends on γ and the growth of the quadratic form $(\mu - r1)^\top \Sigma^{-1}(\mu - r1)$. Moreover, if the spectrum of the matrix Σ^{-1} is bounded from above and away from zero, analysing the convergence of the Euclidean norm of excess returns $||\mu - r1||_2$ is sufficient to characterise the asymptotic behaviour of the expected utility.

When studying the expected utility as a function of the number of risky assets d, we are in effect considering a sequence of markets. This sequence is built with a market containing the first $1, ..., d$ risky assets in the same order and then a new risky asset is included and considered as the $(d+1)$st

asset. In this setting, the drift, the volatility matrix, the covariance matrix and the Brownian motion are denoted, for the market with d risky assets, by $\mu^{(d)}$, $\sigma^{(d)}$, $\Sigma^{(d)}$, and $W^{(d)}$ respectively. A portfolio strategy in the market with d risky assets is denoted by $\pi^{(d)}$.

Proposition 2.2.1. *Let* $\left(\Sigma^{(d)}\right)_{d\geq 1} \subset \mathbb{R}^{d\times d}$ *be such that its eigenvalues* $\lambda_i^{(d)}$, $i = 1, \ldots, d$, *satisfy*

$$\underline{m}_\lambda = \lim_{d\to\infty} \min_{i=1,..,d} 1/\lambda_i^{(d)}, \qquad \overline{m}_\lambda = \lim_{d\to\infty} \max_{i=1,..,d} 1/\lambda_i^{(d)}. \tag{2.2.5}$$

Suppose that $\underline{m}_\lambda > 0$ *and* $\overline{m}_\lambda < \infty$. *Then, for* $\gamma > 1$ *and for all* d *we have that*

$$\begin{aligned} &K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right) \\ &\leq V_\gamma\left((\pi^*)^{(d)}\,|\mu^{(d)}, \sigma^{(d)}\right) \\ &\leq K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right), \end{aligned}$$

and for $\gamma = 1$

$$\begin{aligned} K_1 + \frac{T}{2}\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2 \leq V_1\left((\pi^*)^{(d)}\,|\mu^{(d)}, \sigma^{(d)}\right) \\ \leq K_1 + \frac{T}{2}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2. \end{aligned}$$

Proof of Proposition 2.2.1. The matrix $\left(\Sigma^{-1}\right)^{(d)}$ is symmetric positive definite with spectrum $\left(1/\lambda_i^{(d)}\right)_{i=1,\ldots,d}$ and it can be diagonalised. Since the change of basis of this matrix is orthogonal, the following inequalities hold for each $x \in \mathbb{R}^d$,

$$\underline{m}_\lambda ||x||_2^2 \leq \min_{i=1,\ldots,d} \frac{1}{\lambda_i^{(d)}} ||x||_2^2 \leq x^\top \left(\Sigma^{-1}\right)^{(d)} x \leq \max_{i=1,\ldots,d} \frac{1}{\lambda_i^{(d)}} ||x||_2^2 \leq \overline{m}_\lambda ||x||_2^2.$$

Thus,

$$\begin{aligned} \underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2 \leq \left(\mu^{(d)} - r1^{(d)}\right)^\top \left(\Sigma^{-1}\right)^{(d)} \left(\mu^{(d)} - r1^{(d)}\right) \\ \leq \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2, \end{aligned}$$

and, for $\gamma > 1$,

$$\begin{aligned} & K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right) \\ & \leq V_\gamma\left((\pi^*)^{(d)}\,|\mu^{(d)}, \sigma^{(d)}\right) \\ & \leq K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right). \end{aligned}$$

The argument is similar for $\gamma = 1$. □

As $||\mu^{(d)} - r1^{(d)}||_2$ is an increasing positive sequence in d, it always admits a limit. If this limit is finite, the expected utility is bounded away from zero. If the limit is infinite, the expected utility reaches zero and the positive effect of diversification is fully exploited. The case $\gamma = 1$ is similar.

2.3 Performance of Plug-in Strategies

When asset prices are continuously observed, we can obtain the true volatility matrix σ since the quadratic variation of the log-stock price is observable. However, this is not the case for the drift μ. Indeed, the accuracy of the estimation of the drift depends on the length of the estimation period and not on the frequency of observations. We use the maximum likelihood estimator (MLE) of the drift over the observation period $[-t_{obs}, 0]$ for a constant $t_{obs} > 0$. The MLE for μ_i, $i = 1, \ldots, d$, is given by

$$\hat{\mu}_i = \frac{\log\left(S_i\left(0\right)\right) - \log\left(S_i\left(-t_{obs}\right)\right)}{t_{obs}} + \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2. \tag{2.3.1}$$

Based on the estimator $\hat{\mu}$, one can implement the time-constant plug-in strategy

$$\hat{\pi} = \frac{1}{\gamma}\Sigma^{-1}\left(\hat{\mu} - r1\right). \tag{2.3.2}$$

$\hat{\pi}$ is an unbiased Gaussian estimator of π^*, in particular

$$\hat{\pi} \sim \mathcal{N}\left(\pi^*, V_0^2\right) \text{ with } V_0^2 = \frac{1}{\gamma^2 t_{obs}}\Sigma^{-1}. \tag{2.3.3}$$

Furthermore, the expected utility of the plug-in strategy is given by the moment generating function of the mean-variance term $M_\gamma(\hat{\pi})$ as the following lemma shows.

Lemma 2.3.1. *Let $\gamma > 1$,*

$$V_\gamma(\hat{\pi}|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)\, T M_\gamma(\hat{\pi})\right)\right]. \tag{2.3.4}$$

The lemma is proved in Appendix A.

We now characterise the loss in expected utility due to estimation when implementing the plug-in strategy $\hat{\pi}$, based on the MLE of the drift.

Theorem 2.3.2. *Let $\gamma > 1$ and $t_{obs} > T$. Then the expected utility of the plug-in strategy $\hat{\pi}$ is given by*

$$V_\gamma(\hat{\pi}|\mu,\sigma) = L_\gamma(\hat{\pi},\pi^*)\, V_\gamma(\pi^*|\mu,\sigma) \tag{2.3.5}$$

with

$$L_\gamma(\hat{\pi},\pi^*) = \left(1 + \frac{(1-\gamma)\, T}{\gamma t_{obs}}\right)^{-\frac{d}{2}}. \tag{2.3.6}$$

The theorem is proved in Appendix A.

For the case $\gamma = 1$, Gandy and Veraart (2013) have shown that the loss is linear in d:

$$V_1(\hat{\pi}|\mu,\sigma) = V_1(\pi^*|\mu,\sigma) - L_1(\hat{\pi},\pi^*) \text{ with } L_1(\hat{\pi},\pi^*) = \frac{T}{2t_{obs}} d > 0. \tag{2.3.7}$$

While the loss factor does not depend on the value of the true parameters μ and Σ, it is an increasing function of the number of risky assets d.

Since $\hat{\pi}$ is a consistent estimator of π^*, the expected utility of the plug-in strategy converges to the expected utility of the optimal strategy as the length of the observation period $t_{obs} \to \infty$.

By (2.3.3), the accuracy of the plug-in strategy depends on the length of the observation period and the rate of convergence in (2.3.5) is very slow. For instance, with $T = 1$, $\gamma = 2$ and $d = 200$ risky assets, we need three centuries of observations, $t_{obs} = 300$, to get a loss factor close to one, $L_\gamma(\hat{\pi},\pi^*) \approx 1.18$. We will see from (2.3.9), that this corresponds to a 15.3%

loss in certainty equivalent. Therefore, the loss can be reduced significantly only by taking a very long estimation period and, for a feasible estimation period, using a plug-in strategy $\hat{\pi}$ results in a poor expected utility.

The following corollary gives a sufficient condition for the expected utility to degenerate.

Corollary 2.3.3. *Let $\gamma \geq 1$, $t_{obs} > T$ and suppose that the sequence of eigenvalues of $\Sigma^{(d)}$ verifies (2.2.5). If $||\mu^{(d)} - r1^{(d)}||_2^2$ is in $o(d)$, then*

$$V_\gamma\left(\hat{\pi}^{(d)}|\mu^{(d)}, \sigma^{(d)}\right) \to -\infty \text{ as } d \to \infty.$$

Proof of Corollary 2.3.3. Assume $t_{obs} > T$ so that the expected utility of the plug-in strategy is well-defined. For $\gamma > 1$ we see from (2.3.6) that

$$\lim_{d\to\infty} L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) = \infty.$$

By Proposition 2.2.1 and because the loss factor is positive, the expected utility of the plug-in strategy is bounded by above as follows:

$$\begin{aligned} V_\gamma\left(\hat{\pi}^{(d)}|\mu^{(d)}, \sigma^{(d)}\right) \leq & L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) \\ & \cdot K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda||\mu^{(d)} - r1^{(d)}||_2^2\right). \end{aligned}$$

If the right-hand side of the inequality goes to $-\infty$, the expected utility $V_\gamma(\hat{\pi}|\mu, \sigma)$ degenerates. We have the following equivalences :

$$\begin{aligned} & \lim_{d\to\infty} L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda||\mu^{(d)} - r1^{(d)}||_2^2\right) = -\infty \\ \Leftrightarrow & \lim_{d\to\infty} L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda||\mu^{(d)} - r1^{(d)}||_2^2\right) = \infty \\ \Leftrightarrow & \lim_{d\to\infty} \log\left(L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda||\mu^{(d)} - r1^{(d)}||_2^2\right)\right) = \infty \\ \Leftrightarrow & \lim_{d\to\infty} -\frac{d}{2}\log\left(1 + \frac{(1-\gamma)\,T}{\gamma t_{obs}}\right) + \frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda||\mu^{(d)} - r1^{(d)}||_2^2 = \infty \\ \Leftrightarrow & \lim_{d\to\infty} -\frac{d}{2}\log\left(1 + \frac{(1-\gamma)\,T}{\gamma t_{obs}}\right) - \frac{(\gamma-1)\,T}{2\gamma}\overline{m}_\lambda||\mu^{(d)} - r1^{(d)}||_2^2 = \infty \end{aligned}$$

We set

$$x_d = -\frac{d}{2}\log\left(1 + \frac{(1-\gamma)\,T}{\gamma t_{obs}}\right), \qquad y_d = \frac{(\gamma-1)\,T}{2\gamma}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2.$$

Hence, as $||\mu - r1||_2^2$ is in $o(d)$, $\lim_{d\to\infty} x_d - y_d = \infty$. □

For instance, with $\mu_i^{(d)} - r = 1/i$, $i = 1, \ldots, d$, the sequence $||\mu^{(d)} - r1^{(d)}||_2^2$ has a finite limit and the corollary applies.

It is already well-known that strategies based on the MLE of the drift perform poorly. It is common to obtain extreme positions due to estimation error and, for a high-dimensional problems, the accumulation of error leads to a large loss in expected utility. What is new here is a full description of the loss due to estimation as a function of the coefficient of RRA and the number of risky assets. Additionally we provide in Corollary 2.3.3 a sufficient condition for the degeneracy of the expected utility as $d \to \infty$.

2.3.1 Measures of Economic Loss

Theorem 2.3.2 establishes an analytic relationship between the expected utility obtained from using the optimal strategy with known drift and the expected utility from using a plug-in strategy. In general it is hard to interpret different levels of expected utility, as utility functions describe a preference ordering which is invariant to linear transformations. Therefore we provide some discussion on how one can measure economic loss that is due to using a plug-in strategy rather than an optimal strategy.

Mean-variance Loss Function

For known parameters, problem (2.2.2) is equivalent to maximising the (instantaneous) mean-variance term. When deviating from the optimal strategy, a standard choice to measure economic loss is the mean-variance loss function [5]:

$$L_\gamma^M(\hat{\pi}, \pi^*) = M_\gamma(\pi^*) - \mathbb{E}\left[M_\gamma(\hat{\pi})\right].$$

[5] See, e.g., Kan and Zhou (2007) or Tu and Zhou (2011) and the references therein.

The mean-variance loss is then given by

$$L_{\gamma}^{M}\left(\hat{\pi},\pi^{*}\right)=\frac{d}{2\gamma t_{obs}}. \tag{2.3.8}$$

$L_{\gamma}^{M}\left(\hat{\pi},\pi^{*}\right)$ captures the fact that a smaller fraction of wealth is invested in the risky assets as γ increases[6].

$L_{\gamma}^{M}\left(\hat{\pi},\pi^{*}\right)$ does not measure estimation risk consistently, however, if one considers an investor with CRRA power utility function for $\gamma > 1$. Indeed, by Lemma 2.3.1, the expected utility of final wealth is proportional to the moment generating function of the mean-variance term. In general there is a non-monotonic relation between the moment generating function and the expectation of the mean-variance term. Hence, the equivalence between our setting and the mean-variance approach does not hold when the implemented strategy is random.

Certainty Equivalents and Efficiency

To account for both sources of risk consistently, namely the risk due to the driving Brownian motions and the risk due to parameter uncertainty, the loss due to estimation has be to be quantified in terms of expected utility. A strategy $\hat{\pi}$ is suboptimal if it generates a loss in expected utility, $V_{\gamma}\left(\hat{\pi}|\mu,\sigma\right)\leq V_{\gamma}\left(\pi^{*}|\mu,\sigma\right)$.

We now look at the loss in certainty equivalents and show the relation with the relative loss in expected utility.

Definition 2.3.4. *For the optimal strategy and the plug-in strategy, the* certainty equivalents *are the scalar quantities* $CE_{\gamma}^{\hat{\pi}}$ *and* $CE_{\gamma}^{\pi^{*}}$ *respectively such that*

$$U_{\gamma}\left(CE_{\gamma}^{\hat{\pi}}\right)=V_{\gamma}\left(\hat{\pi}|\mu,\sigma\right)\ \text{ and }U_{\gamma}\left(CE_{\gamma}^{\pi^{*}}\right)=V_{\gamma}\left(\pi^{*}|\mu,\sigma\right).$$

The certainty equivalents are the cash amounts delivering the same utility as the corresponding strategies. From the definition of CRRA utility functions one obtains immediately that, for the certainty equivalents $CE_{\gamma}^{\hat{\pi}}$ and $CE_{\gamma}^{\pi^{*}}$ the following relationship holds true:

[6] Jorion (1986) uses the relative mean-variance loss function $\frac{L_{\gamma}^{M}\left(\hat{\pi},\pi^{*}\right)}{|M_{\gamma}(\pi^{*})|}$. In this case, the loss function does not depend on the coefficient of RRA.

$$\frac{CE_\gamma^{\hat{\pi}}}{CE_\gamma^{\pi^*}} = \begin{cases} L_\gamma\left(\hat{\pi}, \pi^*\right)^{\frac{1}{1-\gamma}}, & \text{for } \gamma > 1, \\ \exp\left(-L_\gamma\left(\hat{\pi}, \pi^*\right)\right), & \text{for } \gamma = 1. \end{cases} \tag{2.3.9}$$

The ratio of certainty equivalents can also be interpreted in terms of the efficiency measure that has been introduced in the literature to compare different expected utilities; see Rogers (2001). Along the lines of (Rogers, 2001, Definition 1) we define the efficiency in our context as follows.

Definition 2.3.5. *The* efficiency $\Theta_\gamma(\pi)$ *of an investor with relative risk aversion* γ *using strategy* π *relative to the Merton investor (who uses the optimal strategy* π^**) is the amount of wealth at time 0 which the Merton investor would need to obtain the same maximised expected utility at time* T *as the investor with strategy* π *who started at time 0 with unit wealth.*

Using the results of Theorem 2.3.2 we obtain that, for CRRA utility functions, the ratio of the certainty equivalents (2.3.9) is exactly the efficiency.

Theorem 2.3.6. *The efficiency of the investor who uses the simple plug-in strategy (2.3.2) is given by*

$$\Theta_\gamma(\hat{\pi}) = \begin{cases} L_\gamma\left(\hat{\pi}, \pi^*\right)^{\frac{1}{1-\gamma}}, & \textit{for } \gamma > 1, \\ \exp\left(-L_1\left(\hat{\pi}, \pi^*\right)\right), & \textit{for } \gamma = 1. \end{cases}$$

Proof of Theorem 2.3.6. Using the results of Theorem 2.3.2, and the expressions (2.2.3) and (2.2.4), we obtain, for $\gamma > 1$,

$$\begin{aligned} &\frac{\Theta_\gamma(\hat{\pi})^{1-\gamma}}{1-\gamma} \exp\left((1-\gamma) rT\right) \exp\left(\frac{(1-\gamma) T}{2\gamma} (\mu - r1)^\top \Sigma^{-1} (\mu - r1)\right) \\ &\quad = \frac{L_\gamma\left(\hat{\pi}, \pi^*\right)}{1-\gamma} \exp\left((1-\gamma) rT\right) \exp\left(\frac{(1-\gamma) T}{2\gamma} (\mu - r1)^\top \Sigma^{-1} (\mu - r1)\right) \\ &\iff \Theta_\gamma(\hat{\pi}) = L_\gamma\left(\hat{\pi}, \pi^*\right)^{\frac{1}{1-\gamma}}. \end{aligned}$$

For $\gamma = 1$, we obtain

$$\begin{aligned}
&\log\left(\Theta_1\left(\hat{\pi}\right)\right) + rT + \frac{T}{2}\left(\mu - r1\right)^\top \Sigma^{-1}\left(\mu - r\right) \\
&\quad = \log\left(1\right) + rT + \frac{T}{2}\left(\mu - r1\right)^\top \Sigma^{-1}\left(\mu - r\right) - L_1\left(\hat{\pi}, \pi^*\right) \\
&\iff \Theta_1\left(\hat{\pi}\right) = \exp\left(-L_1\left(\hat{\pi}, \pi^*\right)\right).
\end{aligned}$$

□

Remark 2.3.7. Theorem 2.3.6 remains true for any (possibly random) strategy constant in time, sufficiently integrable and independent of $W(T)$; see Appendix B. In particular, this is the case of the constrained strategies considered in Chapter 3.

We see that for $\gamma > 1$ the relative loss factor $L_\gamma\left(\hat{\pi}, \pi^*\right)$ is the efficiency raised to power $1/\left(1-\gamma\right)$ and, for $\gamma = 1$, the absolute loss $L_1\left(\hat{\pi}, \pi^*\right)$ is minus the logarithm of the corresponding efficiency. Hence, there is a one-to-one monotonic relation between the relative loss in expected utility and efficiency. Furthermore, for $\gamma > 1$, the loss factor $L_\gamma\left(\hat{\pi}, \pi^*\right)$ is always greater than one and the efficiency is always smaller than one. When there is no estimation risk, both quantities are equal to one.

As the loss factor is increasing in the number of assets and the power $1/\left(1-\gamma\right)$ is negative, the efficiency is sharply decreasing with the number of assets. Namely, the more assets are available the lower the initial wealth of the Merton investor can be to obtain the same expected utility as the plug-in investor. This is illustrated in Figure 2.1.

While the loss factor $L_\gamma\left(\hat{\pi}, \pi^*\right)$ measures loss in expected utility consistently for a fixed level of risk aversion, its magnitude should not be compared across different levels of risk aversion. The expected utility of the plug-in investor is characterised as the product of the loss factor and the expected utility of the Merton investor but both quantities depend on the investor's risk aversion γ. Although the loss factor itself is an increasing function in γ, this fact is not sufficient to draw conclusions on how expected utilities of plug-in investors with different risk-aversion parameters relate to each other.

We therefore look at the efficiency of the plug-in investor as a function of γ. For a fixed number of risky assets, the efficiency is an increasing function of γ. Hence, the plug-in strategy becomes more efficient as γ increases. If

we consider two plug-in investors with different parameters of relative risk aversion γ_1 and γ_2, with $\gamma_1 > \gamma_2$, the more risk averse investor, i.e., the one with risk aversion γ_1, will be more efficient relative to the Merton investor than the plug-in investor who is less risk averse with risk aversion γ_2. The reason for this behaviour is that the more risk averse investor invests a smaller fraction of his wealth in the risky assets. This is in line with the behaviour of the mean-variance loss function in (2.3.8), in which the effect of estimation is also reduced as the coefficient of RRA increases.

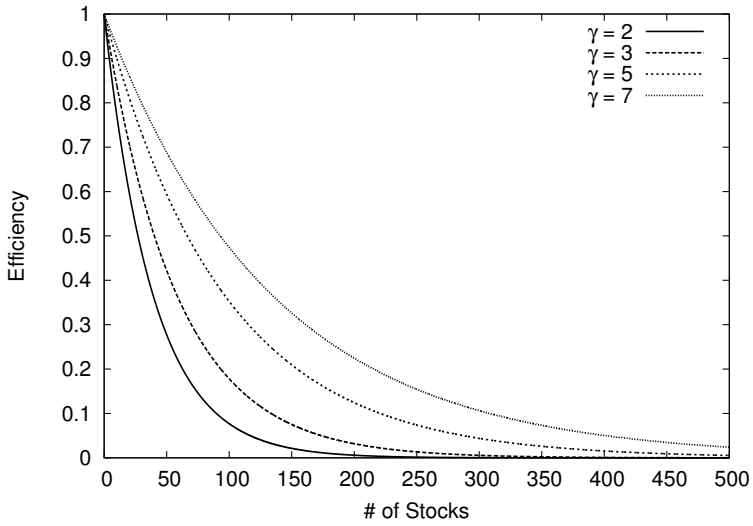


Figure 2.1: Plot of the efficiency of the plug-in investor relative to the Merton investor as a function of the number of risky assets d for different levels of risk aversion γ. For a given γ, the efficiency depends only the number of risky assets, the investment horizon $T = 1$ and the observation period $t_{obs} = 10$.

2.3.2 Drift versus Covariance Estimation

So far we have only considered the estimation problem of the drift and assumed that the matrix Σ is observable. We have justified at the beginning of Section 2.3 that as long as we are in continuous time the quadratic variation of the stock price is observable and hence Σ is known.

As soon as we move to a discrete-time setting the situation changes. If we assume that observing the asset prices continuously is no-longer possible, the covariance matrix Σ also needs to be estimated. Hence any discussion on estimating the covariance matrix is linked to the discussion on discrete

versus continuous-time settings.

The effects of discrete trading have already been studied by Rogers (2001) and a detailed analysis on discrete trading and observations in the context of parameter uncertainty is available in Bäuerle et al. (2013). Note that one cannot just suitably discretise a strategy that is optimal in continuous time to obtain a strategy that is optimal in discrete time. A strategy that is optimal in discrete time has different characteristics, e.g., short selling is forbidden. Furthermore, Bäuerle et al. (2013) show that, with parameter uncertainty on the drift and the covariance, the expected utility of the "discrete trader" does not converge to the expected utility of the "continuous trader", as the time step goes to zero.

Note that this is in contrast to the static Markowitz mean-variance approach, where there is no rebalancing. In a static mean-variance context, the structure and the performance of plug-in strategies using estimators for both the mean and the covariance matrix has been studied in depth; see, e.g., El Karoui (2010). Since these results are already available and we are studying a continuous-time setting, we will not analyse the theoretical problem of the estimation of the covariance matrix any further here.

Chapter 3

The L_1-restricted Portfolio

To avoid the degeneracy of the expected utility due to parameter uncertainty, we reduce the dimension of the portfolio by imposing an L_1-constraint on the investment strategies. For $c \geq 0$, the L_1-constrained problem is

$$\max_{\pi \in \mathcal{A}_c} V_\gamma(\pi|\mu, \sigma), \tag{3.0.1}$$

where $\mathcal{A}_c$ the set of admissible constrained strategies π, as defined in Subsection 2.2.1, such that

$$||\pi(t,w)||_1 = \sum_{i=1}^{d} |\pi_i(t,w)| \leq c \text{ for } m \otimes \mathbb{P} - a.e.\,(t,w),$$

and m is the Lebesgue measure on $[0,T]$.

3.1 Reduction to the Static Problem

Proposition 3.1.1. *Problem (3.0.1) reduces to the static problem*

$$\begin{cases} \max_{\pi \in \mathbb{R}^d} V_\gamma(\pi|\mu, \sigma) \\ \text{subject to } ||\pi||_1 \leq c. \end{cases} \tag{3.1.1}$$

In particular, the optimal strategy π_c^* is deterministic and constant.

We will use the following lemma[1] to prove Proposition 3.1.1.

[1] See also the dual approach of Cvitanic and Karatzas (1992) for power utility functions with $\gamma < 1$, and Karatzas and Shreve (1998) for $\gamma > 1$ and cone constraints.

Lemma 3.1.2. *Suppose $w(t,x)$ is a regular solution of the HJB equation*

(3.1.2)
$$w_t(t,x) + \sup_{\{\nu\in\mathbb{R}^d:||\nu||_1\leq c\}} \{\mathcal{L}^\nu w(t,x)\} = 0, \text{ with } w(T,x) = U_\gamma(x), x \geq 0.$$

where $\mathcal{L}^\nu$ is the differential operator given by

$$\mathcal{L}^\nu w(t,x) = \left(r + \nu^\top(\mu - r1)\right) x w_x + \frac{1}{2}\nu^\top \Sigma \nu x^2 w_{xx}.$$

Then, the stochastic integral

$$\int_0^t w_x(s, X^\pi(s)) X^\pi(s) \pi^\top(s) \sigma dW(s) \tag{3.1.3}$$

is a martingale for any admissible portfolio weight process $\pi \in \mathcal{A}_c$

Proof. We use the ansatz[2] that the solution of the HJB equation is of the form

$$w(t,x) = \phi(t) U_\gamma(x),$$

with

$$\frac{\phi'(t)}{1-\gamma} + \rho\phi(t) = 0, \phi(T) = 1,$$
$$\text{and } \rho = \sup_{\{\nu\in\mathbb{R}^d:||\nu||_1\leq c\}} \left\{r + \nu^\top(\mu - r1) - \frac{\gamma}{2}\nu^\top\Sigma\nu\right\}.$$

To obtain that the stochastic integral (3.1.3) is a true martingale for any admissible portfolio weight process $\pi \in \mathcal{A}_c$, we show that

$$\mathbb{E}\left(\int_0^T \left(w_x(s, X^\pi(s)) X^\pi(s)\right)^2 \pi^\top(s) \Sigma \pi(s) ds\right) < \infty.$$

Let $p = 1 - \gamma < 0$, then

$$\left(w_x(s, X^\pi(s)) X^\pi(s)\right)^2 = \phi^2(t) \left(X^\pi(s)\right)^{2p}.$$

[2]See (Pham, 2009, Subsection 3.6.1) for a related one-dimensional version of the problem.

Since ϕ is continuous on $[0,T]$ and the process $\pi \in \mathcal{A}_c$, we have

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T \left(w_x\left(s, X^\pi\left(s\right)\right)\right)^2 \left(X^\pi\left(s\right)\right)^2 \pi^\top\left(s\right) \Sigma \pi\left(s\right) ds\right) \\
&\leq c^2 ||\Sigma||_\infty \max_{t\in[0,T]} |\phi^2\left(t\right)| \mathbb{E}\left(\int_0^T \left(X^\pi\left(s\right)\right)^{2p} ds\right),
\end{aligned}$$

where $||\Sigma||_\infty = \max\left\{||\Sigma x||_\infty : x \in \mathbb{R}^d \text{ with } ||x||_\infty = 1\right\}$. Hence, as $\left(X^\pi\left(s\right)\right)^{2p}$ is positive, by Fubini's theorem, it remains to show that $\mathbb{E}\left(\left(X^\pi\left(s\right)\right)^{2p}\right)$ is bounded by a continuous function of time and the result will follow. Without loss of generality, we assume that $X(0) = 1$.

$$\begin{aligned}
\left(X^\pi\left(s\right)\right)^{2p} &= \exp\left(2p\left(\int_0^s \pi_u^\top\left(\mu - r1\right) - \frac{1}{2}\pi_u^\top \Sigma \pi_u du + \int_0^s \pi_u^\top \sigma dW_u\right)\right) \\
&= \exp\left(2p\left(\int_0^s \pi_u^\top\left(\mu - r1\right) + \frac{2p-1}{2}\pi_u^\top \Sigma \pi_u du\right)\right) \\
&\quad \cdot \mathcal{E}\left(2p\pi^\top \sigma\right)\left(s\right).
\end{aligned}$$

For $\pi \in \mathcal{A}_c$, the process $\left(\mathcal{E}\left(2p\pi^\top\sigma\right)\left(t\right)\right)_{(0\leq t\leq T)}$ is an exponential martingale, as it verifies the Novikov condition. Hence, we define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ as the Radon-Nykodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_T = \mathcal{E}\left(2p\pi^\top\sigma\right)\left(T\right).$$

As $\mathbb{Q}$ is equivalent to $\mathbb{P}$ and $\pi \in \mathcal{A}_c$, we have for $p < 0$,

$$\begin{aligned}
\mathbb{E}\left(\left(X^\pi\left(s\right)\right)^{2p}\right) &= \mathbb{E}^{\mathbb{Q}}\left(\exp\left(2p\left(\int_0^s \pi_u^\top\left(\mu - r1\right) + \frac{2p-1}{2}\pi_u^\top \Sigma \pi_u du\right)\right)\right) \\
&\leq \exp\left(2p\left(\left(p - \frac{1}{2}\right)c^2||\Sigma||_\infty - c||\mu - r1||_\infty\right)s\right)
\end{aligned}$$

The last term is continuous in the variable s and this concludes the proof. □

Proof of Proposition 3.1.1. Let $v\left(t,x\right)$ be the value function

$$v\left(t,x\right) = \sup_{\pi\in\mathcal{A}_c} \mathbb{E}[U\left(X^{\pi,t,x}\left(T\right)\right)].$$

We want to show that the value function of problem (3.0.1) is equal to the solution of the associated HJB equation $w\left(t,x\right)$. Since $w\left(t,x\right)$ is a regular solution, we can apply Itô's formula, we have the following decomposition.

For $\pi \in \mathcal{A}_c$,

$$\begin{aligned} U_\gamma\left(X^{\pi,t,x}(T)\right) &= w\left(T, X^{\pi,t,x}(T)\right) \\ &= w(t,x) + \int_t^T w_t\left(s, X^{\pi,t,x}(s)\right) + \mathcal{L}^{\pi(s)} w\left(s, X^{\pi,t,x}(s)\right) ds \\ &+ \int_t^T w_x\left(s, X^{\pi,t,x}(s)\right) X^{\pi,t,x}(s)\, \pi^\top(s)\, \sigma dW(s). \end{aligned}$$

Note that the first integral is negative as $w(t,x)$ is the solution of the HJB equation. By Lemma 3.1.2, we know that the stochastic integral is a true martingale. Hence

$$\mathbb{E}\left(U\left(X^{\pi,t,x}(T)\right)\right) \leq w(t,x) \Rightarrow v(t,x) \leq w(t,x) \quad \forall (t,x) \in [0,T] \times [0,\infty).$$

Furthermore, the optimal control of the HJB equation is given by

$$\pi_c^* = \arg\max_{||\nu||_1 \leq c} \left\{ r + \nu^\top (\mu - r1) - \frac{\gamma}{2} \nu^\top \Sigma \nu \right\}.$$

By definition of the value function, we have

$$w(t,x) = \mathbb{E}\left(U\left(X^{\pi_c^*,t,x}(T)\right)\right) \leq v(t,x) \quad \forall (t,x) \in [0,T] \times [0,\infty).$$

This implies $v = w$. Therefore the original optimisation problem (3.0.1) is equivalent to the HJB equation. As the optimal control is deterministic and constant in (3.1.2), the dynamic problem (3.0.1) reduces to the static problem (3.1.1). □

Remark 3.1.3. Note that the summation starts from $i = 1$, i.e., we only restrict the portfolio weights in the risky assets. The bound c of the L_1-constraint controls the level of sparsity in the portfolio.

Let $\tilde{\pi} = (\pi_0, \pi_1, \ldots, \pi_d)^\top$, with $\pi_0 = 1 - \sum_{i=1}^d \pi_i$. With $\pi_i^+ = \max\{\pi_i, 0\}$ and $\pi_i^- = -\min\{\pi_i, 0\}$ we denote by $\tilde{\pi}^{(l)}$ and $\tilde{\pi}^{(s)}$ the total percentages of long and short positions respectively, then

$$\begin{aligned} \tilde{\pi}^{(l)} &= \sum_{i=0}^d \pi_i^+ = \sum_{i=0}^d \left(|\pi_i| + \pi_i\right)/2 = \frac{1}{2}\left(\|\tilde{\pi}\|_1 + 1\right), \\ \tilde{\pi}^{(s)} &= \sum_{i=0}^d \pi_i^- = \sum_{i=0}^d \left(|\pi_i| - \pi_i\right)/2 = \frac{1}{2}\left(\|\tilde{\pi}\|_1 - 1\right). \end{aligned}$$

Then $\tilde{\pi}^{(l)} - \tilde{\pi}^{(s)} = 1$ and $\tilde{\pi}^{(l)} + \tilde{\pi}^{(s)} = \|\tilde{\pi}\|_1 = \|\pi\|_1 + |\pi_0|$. Now with $\sum_{i=1}^{d} |\pi_i| \leq c$, we obtain that

$$\begin{aligned}\tilde{\pi}^{(s)} &= \frac{1}{2}\left(\|\tilde{\pi}\|_1 - 1\right) = \frac{1}{2}\left(\|\pi\|_1 + |1 - \sum_{i=1}^{d} \pi_i| - 1\right) \\ &\leq \frac{1}{2}\left(c + |1 - \sum_{i=1}^{d} \pi_i| - 1\right) \leq \frac{1}{2}\left(c - 1 + 1 + c\right) = c.\end{aligned}$$

Hence, c is an upper bound on the total percentage of short positions held in the portfolio.

3.2 Structure of the Optimal Strategy

We study the effect of the L_1-constraint on the optimal strategy for given parameters μ and σ. This allows us to understand how assets are selected and to characterise the sparsity of the strategy as a function of γ. As the optimal strategy of the initial constrained problem is constant and deterministic, the constrained optimisation reduces to the mean-variance problem with the same L_1-constraint:

(3.2.1)
$$\begin{cases} \max\limits_{\pi \in \mathbb{R}^d} M_\gamma(\pi) \\ \text{subject to } ||\pi||_1 \leq c. \end{cases}$$

The mean-variance term can be rewritten as follows:

(3.2.2)
$$M_\gamma(\pi) = \pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi = -\frac{1}{2}||\sqrt{\gamma}\sigma^\top\pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)||_2^2 + K,$$

where $||\cdot||_2$ is the Euclidean norm and K does not depend on π. The L_1-constraint holds only on the weights of the risky assets. Therefore, we have a standard L_1-constrained ordinary least-square (OLS) problem; see Tibshirani (1996).

Lemma 3.2.1. *Problem (3.2.1) is equivalent to the constrained optimisation*

of the quadratic form:

$$\begin{cases} \min\limits_{\pi\in\mathbb{R}^d} ||\sqrt{\gamma}\sigma^\top\pi - \dfrac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)||_2^2 \\ \textit{subject to } ||\pi||_1 \leq c. \end{cases} \tag{3.2.3}$$

Note that (3.2.3) can be reduced to the case $\gamma = 1$ by considering the covariance matrix $\tilde{\Sigma} = \gamma\Sigma$. In that sense, we can interpret the optimisation problem for general RRA parameter $\gamma \geq 1$ as an optimisation problem in which a higher level of RRA is treated equivalently to larger entries in the covariance matrix for an investor with $\gamma = 1$.

To highlight the fundamental role of the coefficient of RRA γ on the sparsity of the constrained strategy, we provide an analytical solution of (3.2.3) for diagonal volatility matrices.

Theorem 3.2.2. *Suppose* $\Sigma = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_d^2\right)$ *and* $\mu \in \mathbb{R}^d$ *with*

$$|\mu_1 - r| > |\mu_2 - r| > \ldots > |\mu_d - r|.$$

Then $\pi_i^* = \frac{1}{\gamma\sigma_i^2}(\mu_i - r1)$, $i = 1, \ldots, d$, *is a solution to (3.2.1) if* $||\pi^*||_1 \leq c$. *Otherwise, the unique solution is*

$$\pi_c^* = \left(\frac{\operatorname{sgn}(\mu_1 - r)}{\gamma\sigma_1^2}(|\mu_1 - r| - a), \ldots, \frac{\operatorname{sgn}(\mu_k - r)}{\gamma\sigma_k^2}(|\mu_k - r| - a), 0, \ldots, 0\right)^\top, \tag{3.2.4}$$

where sgn *is the sign function and*

$$a = \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}\left(\sum_{i=1}^k \frac{|\mu_i - r|}{\sigma_i^2} - \gamma c\right), \tag{3.2.5}$$

$$k = \min\left\{j = 1, \ldots, d : c \leq \sum_{i=1}^j \frac{1}{\gamma\sigma_i^2}(|\mu_i - r| - |\mu_{j+1} - r|)\right\}, \tag{3.2.6}$$

and $\mu_{d+1} = r$.

Proof of Theorem 3.2.2. The proof of this result is divided in two parts. First, we identify the shape of the optimal constrained weights. Second, we characterise the shrinkage intensity.

Since the constraint is binding, it is equivalent to the minimisation of the Lagrangian:

$$\min_{\pi} \frac{1}{2}||\sqrt{\gamma}\sigma^{\top}\pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)||_2^2 + a||\pi||_1, a \geq 0.$$

As the matrix σ is diagonal, we can optimise term by term. For each $i = 1, \ldots, d$,

$$\min_{\pi_i} \frac{1}{2}\left(\sqrt{\gamma}\sigma_i\pi_i - \frac{\mu_i - r}{\sqrt{\gamma\sigma_i}}\right)^2 + a|\pi_i| = \min_{\pi_i} \frac{\gamma}{2}\sigma_i^2\left(\pi_i - \frac{\mu_i - r}{\gamma\sigma_i^2}\right)^2 + a|\pi_i|.$$

Therefore, for each $i = 1, \ldots, d$, the optimal solution $(\pi_c^*)_i$ is the proximal mapping of the previous minimisation problem. For the absolute value function $|\cdot|$, the proximal mapping corresponds to the soft-thresholding operator[3]. It is given by computing the stationary point of the objective function for $\pi_i > 0$ and $\pi_i < 0$ and we get

$$(\pi_c^*)_i = \frac{\operatorname{sgn}(\mu_i - r)}{\gamma\sigma_i^2}(|\mu_i - r| - a)^+, \text{ for each } i = 1, \ldots, d.$$

We can now compute the parameter a. The argument follows from a discussion of (Osborne et al., 2000, Section 5.2) on L_1-constrained regression with an orthogonal matrix. We order the absolute excess returns by decreasing order

$$|\mu_1 - r| \geq |\mu_2 - r| \geq \ldots \geq |\mu_d - r|.$$

Let $k = \min_{i=1,\ldots,d-1}\{|\mu_{i+1} - r| \leq a\}$, we have that

$$\begin{aligned}
||\pi^*||_1 - c &= \sum_{i=1}^{d} \frac{|\mu_i - r|}{\gamma\sigma_i^2} - \sum_{i=1}^{d} \frac{(|\mu_i - r| - a)^+}{\gamma\sigma_i^2} \\
&= \sum_{i=1}^{d} \frac{|\mu_i - r|}{\gamma\sigma_i^2} I(|\mu_i - r| \leq a) + \sum_{i=1}^{d} \frac{a}{\gamma\sigma_i^2} I(|\mu_i - r| > a) \\
&= \sum_{i=k+1}^{d} \frac{|\mu_i - r|}{\gamma\sigma_i^2} + \sum_{i=1}^{k} \frac{a}{\gamma\sigma_i^2}.
\end{aligned}$$

[3]The notion of proximal mapping is due to Moreau (1965). The terminology of "soft" threshold was first introduced by Donoho and Johnstone (1994).

Using the expression of $||\pi^*||_1$ again, we obtain a. □

Remark 3.2.3. Note that if the volatility is a multiple of the identity matrix, the result follows directly from (Gandy and Veraart, 2013, Theorem 5.2) by considering $\tilde{\sigma} = \sqrt{\gamma}\sigma$ rather than σ in the optimisation problem, which is solved for a logarithmic utility function. Hence, the novelty of this result relies in the more general structure of the volatility matrix, and not on the type of utility function considered.

The argument to find the structure of the constrained weights is as follows. As the optimal unconstrained strategy π^* does not satisfy the L_1-constraint in general, the weights have to be shrunk. For a diagonal volatility matrix, the constrained solution is of the form,

$$(\pi_c^*)_i = \frac{\operatorname{sgn}(\mu_i - r)}{\gamma\sigma_i^2} \left(|\mu_i - r| - a\right)^+, \text{ for each } i = 1, \ldots, d, \tag{3.2.7}$$

where $a > 0$ and $(\cdot)^+$ denotes the positive part[4]. Because of the structure of (3.2.7), we invest only in assets with the highest absolute excess returns and we classify them by decreasing order of absolute excess return. The excess returns are adjusted by the shrinking constant a. The shrinkage intensity has to be decreasing in c, to reflect the fact that, the strategy is less constrained for a large c.

To satisfy the constraint, we have to deviate from π^* but the shrinkage intensity should not change the sign of the position, long or short[5]. Therefore, the weight of an asset is set to zero if its absolute excess return is smaller than a. The index k is then defined as the last asset with an absolute excess return larger than a,

$$k = \min_{j=1,\ldots,d-1} \left\{|\mu_{j+1} - r| \leq a\right\}. \tag{3.2.8}$$

Since the L_1-constraint is binding, we obtain the expression (3.2.5) of the positive shrinkage intensity a as a function of k.

[4]This structure was first characterised in Tibshirani (1996) for orthogonal matrices. See also the corresponding result in Gandy and Veraart (2013), for a multiple of the identity matrix.

[5]For a diagonal matrix, the constrained and the unconstrained solutions will have the same sign component wise. However, when there is enough correlation, the signs can be different; see Tibshirani (1996, Section 2.3).

We can also write k in terms of c and γ. By (3.2.5) and (3.2.8), k is the smallest index such that the following inequality holds,

$$\sum_{i=1}^{k} \frac{|\mu_{k+1} - r|}{\gamma \sigma_i^2} \leq \sum_{i=1}^{k} \frac{|\mu_i - r|}{\gamma \sigma_i^2} - c. \tag{3.2.9}$$

The left-hand side of inequality (3.2.9) corresponds to the L_1-norm of the weights of k risky independent assets with a common drift μ_{k+1} and volatility matrix σ. Therefore, the strategy includes assets with largest excess returns until the difference between the L_1-norm of the corresponding unconstrained weights and the bound c exceeds the L_1-norm of the weights of these k fictitious assets. Rewriting inequality (3.2.9), we get the full characterisation of k in (3.2.6) and the constrained strategy in (3.2.4). The index k increases with γ and the larger is γ, the less sparse is the constrained portfolio.

Finally, we establish some regularity properties of the optimal constrained strategy and the corresponding expected utility as functions of the bound c.

Proposition 3.2.4. *Let π_c^* be the optimal solution of problem (3.1.1).*

1. *The solution map $c \mapsto \pi_c^*$ is continuous on $\mathbb{R}^+$,*
2. *The expected utility map $c \mapsto V_\gamma(\pi_c^* | \mu, \sigma)$ is continuous and concave on $\mathbb{R}^+$.*

Proof of Proposition 3.2.4. The domain of both maps is $\mathbb{R}^+$. Indeed, c has to be non-negative for the constrained problem to be well-defined.

1. From Lemma 3.2.1 it is sufficient to consider problem (3.2.3). By changing the signs of the objective function and (3.2.2), problem (3.2.3) is equivalent to the following minimisation problem:

$$\begin{cases} \min\limits_{\pi} \dfrac{1}{2} \pi^\top D \pi + \pi^\top b \\ \text{subject to } ||\pi_c||_1 \leq c, \end{cases}$$

   where, in our case, $D = \gamma \Sigma$ and $b = -(\mu - r1)$.

   Next, we represent the L_1-constraint as a set of linear inequality constraints. The L_1-constraint is equivalent to the 2^d constraints of the type:

$$(\pm 1, .., \pm 1)\, \pi \leq c.$$

Let $\mathbf{1} = (1, \ldots, 1)^\top$ the 2^d-dimensional vector of ones. In matrix form, we have

$$A\pi \leq c\mathbf{1},$$

where $A = (A_{ij})$ with $A_{ij} \in \{1, -1\}$, $i = 1, \ldots, 2^d$, $j = 1, \ldots, d$, and each possible combination of signs appears only once as a row of A. Therefore, the constrained OLS problem is equivalent to standard quadratic programming problem:

$$\begin{cases} \min\limits_{\pi} \dfrac{1}{2}\pi^\top D\pi + \pi^\top b, \\ \text{subject to } A\pi \leq c\mathbf{1}. \end{cases}$$

Since the matrix D is positive definite, the solution map $c \mapsto \pi_c^*$ is continuous, by Lee et al. (2006, Corollary 3.1).

2. Let $v(x) = K_\gamma \exp((1-\gamma)Tx)$, $x \in \mathbb{R}$. The function v and the mean-variance term function M_γ are both continuous on $\mathbb{R}$ and $\mathbb{R}^d$, respectively. Since the map $c \mapsto \pi_c^*$ is continuous, the map $c \mapsto V_\gamma(\pi_c^*|\mu, \sigma)$ is continuous as a composition of continuous functions:

$$c \mapsto \pi_c^* \mapsto M_\gamma(\pi_c^*) \mapsto v(M_\gamma(\pi_c^*)) = V_\gamma(\pi_c^*|\mu, \sigma).$$

Note that since π_c^* is deterministic the last equality holds.

To show the concavity of the expected utility as a function of c, we define the functions

$$f(\pi) = ||\sqrt{\gamma}\sigma^\top \pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)||_2, \text{ and } G(\pi) = ||\pi||_1.$$

Both functions are convex on $\mathbb{R}^d$. This implies by Luenberger (1969, Proposition 1, Section 8.3) that the function $w(c) = f(\pi_c^*)$ is convex in c. Moreover, by (3.2.2), $M_\gamma(\pi_c^*) = -\frac{1}{2}w(c) + K$ and the map $c \mapsto M_\gamma(\pi_c^*)$ is concave in c. Finally, as $\gamma > 1$ and $K_\gamma < 0$, the function v is concave and increasing. Hence, the expected utility map $c \mapsto V_\gamma(\pi_c^*|\mu, \sigma)$ is concave in c as a composition of a concave and increasing function with a concave function, $V_\gamma(\pi_c^*|\mu, \sigma) = v(M_\gamma(\pi_c^*))$.

□

This result holds for any volatility matrix, and it shows that we can adjust continuously the sparsity of our strategy while keeping the continuity and the concavity of the expected utility as a function of c. In particular, as we relax the constraint, the expected utility of the constrained problem converges to the expected utility of the Merton ratio; see Corollary B.0.4.

3.3 The Constrained Plug-in Strategy: Sparsity and Estimation

We first characterise the expected utility for any time-constant L_1-constrained strategy π_c independent of $W(T)$. Similarly to Lemma 2.3.1, the expected utility is given by the moment generating function of the mean-variance term $M_\gamma(\pi_c)$.

Lemma 3.3.1. *Assume that π_c is a time-constant and possibly random strategy independent of $W(T)$ such that $||\pi_c||_1 \leq c$. Then,*

$$V_\gamma(\pi_c|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)TM_\gamma(\pi_c)\right)\right]. \tag{3.3.1}$$

The lemma is proved in Appendix B.

For the unconstrained estimated strategy, the expected utility can degenerate to $-\infty$ for $\gamma \geq 1$, as the number of available risky assets increases. The first advantage of using an L_1-norm is to rule out the degeneracy of the expected utility as the number of assets grows to infinity.

Proposition 3.3.2. *Assume that $\pi_c^{(d)}$ is a time-constant and possibly random strategy independent of $W^{(d)}(T)$ such that $||\pi_c^{(d)}||_1 \leq c$ for each d. Suppose*

$$\lim_{d\to\infty} \max_{i=1,..,d} |\mu_i^{(d)} - r| < \infty \text{ and } \lim_{d\to\infty} \max_{i,j=1,..,d} \Sigma_{ij}^{(d)} < \infty,$$

then $V_\gamma\left(\pi_c^{(d)}|\mu^{(d)},\sigma^{(d)}\right)$ is bounded from below for $\gamma \geq 1$ as $d \to \infty$.

The statement is proved in Appendix B.

Note that this statement is true for any type of constant L_1-constrained strategy. Hence, this does include plug-in strategies in which drift and possibly even covariances are replaced by estimators.

We see that the L_1-constraint is especially helpful when we face parameter uncertainty for a large number of assets. Indeed, the loss due to estimation accumulates and it is so large that we can potentially gain from holding a sparse portfolio. The key to the performance of the L_1-constraint is the trade-off between the gain due to diversification and the loss due to estimation error.

To understand this in terms of loss functions, let π^* be the Merton strategy and π_c^* the solution of the constrained problem with the true drift μ and bound c. When μ is known, the sparsity of the constrained strategy, π_c^*, implies a loss in expected utility. We define the loss factor $L_\gamma(\pi_c^*, \pi^*)$ by the relationship

$$V_\gamma(\pi_c^*|\mu, \sigma) = L_\gamma(\pi_c^*, \pi^*) V_\gamma(\pi^*|\mu, \sigma).$$

At the beginning of the investment period, the value of the MLE of the drift $\hat{\mu}$ is fixed and we obtain the constrained plug-in strategy $\hat{\pi}_c$ by solving (3.2.3) with $\hat{\mu}$ and the bound c. The resulting strategy $\hat{\pi}_c$ is not normally distributed and it is a biased estimator of π_c^*. Thus, there is loss due to estimation on the constrained strategy. The corresponding loss factor $L_\gamma(\hat{\pi}_c, \pi_c^*)$ is defined by

$$V_\gamma(\hat{\pi}_c|\mu, \sigma) = L_\gamma(\hat{\pi}_c, \pi_c^*) V_\gamma(\pi_c^*|\mu, \sigma).$$

The total loss consists of both the loss due to insufficient diversification and the loss due to estimation error, i.e., the total loss factor $L_\gamma(\hat{\pi}_c, \pi^*)$ is defined by

$$L_\gamma(\hat{\pi}_c, \pi^*) = L_\gamma(\pi_c^*, \pi^*) L_\gamma(\hat{\pi}_c, \pi_c^*)$$

and then we obtain

$$V_\gamma(\hat{\pi}_c|\mu, \sigma) = L_\gamma(\hat{\pi}_c, \pi^*) V_\gamma(\pi^*|\mu, \sigma).$$

By Proposition 3.2.4, the loss factor due to under-diversification is continuous in c. The loss factor due to estimation and the total loss factor are also continuous functions in the bound c as the following proposition shows.

Proposition 3.3.3. *Let t_{obs} be large enough such that*

$$t_{obs}\gamma\Sigma + (1-\gamma)\,Td||\Sigma||_\infty I_d$$

is positive definite. Then, the maps

$$c \mapsto L_\gamma\left(\hat{\pi}_c, \pi_c^*\right) \text{ and } c \mapsto L_\gamma\left(\hat{\pi}_c, \pi^*\right)$$

are continuous on $\mathbb{R}^+$.

Proof Proposition 3.3.3. By (2.3.1), the maximum likelihood of the drift, $\hat{\mu}$ is a $\mathbb{R}^d$-valued random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. To compute the plug-in constrained strategy, we solve problem (3.2.3) with $\hat{\mu}$. In Proposition 3.2.4, we showed that for a fixed value of the drift the solution map is continuous. Hence, the solution map of the plug-in strategy $(\omega, c) \mapsto \hat{\pi}_c$ is $\mathcal{F}$-measurable and continuous in c a.s..

As in Proposition 3.2.4, let $v(x) = K_\gamma \exp\left((1-\gamma)\,Tx\right)$, with $K_\gamma < 0$. Our objective is to show that $|v\left(M_\gamma\left(\hat{\pi}_c\right)\right)| \in L^1\left(\Omega, \mathcal{F}, \mathbb{P}\right)$. Indeed, since the map $c \mapsto v\left(M_\gamma\left(\hat{\pi}_c\right)\right)$ is continuous, we can then establish the continuity of the map $c \mapsto \mathbb{E}\left(v\left(M_\gamma\left(\hat{\pi}_c\right)\right)\right) = V_\gamma\left(\hat{\pi}_c|\mu, \sigma\right)$ by the dominated convergence theorem.

We start by computing a lower bound for the mean-variance term $M_\gamma\left(\hat{\pi}_c\right)$ (inequalities hold a.s.):

$$M_\gamma\left(\hat{\pi}_c\right) = \hat{\pi}_c^\top\left(\mu - r1\right) - \frac{\gamma}{2}\hat{\pi}_c^\top\Sigma\hat{\pi}_c \geq -||\hat{\pi}_c||_1||\mu - r1||_\infty - \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}_c||_1^2.$$

Since $||\hat{\pi}_c||_1 \leq ||\hat{\pi}||_1$, we deduce that

$$M_\gamma\left(\hat{\pi}_c\right) \geq -||\hat{\pi}||_1||\mu - r1||_\infty - \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}||_1^2,$$

and we define

$$Y = -||\hat{\pi}||_1||\mu - r1||_\infty - \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}||_1^2. \tag{3.3.2}$$

Because the function v is increasing and negative, $v(Y) \leq v\left(M_\gamma\left(\hat{\pi}_c\right)\right)$ and $|v\left(M_\gamma\left(\hat{\pi}_c\right)\right)| \leq |v(Y)|$.

To apply the dominated convergence theorem, we need to show that

$|v(Y)| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Note that

$$\mathbb{E}(|v(Y)|) = -K_\gamma \mathbb{E}\left[\exp\left((\gamma-1)T\left(||\hat{\pi}||_1||\mu - r1||_\infty + \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}||_1^2\right)\right)\right]$$

and by Cauchy-Schwarz inequality in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ we obtain

$$(\mathbb{E}[|v(Y)|])^2 \leq K_\gamma^2 \mathbb{E}\left[\exp\left(2(\gamma-1)T||\hat{\pi}||_1||\mu - r1||_\infty\right)\right] \tag{3.3.3}$$

$$\cdot\, \mathbb{E}\left[\exp\left((\gamma-1)\gamma T||\Sigma||_\infty||\hat{\pi}||_1^2\right)\right]. \tag{3.3.4}$$

Therefore, we have to show that the two expectations on the right hand side of the inequality are finite. Recall that $\hat{\pi} \sim \mathcal{N}_d\left(\pi^*, V_0^2\right)$. We set $V_0^2 = \tilde{V}\tilde{V}^\top$ with $\tilde{V} \in \mathbb{R}^{d\times d}$. Let $z \sim \mathcal{N}(0,1)$, $Z \sim \mathcal{N}_d(0, I_d)$,

$$a = 2(\gamma-1)T||\mu - r1||_\infty||\tilde{V}||_\infty \text{ and } C = \exp\left(2(\gamma-1)T||\mu - r1||_\infty||\pi^*||_1\right).$$

For $\gamma > 1$ the first expectation in (3.3.3) can be bounded as follows,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(2(\gamma-1)T||\hat{\pi}||_1||\mu - r1||_\infty\right)\right] \\
&= \mathbb{E}\left[\exp\left(2(\gamma-1)T||\tilde{V}Z + \pi^*||_1||\mu - r1||_\infty\right)\right], \\
&\leq C\mathbb{E}\left[\exp\left(2(\gamma-1)T||\mu - r1||_\infty||\tilde{V}||_\infty||Z||_1\right)\right], \\
&= C\left(\mathbb{E}\left[\exp\left(2(\gamma-1)T||\mu - r1||_\infty||\tilde{V}||_\infty|z|\right)\right]\right)^d, \\
&= C\left(2\exp\left(\frac{a^2}{2}\right)\Phi_z(a)\right)^d, \text{ where } \Phi_z \text{ the CDF of } z, \\
&< \infty.
\end{aligned}$$

Since $||\hat{\pi}||_1^2 \leq d||\hat{\pi}||_2^2$, the second expectation in (3.3.3) can be bounded as follows:

$$\mathbb{E}\left[\exp\left((\gamma-1)\gamma T||\Sigma||_\infty||\hat{\pi}||_1^2\right)\right] \leq \mathbb{E}\left[\exp\left((\gamma-1)\gamma Td||\Sigma||_\infty||\hat{\pi}||_2^2\right)\right],$$

which is finite by Lemma A.0.3 because $\left(V_0^2\right)^{-1} - (\gamma-1)\gamma Td||\Sigma||_\infty I_d$ is positive definite as $\left(V_0^2\right)^{-1} = \gamma^2 t_{obs}\Sigma$. This condition is equivalent to $t_{obs}\gamma\Sigma + (1-\gamma)Td||\Sigma||_\infty||I_d$ being positive definite. □

The proof consists in showing that the map $c \mapsto V_\gamma(\hat{\pi}_c|\mu, \sigma)$ is continuous. Because the term $(1-\gamma)Td||\Sigma||_\infty$ is negative and grows linearly

with d, Proposition 3.3.3 can be applied only if t_{obs} is roughly larger than d. If the L_1-constraint effectively shrinks all the weights towards zero, i.e., $|(\hat{\pi}_c)_i| \leq |\hat{\pi}_i|$, for $i = 1, \ldots, d$, a.s., we can get rid of the dependency on d. In this case, the continuity holds if $t_{obs}\gamma\Sigma + (1-\gamma)\, T||\Sigma||_2 I_d$ is positive definite. This property is in particular true for a diagonal covariance matrix.

Remark 3.3.4. If $||\hat{\pi}_c||_2 \leq ||\hat{\pi}||_2$ *a.s.*, then

$$\hat{\pi}_c^\top \Sigma \hat{\pi}_c \leq \lambda_{max}(\Sigma)\, ||\hat{\pi}_c||_2^2 \leq \lambda_{max}(\Sigma)\, ||\hat{\pi}||_2^2 \quad a.s.,$$

with $\lambda_{max}(\Sigma) = \max\left\{||\Sigma x||_2 : x \in \mathbb{R}^d \text{ with } ||x||_2 = 1\right\}$. Therefore, we can set the lower bound in (3.3.2) as

$$Y = -||\hat{\pi}||_1 ||\mu - r1||_\infty - \frac{\gamma}{2}\lambda_{max}(\Sigma)\, ||\hat{\pi}||_2^2.$$

Then, $|v(Y)| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ if $t_{obs}\gamma\Sigma + (1-\gamma)\, T\lambda_{max}(\Sigma)\, I_d$ is positive definite.

As we relax the constraint, the strategy π_c^* is more diversified and the term $L_\gamma(\pi_c^*, \pi^*)$ converges to one, i.e.,

$$L_\gamma(\pi_c^*, \pi^*) \to 1 \text{ as } c \to \infty. \tag{3.3.5}$$

The loss factor due to estimation error, $L_\gamma(\hat{\pi}_c, \pi_c^*)$, behaves oppositely. As c increases, more stocks are included in the strategy and the estimation error forces $L_\gamma(\pi_c^*, \hat{\pi}_c)$ to move away from one. By Proposition 3.3.3, the loss factor $L_\gamma(\hat{\pi}_c, \pi_c^*)$ and the total loss factor $L_\gamma(\hat{\pi}_c, \pi^*)$ both converge to the loss factor of the unconstrained plug-in strategy[6], i.e.,

$$L_\gamma(\hat{\pi}_c, \pi_c^*) \to L_\gamma(\hat{\pi}, \pi^*) \text{ and } L_\gamma(\hat{\pi}_c, \pi^*) \to L_\gamma(\hat{\pi}, \pi^*) \text{ as } c \to \infty. \tag{3.3.6}$$

For t_{obs} and c finite, the loss factors are bigger than one and the aim is to find a bound such that the total loss factor is closest to one. The existence of an optimal bound depends on the structure of the true parameters and the accuracy of the estimator $\hat{\mu}$. We study the behaviour of the loss factors in more detail in the next section.

Remark 3.3.5. As in the unconstrained case, the measure of efficiency, introduced in Definition 2.3.5, is a function of the loss factor of the constrained

[6]See Corollary B.0.4 in the Appendix.

plug-in strategy. Since the expected utility is given by (3.3.1), the following relationship holds true for $\gamma > 1$:

$$\Theta_\gamma\left(\hat{\pi}_c\right) = L_\gamma\left(\hat{\pi}_c, \pi^*\right)^{\frac{1}{1-\gamma}}.$$

3.4 Simulation Study

In this section, we investigate the structure and the performance of the L_1-constrained portfolio when risky assets are correlated. We consider a volatility matrix that is non-diagonal. For this situation we do not have an analytic form for the constrained strategy and therefore we use simulations to compute its expected utility. Our data set consists of a random sample of $d = 250$ stocks that have been listed at least once on the S&P 500 and had daily returns for all trading days between January 2001 and December 2011. There is no problem of survivorship bias because our main goal is to show the existence and the uniqueness of an optimal bound c for a given universe of stocks. Throughout this section, we assume an initial normalised endowment of $X(0) = 1$, an annual risk-free rate of $r = 0.02$ and an investment horizon of $T = 1$ year.

3.4.1 Methodology

Based on the daily log-returns of the stocks, we compute the following unbiased estimators $\tilde{\mu}$, $\tilde{\Sigma}$ of μ and Σ respectively:

$$\tilde{\mu} = \frac{1}{\Delta}\tilde{\xi} + \frac{1}{2}\operatorname{diag}\left(\tilde{\Sigma}\right), \tag{3.4.1}$$

$$\tilde{\Sigma}_{ij} = \frac{1}{\Delta\left(N-1\right)}\sum_{k=0}^{N-1}\left[R_i\left(k\right) - \tilde{\xi}_i\right]\left[R_j\left(k\right) - \tilde{\xi}_j\right], \text{ for } i,j = 1,\ldots,d, \tag{3.4.2}$$

with $\tilde{\xi}_i = \frac{1}{N}\sum_{k=0}^{N-1} R_i\left(k\right)$ and $R_i\left(k\right) = \log\left(\frac{S_i((k+1)\Delta)}{S_i(k\Delta)}\right)$. The time step is defined by $\Delta = t_{obs}/N$, with $t_{obs} = 11$ years and $N = 2767$ days.

We then assume a standard Merton market as introduced in Section 2.2 where we set $\mu = \tilde{\mu}$ for the drift and $\Sigma = \tilde{\Sigma}$ for the covariance matrix. Furthermore, we evaluate the estimator of the drift $\hat{\mu}$ by sampling from its

law, i.e., we use the fact that $\hat{\mu} \sim \mathcal{N}_d(\mu, \Sigma/t_{obs})$. In this setting, we compute the investment weights and the associated expected utility for :

- π^* the unconstrained strategy using the true μ,
- $\hat{\pi}$ the unconstrained plug-in strategy using the estimator $\hat{\mu}$,
- π_c^* the L_1-constrained strategy using the true μ,
- and $\hat{\pi}_c^*$ the L_1-constrained plug-in strategy using the estimator $\hat{\mu}$.

For a non-diagonal volatility matrix, we do not have analytical results for the L_1-constrained strategy. Therefore, we compute the optimal weights numerically. To do so, we plug the true or the estimated drift in (3.2.1) and we solve the quadratic optimisation problem with the LARS algorithm; see Efron et al. (2004). This algorithm generates the optimal portfolio weights for all binding bounds.

3.4.2 Computation of the Loss Function

We want to compute the expected utility of the unconstrained and the constrained strategies both when the drift is known and when it is estimated. For the unconstrained case, we know the explicit form of the expected utility associated with the Merton ratio (2.2.3) and the plug-in strategy (2.3.2). For the constrained case with known drift, the strategy is deterministic and the associated expected utility can be computed directly by using (3.3.1).

When the drift is estimated, the constrained strategy $\hat{\pi}_c$ is random, and we approximate its expected utility using a Monte-Carlo method. More specifically, we sample M i.i.d. realisations of the MLE of the drift. Then, for each realisation, we solve for the associated constrained strategy and we obtain i.i.d. realisations $Y_{\gamma,c}^1, \ldots, Y_{\gamma,c}^M$ of $Y_{\gamma,c}$ given by

$$Y_{\gamma,c} = \exp\left((1-\gamma)\, T\left(\hat{\pi}_c^\top (\mu - r1) - \frac{\gamma}{2}\hat{\pi}_c^\top \Sigma \hat{\pi}_c\right)\right).$$

For K_γ given in (2.2.4), we define the Monte-Carlo estimator of the expected utility by

$$\bar{V}_{\gamma,M}(\hat{\pi}_c|\mu,\sigma) = K_\gamma \frac{1}{M}\sum_{i=1}^{M} Y_{\gamma,c}^i,$$

which approximates $V_\gamma(\hat{\pi}_c|\mu,\sigma) = K_\gamma \mathbb{E}(Y_{\gamma,c})$; see (3.3.1). The accuracy of the method is measured by the standard deviation of the Monte-Carlo estimator given by

$$\sqrt{\mathbf{Var}\left(\bar{V}_{\gamma,M}(\hat{\pi}_c|\mu,\sigma)\right)}.$$

For a fixed c, the standard deviation of the random variable $\bar{V}_{\gamma,M}(\hat{\pi}_c|\mu,\sigma)$ is a function of γ. Therefore, the number of realisations necessary to attain a given accuracy varies with γ. In this section, the number of realisations, M, is fixed to 5000 for all $\gamma \in [1,7]$[7].

Furthermore, the loss in expected utility due to estimation $L_\gamma(\hat{\pi}_c, \pi_c^*)$ is approximated by the ratio of the Monte-Carlo estimator $\bar{V}_{\gamma,M}(\hat{\pi}_c|\mu,\sigma)$ and the expected utility of the constrained strategy with known parameters $V_\gamma(\pi_c^*|\mu,\sigma)$. To remove the influence of the multiplicative factor $V_\gamma(\pi_c^*|\mu,\sigma)$ on the accuracy of the estimation of the loss factors, we apply the logarithmic transformations:

$$\ell_\gamma(\pi_c^*, \pi^*) = \log\left(L_\gamma(\pi_c^*, \pi^*)\right), \quad \ell_\gamma(\hat{\pi}_c, \pi_c^*) = \log\left(L_\gamma(\hat{\pi}_c, \pi_c^*)\right).$$

As mentioned in Section 3.3, the total loss in expected utility is the product of the loss due to under-diversfication measured in terms of the factor $L_\gamma(\pi_c^*, \pi^*)$, and the loss due to estimation, measured in terms of the factor $L_\gamma(\hat{\pi}_c, \pi_c^*)$. The logarithm of the total loss factor is then given by

$$\ell_\gamma(\hat{\pi}_c, \pi^*) = \ell_\gamma(\pi_c^*, \pi^*) + \ell_\gamma(\hat{\pi}_c, \pi_c^*).$$

Computing the logarithm enables us to have a natural interpretation of the loss factors. Indeed, in this additive setting, the log-loss equals to zero when there is no loss in expected utility. By (3.3.5) and (3.3.6), the log-loss factors converge to zero and to the log-loss factor of the unconstrained plug-in strategy, respectively,

$$\ell_\gamma(\pi_c^*, \pi^*) \to 0, \ \ell_\gamma(\hat{\pi}_c, \pi_c^*) \to \ell_\gamma(\hat{\pi}, \pi^*), \text{ and } \ell_\gamma(\hat{\pi}_c, \pi^*), \to \ell_\gamma(\hat{\pi}, \pi^*).$$

as $c \to \infty$. Finally, as the log-convexity implies the convexity of the loss function itself, it is sufficient to study the convexity of the log-loss.

[7] See discussion in Appendix C.

3.4.3 Existence of an Optimal Bound

Figure 3.1 depicts the profile of the log-loss factor as a function of c for $\gamma = 5$. We see that the log-loss factors are continuous in c. Furthermore, the total log-loss, $\ell_\gamma(\hat{\pi}_c, \pi^*)$, is convex in c and it is minimised at $c^* = 12$. Hence, the total loss factor $L_\gamma(\hat{\pi}_c, \pi^*)$ is also convex and minimised at c^*. Equivalently, the expected utility of the constrained plug-in strategy is maximised at this optimal bound c^*.

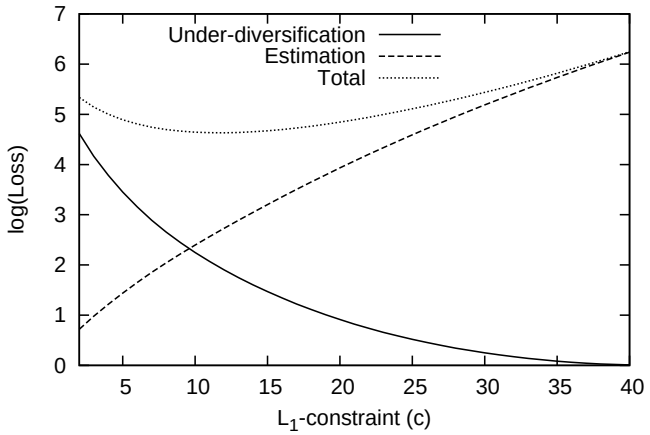


Figure 3.1: Plot of the log-loss factor due to under-diversification $\ell_\gamma(\pi_c^*, \pi^*)$, due to estimation $\ell_\gamma(\hat{\pi}_c, \pi_c^*)$ and the total loss factor $\ell_\gamma(\hat{\pi}_c, \pi^*)$ as a function of the bound of the L_1-constraint c with $\gamma = 5$.

RRA	Optimal	L_1-constrained		Unconstrained	
γ	c	$\ell_\gamma(\hat{\pi}_c, \pi^*)$	$\Theta_\gamma(\hat{\pi}_c)$	$\ell_\gamma(\hat{\pi}, \pi^*)$	$\Theta_\gamma(\hat{\pi})$
2	29	2.88	0.06	5.82	0.003
3	20	3.86	0.15	7.82	0.02
5	12	4.63	0.31	9.44	0.10
7	8	4.97	0.44	10.14	0.19

Table 3.1: Comparison of the log-loss in expected utility and the efficiency between the constrained and the unconstrained case, for different values of risk aversion γ. The log-loss factor is equal to zero when there is no loss.

Table 3.1 shows that the optimal bound decreases sharply with γ. Again, as γ increases, we need to use a more restrictive bound to optimally control the loss factor. For each γ, the loss due to underdiversification represents between 41% and 44% of the total loss factor, while loss due to estimation

represents between 56% and 59%. Moreover, the L_1-constraint greatly helps the log-loss to be closer to zero. The logarithm of the loss is reduced by roughly 50% at the optimal bound and the loss factor itself is reduced by at least 95.7%.

In terms of efficiency, the L_1-constraint also helps significantly. For small levels of risk aversion, e.g., $\gamma = 2, 3$, the efficiency is improved dramatically by imposing an L_1-constraint. For high level of risk aversion, e.g., $\gamma = 7$, the effect of estimation is less important, as a smaller fraction is invested in the risky assets. In this case, holding a L_1-sparse portfolio doubles the efficiency of the investor.

3.4.4 Structure and Stability of the L_1-constrained Strategy

Regarding the sparsity of the constrained strategy, the results of Section 3.2 extend to our general covariance matrix. Figure 3.2 shows that the number of stocks invested in increases with γ. For instance, at $c = 7$, we invest in 44 stocks with $\gamma = 2$ and in 93 stocks with $\gamma = 6$.

On the one hand, if γ is large, we are less constrained as a smaller fraction is invested in each selected stock, and this enables us to hold a more diversified portfolio. On the other hand, if γ is small, we still take some relatively strong positions at the cost of having a very sparse portfolio.

As we relax the constraint, we notice that the number of stocks invested in increases stepwise. The steps are especially long for small γ as we take large positions. In effect, c needs to be increased significantly until new stocks are added to the portfolio.

For independent stocks, we should invest in stocks with the highest absolute excess returns up to a certain index k, which is increasing in γ (Theorem 3.2.2). The weights of the selected stocks are shrunk towards zero, while the other weights are set to zero. If stocks with highest absolute returns have positive excess returns, the portfolio consists exclusively of long positions and the L_1-constraint acts only as a restriction of the number of stocks.

Table 3.2 reports the composition on the L_1-constrained portfolio for correlated stocks, when the drift μ is known. Because of the structure of the covariance matrix, stocks with negative returns are selected although their absolute excess return is close to zero. Indeed, the L_1-constraint selects the stocks with highest returns and then jumps to stocks with negative excess returns. In contrast to the no-short sale constraint, we still keep a

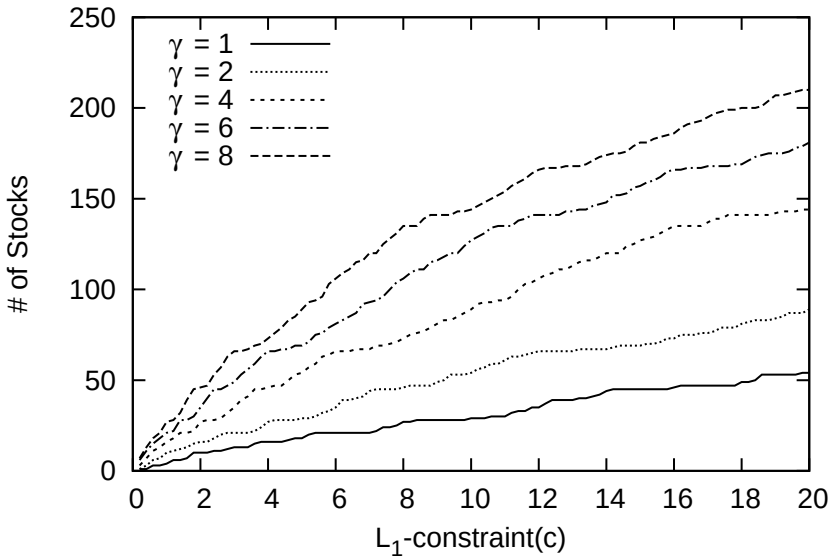


Figure 3.2: Number of stocks invested in as a function of the bound c of the L_1-constraint for different RRA coefficients γ. We have chosen $r = 0.02$ for the annual risk-free rate.

limited proportion of short positions. The short positions represent 47% of the amount of the L_1-norm of the unconstrained portfolio while they are reduced to 30 % for the constrained case when we use L_1-bound $c = 3$. Hence, the constraint controls both the sparsity and the proportion of short positions; see also Remark 3.1.3.

In Table 3.3, we compare the structure of the plug-in strategies, when the drift is estimated using (3.4.1) and $\gamma = 2$. For the L_1-constrained strategy, the bound c is chosen to be the optimal bound minimising the loss in expected utility, as in Table 3.1. We report the expected value of the number of stocks invested in, the number of shorts positions, and the fraction in L_1-norm of short positions. At the optimal bound, the number of stocks and short positions is reduced from 250 to approximately 86 stocks and from approximately 121 to 41 positions, respectively.

For the unconstrained strategy, the number and the fraction in L_1-norm of short positions both represent 48.36% of the positions and the L_1-norm respectively. We demonstrated in Table 3.2 that, by taking a restrictive bound, the proportion of short positions is reduced substantially. At the optimal bound, this is not the case. Indeed, the number of short positions

Stock	Excess Return		unconstrained weight	constrained weight
index i	Rank	$\mu_i - r$	π_i^*	$(\pi_c^*)_i$
238	1	0.60	0.57	0.39
216	2	0.43	0.18	0.13
199	3	0.41	0.57	0.33
150	4	0.39	0.52	0.18
234	5	0.37	0.66	0.21
224	6	0.36	0.57	0.14
112	7	0.34	0.40	0.18
211	8	0.33	0.52	0.02
168	9	0.32	1.00	0.28
9	12	0.29	0.48	0.04
183	15	0.27	0.58	0.17
175	17	0.25	1.44	0.06
219	222	0.01	-0.27	-0.03
64	236	-0.02	-1.19	-0.17
177	238	-0.02	-0.55	-0.08
245	240	-0.03	-1.06	-0.25
93	241	-0.03	-0.18	-0.03
120	242	-0.03	-0.10	-0.02
249	245	-0.03	-0.09	-0.02
198	249	-0.08	-0.17	-0.11
222	250	-0.10	-0.38	-0.17

Table 3.2: Characteristics of the 21 stocks selected for the constrained strategy π_c^* with known parameters. The first column corresponds to the general index of stock i. The second columns corresponds to the global rank in term of excess returns. The third column shows the associated excess return. The fourth and the fifth columns report the weights. Parameters: Bound of the L_1-constraint $c = 3$, coefficient of RRA $\gamma = 2$, annual risk-free rate $r = 0.02$.

and the fraction of short positions represents 47.11% and 45.75% of the L_1-norm[8]. Therefore, at the optimal level of diversification, the magnitude of the short positions between the unconstrained and the constrained strategies is essentially the same. This shows the superiority of the L_1-norm over the no-short sale constraint. We are able to control parameter uncertainty while keeping short positions representing almost half of the portfolio in L_1-norm.

To measure the stability of the previous quantities, we also report their standard deviation. The unconstrained strategy invests in all available, i.e., here 250, stocks. For the L_1-constrained strategy with a fixed bound, the number of stocks invested is random as it depends on the values of $\hat{\mu}$. The standard deviation of stocks invested in is 5.78, or equivalently it represents

[8]We find similar results for $\gamma = 3, 5, 7$.

	Expected value of			Weights
$\gamma = 2$	$\#\{i : (\hat{\pi}_c)_i \neq 0\}$	$\#\{i : (\hat{\pi}_c)_i < 0\}$	$\|\|\hat{\pi}_c^-\|\|_1/\|\|\hat{\pi}_c\|\|_1$ (%)	std. dev.
Merton	250 (0.00)	120.92 (4.81)	48.36 (0.47)	0.66
L_1, $c = 29$	86.39(5.78)	40.70 (4.27)	45.75 (1.71)	0.21

Table 3.3: This table reports the expected value (and standard deviation) of the number of stocks invested in, the number of shorts positions, the fraction in L_1-norm of short positions and the average of the standard deviation of the weights. The quantities presented are the sample mean and standard deviation over $M = 5000$ realisations. The optimal bound c of the L_1-constrained strategy is chosen as in Table 3.1. The unconstrained Merton plug-in strategy corresponds to $c = \infty$. Parameters: Coefficient of RRA $\gamma = 2$, annual risk-free rate $r = 0.02$.

6.7% of the expected number of stocks held in the portfolio. Compared to the unconstrained strategy, the variability of the number and the fraction in L_1-norm of short positions is proportionally higher for the L_1-constrained strategy. While the structure of the portfolio is more sensitive to the estimation of the drift, the overall stability of the portfolio is improved. Indeed, the average standard deviation of the weights is 68.2% smaller for the constrained strategy. Since the L_1-norm of the weights is bounded, extreme positions are forbidden and in turn variability is reduced.

Chapter 4

Out-of-Sample Study

Our goal is to investigate the out-of-sample structure and performance of the unconstrained and the optimally constrained plug-in strategies using empirical data. Therefore we will no longer have any parametric assumptions on the evolution of the asset prices.

From the theoretical considerations in the previous chapters, we have learned the following: We should rebalance our portfolio as frequently as possible (i.e., continuously in the best case). We will therefore trade daily on daily data. To reduce the effect of parameter uncertainty, we should choose a suitable bound c for the L_1-constraint. This is what we do by selecting the bound maximising the utility as we will discuss in detail in Section 4.1. Finally, we measure the performance of the strategies in terms of (expected) utility of terminal wealth.

When looking at the out-of-sample results, we need to keep in mind that the performance of the strategies is now also affected by the effects of discrete trading and discrete observations which was not the case in the previous sections. This is a problem that arises with all continuous-time models.

Our data set consists of the stocks that have been listed at least once in the S&P 500 and had daily returns for all trading days between January 2001 and December 2011.

We test our method between 2006 and 2011. At the beginning of each year, we select randomly a sample of 250 stocks and we fix it as the universe of stocks to invest in. Based on the five previous years of daily returns, we estimate the drift and volatility matrix of these stocks. We calibrate the

optimal bound of the L_1-constraint as outlined in Section 4.1.

Trading now takes place over intervals of a length of one month. At the beginning of each month, we assume a normalised initial endowment of one unit of cash and an annual risk-free rate of $r = 0.02$. At the end of the month we record the terminal utility. We average the terminal utilities that we obtain from trading 24 times consecutively over a one month interval, i.e., a two years period. We consider three consecutive time periods of two-years length and compare the performance of the plug-in strategies to the equally-weighted portfolio, which is a hard benchmark to beat; see DeMiguel et al. (2009b).

4.1 Choosing the L_1-bound

To find a suitable (and ideally optimal) bound for the L_1-constraint c, we present two alternative methods, namely a method that we call the *leave-one-block-out method* (LOB) and a *cross-validation method* (CV).

For both methods we start as follows. On the first trading day of each month, we divide the multivariate time series of daily returns of the five previous years into 60 blocks of one month. Based on the 59 first blocks, we estimate the drift and the volatility matrix using (3.4.1) and (3.4.2). Next, we compute the constrained plug-in strategy in the interval $[0, ||\hat{\pi}||_1]$ for each value of the bound c that is on a grid with grid size $\Delta c = 0.1$. Finally, we invest (rebalancing daily) on the remaining block with respect to the constrained plug-in strategy; see also (4.2.1).

In the leave-one-block-out (LOB) method, we select the bound c maximising the utility of final wealth. In the cross-validation (CV) method, we repeat this procedure 1000 times by taking a random sample with replacement of the 60 blocks and we select the bound c maximising the average utility of final wealths.

As the method of estimation is the same for all months, estimation risk is constant through time and the variation of the optimal bound depends mainly on the market conditions of the calibration period.

Table 4.1 contains the numerical results for the optimal L_1-bounds using the two different methods. For the LOB method, the optimal bound varies widely. While the first quartile stays small over all periods, the median (mean) and the third quartile are of different magnitude within and across

each period. Moreover, there are five months in 2008, and three months in 2009, with an optimal bound equal to zero. During these months, the strategy consists in holding only the risk-free asset.

In 2006-2007 and 2010-2011, the evolution of the optimal bound calibrated with the CV method is stable as the difference between the first quartile and the third quartile is small. In 2008-2009, there is a large variation of the optimal bound because it is a period of transition. Indeed, the mean over time of the optimal bound is 20.62 in 2008 and 5.03 in 2009. Furthermore, we observe a substantial drop in the median (mean) of the optimal bound, from 12.00 (13.39) in 2006-2007 to 3.25 (3.27) in 2010-2011. In 2010-2011, the CV method incorporates the fact that preceding years were very volatile and it delivers a more conservative and restrictive bound.

	Optimal L_1-Bound		
	1st Quartile	Median (Mean)	3rd Quartile
2006-2007			
LOB	0.78	21.65 (65.00)	68.42
CV	10.85	12.00 (13.39)	15.02
2008-2009			
LOB	0.00	1.80 (29.78)	23.20
CV	3.93	10.90 (12.82)	20.32
2010-2011			
LOB	0.65	5.10 (44.63)	91.15
CV	2.20	3.25 (3.27)	4.60

Table 4.1: This table reports the summary statistics of the optimal bounds c for the leave-one-block out (LOB) and the cross validation (CV) methods for the three time periods that we consider. Parameters: Coefficient of RRA $\gamma = 2$, investment horizon $T = 1/12$, and annual risk-free rate $r = 0.02$

4.2 Performance of the Plug-in Strategies

At the beginning of each month we pick the optimal L_1-bound which was identified with one of our two methods. We estimate the drift and the volatility matrix using (3.4.1) and (3.4.2) based on the past five years of daily returns. Then we compute the unconstrained and the constrained plug-in strategies. The strategies are then held constant over a period of one month, i.e., $T = 1/12$. This means we rebalance the positions daily to

keep the weights constant over the investment period. For each month M and strategy π^M which is constant in time over the month, the dynamics of the wealth denoted by X^M is given by

$$X^M(t+1) = X^M(t)\left(1 + r_d + \sum_{i=1}^{d} \pi_i^M \left(r_i(t+1) - r_d\right)\right), \tag{4.2.1}$$

where r_d is daily risk-free rate and $r_i(t+1)$ is the simple net return of asset i between day t and day $t+1$. Since there are 21 trading days in one month, we have $t = 0, \ldots, 20$. At the end of each month M, we obtain a utility of final wealth $U_\gamma\left(X^M(20)\right)$.

In Table 4.2, we report the summary statistics of the utility of final wealth for three blocks of two years, i.e., $M = 1, \ldots, 24$ in each block. Over all periods, the unconstrained plug-in strategy performs poorly, because of its extreme variance. For instance, the unconstrained strategy gives its highest return in August 2007, and directly leads to bankruptcy a month later. Furthermore, the standard deviation of monthly returns of the wealth is at its peak in 2006-2007 with a magnitude of 1219.95%. Although on a smaller scaler, its remains between 89.61% and 120.02% over the subsequent periods. As a result, from January 2008 to May 2009, the wealth reaches zero in almost every month. As we measure the performance with a power utility function with $\gamma > 1$, hitting zero for the wealth translates into infinitely negative utility.

The constrained plug-in strategy calibrated with a bound with the LOB method is halfway between the unconstrained strategy and the constrained strategy calibrated with the CV method. It is the most successful strategy for two periods, with a utility of -0.74 in 2006-2007 and of -0.81 in 2010-2010. This performance comes at the cost of a very large variance of monthly returns. In terms of standard deviation of utility of final wealth and monthly returns, the strategy is similar to the unconstrained portfolio. Therefore, amidst the financial crisis of 2008-2009, the wealth also hits zero.

The constrained plug-in strategy calibrated with the CV method performs better than the unconstrained strategy in all periods and it is more stable over time than both the unconstrained and the L_1-constrained with LOB strategies. Although, on average, its returns have a larger standard deviation than the equally-weighted portfolio, the mean utility of final wealth

is comparable in all periods. The utility of the constrained strategy is higher in 2006-2007, measuring -0.97 versus -0.99. In 2008-2009, the large proportion of estimated short positions hurts the performance of the constrained strategy and it has a smaller mean utility of -1.46 versus -1.01. Finally the mean utility of final wealth is equal for both strategies in 2010-2011 with a value of -0.99.

Given the nature of the out-of-sample study trading takes place in discrete time. It is well-known that strategies that are optimal in continuous-time cannot just be discretised to obtain strategies that are optimal in discrete-time; see Rogers (2001) and Bäuerle et al. (2013) for extensive discussion. In particular, short-selling is never optimal when trading takes place in discrete-time. Since in our study the L_1-constraint does not rule out short-selling, the superiority of the L_1-constrained portfolio over the unconstrained portfolio is not just a consequence of no short-selling.

	Utility of Final Wealth			Return of Portfolio (%)		
	Min.	Mean (std.dev.)	Max	Min.	Mean (std.dev.)	Max.
			2006-2007			
Merton	$-\infty$	$-\infty$	-0.02	-100	290.84 (1219.95)	5765.79
L_1, c with LOB	-1.00	-0.74 (0.32)	-0.02	0.17	355.86 (1201.51)	5765.79
L_1, c with CV	-1.38	-0.97 (0.26)	-0.43	-27.36	12.88 (41.37)	134.02
EWE	-1.08	-0.99 (0.03)	-0.95	-7.44	1.01 (3.14)	5.09
			2008-2009			
Merton	$-\infty$	$-\infty$	-0.29	-100	-60.07 (89.61)	245.44
L_1, c with LOB	$-\infty$	$-\infty$	-0.12	-100	66.57 (170.09)	744.64
L_1, c with CV	-4.95	-1.46 (0.99)	-0.29	-79.81	0.55 (72.24)	242.19
EWE	-1.27	-1.01 (0.13)	-0.74	-21.35	0.59 (12.95)	34.50
			2010-2011			
Merton	-10.11	-2.43 (2.55)	-0.20	-90.11	6.39 (120.02)	398.11
L_1, c with LOB	-1.00	-0.81 (0.24)	-0.20	0.17	48.25 (99.01)	398.11
L_1, c with CV	-1.17	-0.99 (0.08)	-0.86	-14.28	1.48 (8.22)	15.72
EWE	-1.13	-0.99 (0.06)	-0.87	-11.56	1.07 (6.19)	15.21

Table 4.2: This table reports the summary statistics of the utility and the monthly returns of final wealth out-of-sample. The quantities presented are computed over each block of 24 months. The optimal bounds c of the L_1-constrained strategies are calibrated using the leave-one-block-out (LOB) and the cross-validation (CV) methods. Parameters: Coefficient of RRA $\gamma = 2$, initial wealth $X^M(0) = 1$, investment horizon $T = 1/12$, and annual risk-free rate $r = 0.02$.

4.3 Structure and Stability of the Strategies

In Table 4.3, we report the mean and standard deviation of stocks invested in, of the short positions, of the fraction of short positions in L_1-norm and the mean of the monthly portfolio turnover for four differernt investment strategies: the unconstrained Merton strategy, referred to as *Merton* in Table 4.3, the L_1-constrained strategy where the bound was computed using the LOB-method, the L_1-constrained strategy where the bound was computed using the CV-method, and the equally weighted portfolio, referred to as *EW*.

The monthly portfolio turnover is defined as

$$\text{Monthly Turnover } = \frac{1}{20}\sum_{t=2}^{21}\sum_{i=1}^{d}\left(\left|\pi_i - \pi_i\left(t^-\right)\right|\right),$$

where π_i is the constant target weight for asset i, and $\pi_i\left(t^-\right)$ is the fraction of wealth (i.e., the weight) invested in asset i just before rebalancing at time t. All portfolios are rebalanced daily and 21 corresponds to the number of trading days in a month. If the wealth hits zero during the month, we set the turnover to ∞ and stop the rebalancing.

For the unconstrained strategy, the structure remains unchanged throughout the whole test period. The mean fraction of short positions of the L_1-norm stays between 120 and 124, the mean ratio $||\pi^-||_1/||\pi||_1$ stays between 47% and 50% and the portfolio turnover is infinite both in 2006-2007 and 2008-2009. It is finite in 2010-2011, but still large.

For the constrained plug-in strategy with the LOB method, the mean number of risky assets invested in and that of short positions is the largest in 2006-2007 and 2010-2011. In 2008-2009, these quantities are significantly smaller, because there are actually several months where there is no investment in risky assets; see also Table 4.1. Nevertheless, on average, the mean ratio $||\pi^-||_1/||\pi||_1$ stays stable over each period, with a larger standard deviation in 2008-2009. In terms of turnover, the strategy performs poorly. The turnover is very high in 2006-2007 and it is infinite in 2008-2009. In 2010-2011, it is reduced but still much larger than the turnover of the constrained plug-in strategy calibrated with the CV method and the equally-weighted portfolio. This turnover is high because of the variability of the optimal bound to different market conditions.

For the constrained plug-in strategy calibrated with the CV method, the mean number of risky assets invested in and that of short positions decreases in each period. The mean fraction of short positions of the L_1-norm does not. The mean ratio $||\pi^-||_1/||\pi||_1$ is actually the largest in 2008-2009, with a value of 44.74%, and it is similar in 2006-2007 and 2010-2011, with 35.48% and 33.47%, respectively. For 2008-2009, the mean of estimated excess returns decreases sharply during the year 2008 and it implies a significantly larger fraction of short positions in the constrained portfolio.The constrained plug-in strategy calibrated with the CV method is more stable than the constrained plug-in strategy calibrated with LOB method. The standard deviation of each measure is reduced by at least 40%. Finally, although its turnover is much larger than the equally-weighted portfolio turnover in 2006-2007 and 2008-2009, the turnovers are comparable during the period 2010-2011.

4.4 Transaction Costs

For a complete analysis of the out-of-sample performance it is interesting to consider transaction costs as well. It is well-known that if transaction costs are included, holding the Merton ratio and hence having to continuously rebalance the portfolio is no-longer optimal. In particular, the investor can find herself in a position where it is optimal not to trade at all. This would happen if her positions would be within the so-called no-trading region which is used to characterise the optimal trading strategies; see, e.g., Davis and Norman (1990). Nevertheless, we can still investigate what transaction costs one would have to pay when using the trading strategies that we have derived in our setting without transaction costs.

Keim and Madhavan (1998) distinguish implicit costs such as the bid-ask spread which are proportional to the cash amount of shares traded, and explicit trading costs, such as brokerage commissions, which are fixed costs per share traded. They measure both types of trading costs as a percentage of the face value of investment. However, Goldstein et al. (2009) argue that brokerage commissions should not be measured as a percentage of face value. Indeed, brokers ignore most available prices and they charge commissions in exact cents per share. Moreover, large trades tends to have higher commissions fees per share and they may also have larger proportional

	Mean Value of			
	$\#\{i : \pi_i \neq 0\}$	$\#\{i : \pi_i < 0\}$	$\|\|\pi^-\|\|_1/\|\|\pi\|\|_1$ (%)	Turnover
		2006-2007		
Merton	250 (0.00)	123.13 (5.01)	47.92 (0.71)	∞
L_1, LOB	90.67 (95.38)	42.58 (47.22)	32.25 (19.38)	12.31
L_1, CV	43.58 (9.82)	17.96 (4.84)	35.48 (4.00)	0.78
EW	250 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01
		2008-2009		
Merton	250 (0.00)	120.08 (5.73)	49.39 (0.36)	∞
L_1, LOB	48.75 (75.00)	22.58 (36.21)	34.82 (21.39)	∞
L_1, CV	32.46 (18.69)	14.67 (8.46)	44.74 (5.39)	1.66
EW	250 (0.00)	0.00 (0.00)	0.00 (0.00)	0.02
		2010-2011		
Merton	250 (0.00)	121.71 (5.75)	49.14 (0.24)	28.65
L_1, LOB	83.79 (99.88)	41.38 (50.08)	31.58 (19.66)	4.95
L_1, CV	18.92 (8.73)	9.67 (4.48)	33.47 (10.95)	0.08
EW	250 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01

Table 4.3: This table reports the mean value (standard deviation) of the number of stocks invested in, of the number of shorts positions, of the fraction in L_1-norm of short positions, and the mean monthly turnover for four different investment strategies. The mean is computed over each block of 24 months. The optimal bound c of the L_1-constrained strategy is calibrated with the leave-one-block out (LOB) and the cross validation (CV) methods. Parameters: Coefficient of RRA $\gamma = 2$, investment horizon $T = 1/12$, and annual risk-free rate $r = 0.02$.

transactions costs because of their potential price impact.

Assuming that proportional trading costs are constant and equal for long and short positions, the total trading costs for one month are defined as

$$\begin{aligned}\text{Monthly Trading Costs} = &\sum_{t=2}^{21}\sum_{i=1}^{d} \kappa S_i(t)\left|N_i(t) - N_i\left(t^-\right)\right| \\ &+ C \cdot \#\{i : N_i(t) \neq 0\},\end{aligned}$$

where κ is the coefficient of charged proportion, $N_i(t)$ is the number of stock i held at time t and $N_i(t^-)$ is number of stocks i held just before rebalancing at time t, and C is the commission price per share.

On average, κ represents 0.3% of the face value invested in each stock

and fixed costs commissions C vary between 2 and 5 cents per share. For simplicity, we take $\kappa = 0.3\%$ and $C = 0.025$. As a benchmark, we consider a small investor, with a coefficient of RRA $\gamma = 2$ and an initial wealth of $X_0^M = 25000$, so that trades do not have any price impact. Note that total trading costs are not proportional to the initial wealth invested. Their dependence on the initial wealth is especially important for the plug-in strategies as the number of shares varies widely.

During the financial crisis 2008-2009, the volatility of the markets leads to a large variability of the positions held in the portfolio even for the constrained plug-in strategies; see Table 4.3. As a result, the proportional transaction costs are much higher for these strategies compared to the equally-weighted portfolio. During the period 2006-2007 and 2010-2011, the constrained strategies perform on average at least as well as the equally-weighted portfolio[1]. While having a significantly higher performance during these years, the constrained strategy calibrated with the LOB method faces large transaction costs. For instance, the average of total trading costs are 6625.08 and 9963.01 units of cash in 2006 and 2010 respectively. In the same years, the total transaction costs of the constrained strategy calibrated with CV and the equally-weighted portfolio are 664.34 and 140.35 in 2006, and 95.42 and 140.68 in 2010. Hence, in 2010, both strategies deliver the same average utility and the transaction costs are on average smaller for the constrained strategy.

In summary, outside the financial crisis L_1-constrained strategies deliver on average a higher utility than the equally-weighted portfolio. Regarding transaction costs, the constrained strategies have large proportional costs and low fixed commissions. Furthermore, because of fixed commissions the equally-weighted portfolio and the unconstrained portfolio suffer from complete diversification. The trade-off between the two types of costs depends on the size of the positions and the sparsity of the strategies. For the constrained strategy with LOB, total trading costs remain large, while the constrained strategy with CV can have lower trading costs than the equally-weighted portfolio. Note that the effect of proportional trading costs will be reduced for a higher level of risk aversion, as the weights in the risky assets, and in turn the number of shares, decreases with γ.

[1]In Table 4.2, the utilities are given for an investor with a normalised initial wealth $X^M(0) = 1$. For $X^M(0) = 25000$, the results differ by the multiplicative factor $(X(0))^{1-\gamma}$.

4.5 A Note on the Estimation of the Covariance Matrix

In our out-of-sample study we need to estimate the covariance matrix as well. A numerical problem emerges here, since the sample covariance matrix tends to be ill-conditioned and therefore the computation of Σ^{-1} can be numerically unstable. Successful regularisation techniques such as the shrinkage approach of Ledoit and Wolf (2004a) have been proposed to control the conditioning of the matrix.

In our out-of-sample study, we have therefore also investigated the performance of the plug-in strategy based on the covariance estimator proposed by Ledoit and Wolf (2004a). While it performs better than the simple unconstrained plug-in strategy in 2010-2011, the wealth also hits zero in 2006-2007 and 2008-2009, as for the simple unconstrained plug-in strategy.

In certain months, the method of estimation of Ledoit and Wolf (2004a) can have a positive first order effect. For example, in the period 2010-2011, the minimum utility is reached in both cases in January 2011, with a utility of -10.11 without shrinkage and -6.52 with shrinkage. However, with the measure of average utility, there is no difference during 2006-2007 and 2008-2009, because the wealth hits zero at least once. As our measure of performance is non symmetric, it smooths the difference between good realisations and it assigns very low values to bad realisations.

We also tested the performance of the L_1-constrained strategies with the covariance estimator of Ledoit and Wolf (2004a). The values of the optimal bound are very close to the results reported earlier and the performance of the strategy is essentially the same.

For both types of plug-in strategies, we start with the sample covariance matrix of log-returns and it is shrunk towards the covariance matrix based on the constant correlation model as discussed in Ledoit and Wolf (2004a). We provide our numerical results in Table 4.4.

4.6 Conclusion

For a coefficient of RRA bigger than one, we have shown that the loss in expected utility due to parameter uncertainty depends on the coefficient of RRA and the number of risky assets in a highly non-linear fashion. In

minimum/mean/maximum of utility over 24 months			
$\gamma = 2$	2006-2007	2008-2009	2010-2011
unconst	-∞/ -∞/-0.02	-∞/-∞/-0.29	-10.11/-2.43/-0.20
unconst, L&W	-∞/-∞/ -0.03	- ∞/-∞/-0.11	-6.52/-1.80/-0.20
L_1	-1.38/-0.97/-0.43	-4.95/ -1.46/-0.29	-1.17/-0.99/-0.86
L_1, L&W	-1.36/-0.97/-0.43	- 4.80/ -1.48/-0.26	-1.16/-0.99/-0.85

Table 4.4: This table reports the minimum, mean and maximum of utility of terminal wealth averaged over 24 months for three time periods. Four different strategies are considered: The unconstrained plug-in strategy using (3.4.1) and (3.4.2), denoted by unconst, unconstrained plug-in strategy using (3.4.1) and the method proposed by Ledoit and Wolf (2004a) for the covariance matrix, denoted by unconst, L&W, and the corresponding strategies where we imposed an L_1-constraint, denoted by L_1 and L_1, L&W, respectively.

particular, as the number of risky assets increases, the loss can become infinite. Therefore, the challenge is to reduce the number of risky assets to limit the exposure to estimation risk when implementing the plug-in strategy. Putting an L_1-constraint on the weights of the plug-in strategy induces sparsity in the portfolio and it is an efficient method to reduce the negative effect of parameter uncertainty on its performance.

By characterising the structure of the L_1-constrained strategy for independent stocks, we show that the level of sparsity is determined by the coefficient of RRA. For a general covariance matrix structure, we demonstrate, based on a simulation study, that there exists an optimal bound minimising the loss due to estimation for each level of risk aversion. Hence, estimation risk can be efficiently controlled with the L_1-constraint by taking into account the level of risk aversion of the investor.

Based on a CRRA utility maximisation framework, we provide an economical justification for using the L_1-constraint as a way to reduce estimation risk. Indeed, for each level of risk aversion, we choose the appropriate level of sparsity and we attain the optimal trade-off between the gain of diversification and the loss due to estimation risk. Finally, we show that L_1-constrained strategies can successfully be applied on empirical data.

Appendix A

Proofs of Chapter 2

Lemma A.0.1. *Let π be constant in time, sufficiently integrable and independent of $W(T)$, then*

$$V_\gamma(\pi|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)TM_\gamma(\pi)\right)\right].$$

Proof of Lemma A.0.1. For $\gamma > 1$, we rewrite the expected utility

$$V_\gamma(\pi|\mu,\sigma) = \frac{X(0)^{1-\gamma}}{1-\gamma}\mathbb{E}\Bigg[\exp\Bigg((1-\gamma)\left(r + \pi^\top(\mu - r1) - \frac{1}{2}\pi^\top\Sigma\pi\right)T + (1-\gamma)\pi^\top\sigma W(T)\Bigg)\Bigg]$$

as

$$V_\gamma(\pi|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)T\left(\pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi\right)\right)Z(T)\right]$$

with

$$Z(T) = \exp\left((1-\gamma)\pi^\top\sigma W(T) - \frac{(1-\gamma)^2}{2}\pi^\top\Sigma\pi T\right) \text{ and } \mathbb{E}\left[Z(T)|\,\pi\right] = 1.$$

Hence,

$$V_\gamma(\pi|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)T\left(\pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi\right)\right)\right].$$

□

Lemma A.0.2. *Let π be constant in time, sufficiently integrable and inde-*

pendent of $W(T)$*, then*

$$V_\gamma(\pi|\mu,\sigma) = \mathbb{E}\left[\exp\left(-\frac{(1-\gamma)\gamma T}{2}(\pi-\pi^*)^\top \Sigma(\pi-\pi^*)\right)\right] V_\gamma(\pi^*|\mu,\sigma).$$

Proof of Lemma A.0.2. From Lemma A.0.1 we obtain

$$V_\gamma(\pi|\mu,\sigma) = K_\gamma \mathbb{E}[\exp((1-\gamma)TM_\gamma(\pi))]. \tag{A.0.1}$$

We write the mean-variance term as

$$\begin{aligned}
M_\gamma(\pi) &= \pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi. \\
&= \pi^\top(\mu - r1) - \frac{\gamma}{2}(\pi-\pi^*)^\top\Sigma(\pi-\pi^*) + \frac{\gamma}{2}\pi^{*\top}\Sigma\pi^* - \gamma\pi^\top\Sigma\pi^* \\
&= \pi^\top(\mu - r1) - \frac{\gamma}{2}(\pi-\pi^*)^\top\Sigma(\pi-\pi^*) + \frac{\gamma}{2}\pi^{*\top}\Sigma\pi^* - \pi^\top(\mu - r1) \\
&= -\frac{\gamma}{2}(\pi-\pi^*)^\top\Sigma(\pi-\pi^*) + \frac{\gamma}{2}\pi^{*\top}\Sigma\pi^* \\
&= -\frac{\gamma}{2}(\pi-\pi^*)^\top\Sigma(\pi-\pi^*) + \frac{1}{2\gamma}(\mu - r1)^\top\Sigma^{-1}(\mu - r1).
\end{aligned}$$

Therefore, by (2.2.3) and (A.0.1),

$$V_\gamma(\pi|\mu,\sigma) = L_\gamma(\pi,\pi^*)\, V_\gamma(\pi^*|\mu,\sigma),$$

with

$$L_\gamma(\pi,\pi^*) = \mathbb{E}\left(\exp\left(-\frac{(1-\gamma)\gamma T}{2}(\pi-\pi^*)^\top\Sigma(\pi-\pi^*)\right)\right).$$

□

Proof of Lemma 2.3.1. By (2.3.1) and (2.3.2), $\hat{\mu}$ and $\hat{\pi}$ are independent $W(T)$. Hence, we can apply Lemma A.0.1 and the result follows. □

Lemma A.0.3. *Let* $Y \sim \mathcal{N}_d(0,\Lambda)$*, where* Λ *is symmetric and positive definite. Let* $B \in \mathbb{R}^d$ *and let* $C \in \mathbb{R}^{d\times d}$ *be a symmetric positive definite matrix such that* Λ *and* C *commute. Let* $t \in \mathbb{R}$ *and suppose that* $\Lambda^{-1} - tC$ *is*

symmetric and positive definite. Then,

$$\mathbb{E}\left[\exp\left(t\left(B^\top Y+\frac{1}{2}Y^\top CY\right)\right)\right]=\exp\left(\frac{1}{2}t^2B^\top\left(\Lambda^{-1}-tC\right)^{-1}B\right)\cdot\frac{1}{\sqrt{|I-t\Lambda C|}}.$$

Proof of Lemma A.0.3.

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t\left(B^\top Y+\frac{1}{2}Y^\top CY\right)\right)\right]\\
&=\frac{1}{\sqrt{|\Lambda|}}\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\exp\left(t\left(B^\top y+\frac{1}{2}y^\top Cy\right)\right)\exp\left(-\frac{1}{2}y^\top\Lambda^{-1}y\right)dy\\
&=\frac{1}{\sqrt{|\Lambda|}}\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\exp\left(tB^\top y\right)\exp\left(-\frac{1}{2}y^\top\left(\Lambda^{-1}-tC\right)y\right)dy\\
&=\frac{\sqrt{|\left(\Lambda^{-1}-tC\right)^{-1}|}}{\sqrt{|\Lambda|}}\mathbb{E}\left[\exp\left(tB^\top Z\right)\right],\ \text{with } Z\sim\mathcal{N}_d\left(0,\left(\Lambda^{-1}-tC\right)^{-1}\right)\\
&=\frac{1}{\sqrt{|I-t\Lambda C|}}\exp\left(\frac{1}{2}t^2B^\top\left(\Lambda^{-1}-tC\right)^{-1}B\right).
\end{aligned}$$

□

Proof of Theorem 2.3.2. By (2.3.1) and (2.3.2), $\hat{\pi}$ is independent of $W(T)$. By Lemma A.0.2, we need to compute the loss factor

$$\mathbb{E}\left(\exp\left(-\frac{(1-\gamma)\gamma T}{2}\left(\hat{\pi}-\pi^*\right)^\top\Sigma\left(\hat{\pi}-\pi^*\right)\right)\right).$$

Since $\hat{\pi}-\pi^*\sim N\left(0,V_0^2\right)$ we can apply Lemma A.0.3 with $\Lambda=V_0^2=\frac{1}{\gamma^2 t_{obs}}\Sigma^{-1}$, $B=0$, $C=\Sigma$, and $t=-(1-\gamma)\gamma T$, and we obtain

$$\begin{aligned}
V_\gamma\left(\hat{\pi}|\mu,\sigma\right)&=\frac{1}{\sqrt{|I+(1-\gamma)\gamma TV_0^2\Sigma|}}V_\gamma\left(\pi^*|\mu,\sigma\right)\\
&=\frac{1}{\sqrt{\left(1+\frac{(1-\gamma)}{\gamma t_{obs}}T\right)^d}}V_\gamma\left(\pi^*|\mu,\sigma\right).
\end{aligned}$$

Note that Λ and C do indeed commute here. For $\gamma>1$, the term $1+\frac{(1-\gamma)}{\gamma t_{obs}}T$ is strictly positive because of the assumption $t_{obs}>T$. □

Appendix B

Proofs of Chapter 3

Proof of Lemma 3.3.1. The expected utility of any L_1-constrained strategy π_c is well defined because the L_1-norm of π_c is bounded by c. The result then follows from Lemma A.0.2. □

Proof of Proposition 3.3.2. Let $m = -c||\mu^{(d)} - r1^{(d)}||_\infty - \frac{\gamma c^2}{2}||\Sigma^{(d)}||_\infty$. Since $||\pi_c^{(d)}||_1 \leq c$,

$$m \leq (\pi_c^{(d)})^\top \left(\mu^{(d)} - r1^{(d)}\right) - \frac{\gamma}{2}\left(\pi_c^{(d)}\right)^\top \Sigma^{(d)}\pi_c^{(d)},$$

where $||\Sigma^{(d)}||_\infty = \max\left\{||\Sigma^{(d)}x||_\infty : x \in \mathbb{R}^d \text{ with } ||x||_\infty = 1\right\}$.

Thus, for $\gamma \neq 1$, we have, using Lemma 3.3.1,

$$V_\gamma\left(\pi_c^{(d)}|\mu^{(d)}, \sigma^{(d)}\right) \geq \frac{1}{1-\gamma}X(0)^{1-\gamma}\exp\left((1-\gamma)T(r+m)\right).$$

For $\gamma = 1$,

$$V_1\left(\pi_c^{(d)}|\mu^{(d)}, \sigma^{(d)}\right) \geq \log\left(X(0)\right) + T(r+m).$$

□

Corollary B.0.4. *Under the assumptions of Proposition 3.3.3, we have the following convergence properties of the loss factors*

$$L_\gamma(\pi_c^*, \pi^*) \to 1 \text{ as } c \to \infty,$$

$$L_\gamma(\hat{\pi}_c, \pi_c^*) \to L_\gamma(\hat{\pi}, \pi^*) \text{ and } L_\gamma(\hat{\pi}_c, \pi^*) \to L_\gamma(\hat{\pi}, \pi^*) \text{ as } c \to \infty.$$

Proof Corollary B.0.4. Note that $\pi_c^* \to \pi^*$, as $c \to \infty$. Since π_c^* and π^* are deterministic, by Lemma A.0.1, $V_\gamma(\pi_c^*|\mu, \sigma) = v(M_\gamma(\pi_c^*))$, and $V_\gamma(\pi^*|\mu, \sigma) = v(M_\gamma(\pi^*))$, where $v(x) = K_\gamma \exp((1-\gamma)Tx)$ is continuous. Hence,

$$V_\gamma(\pi_c^*|\mu, \sigma) \to V_\gamma(\pi^*|\mu, \sigma), \text{ as } c \to \infty$$

and

$$L_\gamma(\pi_c^*, \pi^*) \to 1 \text{ as } c \to \infty.$$

For the constrained plug-in strategy, $\hat{\pi}_c \to \hat{\pi}$, as $c \to \infty$, and the proof boils down to showing that

$$V_\gamma(\hat{\pi}_c|\mu, \sigma) = \mathbb{E}(v(M_\gamma(\hat{\pi}_c))) \to \mathbb{E}(v(M_\gamma(\hat{\pi}))) = V_\gamma(\hat{\pi}|\mu, \sigma), \text{ as } c \to \infty.$$

This is true, as we have shown in the proof of Proposition 3.3.3 that, for all $c \geq 0$, $|v(M_\gamma(\hat{\pi}_c))| \leq |v(Y)| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, where Y is given in (3.3.2). Therefore, the dominated convergence theorem can be applied to the sequence of random variables $v_n = v(M_\gamma(\hat{\pi}_{c_n}))$, where c_n is any sequence converging to $+\infty$. □

Appendix C

Number of Paths for the Monte-Carlo Method

Suppose that we are to estimate $\mu_Y = \mathbb{E}(Y)$ by $\bar{Y}_M = \frac{1}{M}\sum_{i=1}^{M} Y^i$, where $Y^i, i = 1, \ldots, M$ are i.i.d realisations of Y. By the Central Limit Theorem, the approximate $(1-\epsilon)$-confidence interval (for $0 < \epsilon < 1$) for μ_Y is given by

$$\left[\bar{Y}_M - a\frac{\sigma_Y}{\sqrt{M}}, \bar{Y}_M + a\frac{\sigma_Y}{\sqrt{M}}\right],$$

where $\Phi(a) = 1 - \frac{\epsilon}{2}$ and $\sigma_Y^2 = \mathbf{Var}(Y)$. Now suppose that we want to estimate $\mu_Z = \mathbb{E}(Z)$, with $Z = kY$, k being a constant. In this case, the accuracy of the Monte Carlo estimator of μ_Z, is given by $\frac{|k|\sigma_Y}{\sqrt{M}}$. Indeed, the $(1-\epsilon)$-confidence interval is

$$\left[\bar{Z}_M - a\frac{\sigma_Z}{\sqrt{M}}, \bar{Z}_M + a\frac{\sigma_Z}{\sqrt{M}}\right] = \left[k\bar{Y}_M - a\frac{|k|\sigma_Y}{\sqrt{M}}, k\bar{Y}_M + a\frac{|k|\sigma_Y}{\sqrt{M}}\right].$$

Therefore, the constant k is shrinking or widening the confidence interval of the Monte Carlo estimation, depending on whether it is smaller or bigger than one. Furthermore,

$$\begin{aligned}&\mathbb{P}\left(\mu_Z \in \left[k\bar{Y}_M - a\frac{|k|\sigma_Y}{\sqrt{M}}, k\bar{Y}_M + a\frac{|k|\sigma_Y}{\sqrt{M}}\right]\right)\\ &= \mathbb{P}\left(\mu_Y \in \left[\bar{Y}_M - a\frac{\sigma_Y}{\sqrt{M}}, \bar{Y}_M + a\frac{\sigma_Y}{\sqrt{M}}\right]\right).\end{aligned}$$

For k large, μ_Z is a larger quantity than μ_Y, as $\mu_Z = k\mu_Y$. For a fixed number of realisations, we just estimate a bigger quantity with a bigger error.

If we want to estimate the expected utility for a plug-in strategy, we sample M i.i.d realisations $Y^1, \ldots, Y^M$ of

$$Y_\gamma = \exp\left((1-\gamma)\, T\left(\hat{\pi}^\top(\mu - r1) - \frac{\gamma}{2}\hat{\pi}^\top \Sigma \hat{\pi}\right)\right).$$

Next, for K_γ given in (2.2.4), we define the Monte-Carlo estimator of the expected utility by

$$\bar{V}_{\gamma,M}(\hat{\pi}|\mu,\sigma) = K_\gamma \bar{Y}$$

with

$$\bar{Y} = \frac{1}{M}\sum_{i=1}^{M} Y^i.$$

The variance of the Monte-Carlo is then given by

$$\mathbf{Var}\left(\bar{V}_{\gamma,M}\right) = K_\gamma^2 \mathbf{Var}\left(Y_\gamma\right)/M.$$

The constant and the random variable depend non-linearly on the parameter γ and the variance of $\bar{V}_{\gamma,M}$ is non-monotonic in γ. Hence, the number of realisations to reach a given level of accuracy varies greatly with γ.

As we do not know $\mathbf{Var}(Y_\gamma)$, we have to replace it in the confidence interval by an estimator. We choose

$$s_M^2 = \frac{1}{M-1}\sum_{i=1}^{M}\left(Y^i - \bar{Y}\right)^2.$$

s_M^2 is an unbiased estimator of $\mathbf{Var}(Y_\gamma)$. In our case, we take $M = 5000$, so that the level of accuracy is such that, for all $\gamma \in [1,7]$,

$$\frac{K_\gamma s_M}{\sqrt{M}} \approx \bar{V}_{\gamma,M}(\hat{\pi}_c|\mu,\sigma)/100.$$

The same type of issue occurs for the computation of the loss. We define

the loss factor by

$$V_\gamma\left(\hat{\pi}|\mu,\sigma\right) = L_\gamma\left(\hat{\pi},\pi^*\right) V_\gamma\left(\pi^*|\mu,\sigma\right)$$

or equivalently by

$$L_\gamma\left(\hat{\pi},\pi^*\right) = \frac{V_\gamma\left(\hat{\pi}|\mu,\sigma\right)}{V_\gamma\left(\pi^*|\mu,\sigma\right)}.$$

Then, to compute this loss numerically, one would use the estimate

$$\bar{L}_{\gamma,M}\left(\hat{\pi},\pi^*\right) = \frac{\bar{V}_{\gamma,M}\left(\hat{\pi}|\mu,\sigma\right)}{V_\gamma\left(\pi^*|\mu,\sigma\right)}.$$

Again here the accuracy of $\bar{L}_{\gamma,M}\left(\hat{\pi},\pi^*\right)$ is very different from the accuracy of the estimation of the expected utility. As $|V_\gamma\left(\pi^*|\mu,\sigma\right)| << 1$,

$$\frac{1}{|V_\gamma\left(\pi^*|\mu,\sigma\right)|}\sqrt{\mathbf{Var}\left(\bar{V}_{\gamma,M}\right)} >> \sqrt{\mathbf{Var}\left(\bar{V}_{\gamma,M}\right)},$$

and the necessary number of steps would have to be changed accordingly. It could be that, even if the expected utility is accurately estimated, the loss factor is not. Hence, we compute the logarithm of $\bar{L}_{\gamma,M}\left(\hat{\pi},\pi^*\right)$. In this case, the logarithm of the ratio is equal to the difference of the logarithms and the expected utility $\bar{V}_{\gamma,M}\left(\hat{\pi}|\mu,\sigma\right)$ is estimated independently of $V_\gamma\left(\pi^*|\mu,\sigma\right)$.

Part II

A Bayesian Approach to Risk Assessment in Banking Networks

Chapter 5

Estimation of Bilateral Exposures

I failed to see the drift of Bruno's argument. "Surely anybody *could be knocked down," I said: "thick or thin wouldn't matter."*

- Lewis Carroll, *Sylvie and Bruno Concluded*

5.1 Introduction

The stability of the interbank lending market is of central interest for regulators. The interconnectivity of banks' liabilities and assets may lead to a domino effect of defaults spreading through the whole banking sector. To manage and prevent such scenarios, it is fundamental to understand the default mechanisms triggering contagion of banks' defaults. The literature on banking networks has shown that the spread and the depth of contagious defaults depend highly on the connections of each bank and the size of each liability; see Upper (2011) and references therein. However, most data sets available, such as balance sheets reported by banks, only provide the total exposure of each bank to the interbank market and computing individual liabilities requires an estimation procedure.

Gandy and Veraart (2015) estimate liabilities with a Bayesian approach. They assume that each liability follows a exponentially distributed prior distribution, conditionally on a prior probability that the liability exists. As the posterior distribution of liabilities is not available analytically, the authors build a Gibbs sampler to draw samples from the joint distribution of

liabilities conditionally on observed total exposures. Their method enables to study statistical properties about the stability of the unobservesd network by performing the same stress test on different realisations of the liability matrix. In particular, their results show that the structure of the network and the magnitude of default costs determine the likelihood of default of banks affected by contagion.

In this chapter, we extend the approach of Gandy and Veraart (2015) by assuming that liabilities follow general Gamma distributions with different positive modes and asymptotic behaviours at the origin. In effect, we obtain fully heterogeneous networks while keeping tractable prior assumptions. From a Bayesian perspective, it is necessary to develop a model which handles different priors so that additional information about liabilities can be incorporated through a specific prior. Our empirical study shows that there is a strong relation between the influence of the choice of the prior and the size of interbank assets. Furthermore, we provide a new insight on the importance of estimating individual liabilities by identifying in which case the prior has a significant effect on the structure of the network.

Balance sheet contagions are studied by representing the interbank market as a network in which nodes represent banks and directed edges correspond to interbank liabilities. Liabilities are then described in terms of a matrix whose entries are non-negative. The leading method to build a liability matrix, known as the minimum relative entropy method, consists in minimising the Kullback-Leibler divergence with respect to a prior matrix; see Elsinger et al. (2013). The minimisation problem is constrained by the observed total interbank assets and liabilities of each bank. Moreover, additional constraints which can be expressed as linear transformation of the liability matrix are easily incorporated. As this method is flexible with respect to the level of information one would like to put into the optimisation problem and the construction of the liability matrix, it has been used extensively for stress tests on the banking sector of several European countries; see Wells (2004) for Germany, Elsinger et al. (2006) for the UK, Degryse and Nguyen (2007) for Belgium[1].

[1]If no additional information is available, the prior matrix is usually built by dividing the aggregate exposure proportionally over all banks, which results in maximising the number of edge in the network. Such structure is at odds with observed financial networks, which tend to be sparse as reported by Upper and Worms (2004) for the German interbank market and Cocco et al. (2009) for the Portuguese interbank market. For both markets,

Based on a complete data set, including individual bilateral exposures, of the Italian banking sector, Mistrulli (2011) shows that the minimum entropy method underestimates the number of banks affected by contagion. Nevertheless, he also shows that under certain configurations of the network and capitalisation of banks, the minimum entropy method may actually overestimate the contagion effect in the case of very low recovery rates[2].

Based on a complete data set of the Brazilian banking sector, Cont et al. (2013) show that size of total interbank liabilities and global measures of network connectedness do not explain fully the variability of the propensity of contagion. Local measures of the exposure of creditors to a specific bank have a strong predictive power. Hence, local properties of the network have an effect on the overall stability of the network and the estimation of individual liabilities is fundamental to develop realistic stress tests from an incomplete data set.

In Section 5.2, we present the construction of the network and recall the main assumptions on the interbank assets and liabilities, ensuring the existence of a well-structured network. We then introduce the probabilistic model. The choice of a Gamma prior distribution is justified by the flexibility of the shape of distributions, that the model is able to handle.

In Section 5.3, we describe the operation of the Gibbs Sampler. The underlying idea of the Gibbs sampler is to update at each step, a submatrix of the liability matrix conditionally on the row and columns sums. Although the size and the indices are chosen randomly, the elements of the submatrix are selected in a cyclic way so that the multidimensional problem is reduced to the sampling of one single random variable. In Proposition 5.3.1, the distribution of this random variable is characterised by a discrete part and a continuous part. We sample from the continuous part of the random variable using an inverse transform method. In general, the cumulative distribution is not available in closed form and we provide an analytical expression of the cumulative distribution in the case of natural shape and homogeneous rate parameters in Proposition 5.3.2. The cumulative distribution is then

lending is mainly domestic and Cocco et al. (2009) show that lending relationships over time are determinant to access the interbank market.

[2]For the Dutch banking sector, Van and Liedorp (2006) use a cross-entropy minimization problem using additional information such as large data exposure and a bank survey on bilateral exposure they show that that minimum entropy method always results in less banks being affected and in smaller losses.

inverted using an hybrid Bisection-Newton method.

In Section 6.1, we investigate the properties of the conditional posterior distribution based on a data set of 11 German banks. We assume that each liability has the same prior distribution and that liabilities between two banks have the same prior probability of being strictly positive. Furthermore, as the total interbank liability of each bank is not available from the balance sheet data, we assume that it is a perturbation of the interbank assets. In this setting, the size of interbank assets determines the characteristics of the marginal posterior distribution of each liability. For liabilities of small banks, the discrete part of the posterior dominates the continuous part so that small banks tend to be less connected than large banks.

Next, we stress the network by assuming a fixed fall in external assets for all bank. We compute the default probability of each bank suffering of contagious default and the mean loss given default of each defaulting bank. The vector of default probability indicates how widely the external shock propagates through the system while, by quantifying the size of default, the mean loss given default vector measures the depth of the shock. Our results demonstrate that fundamentally insolvent banks have larger default size, and, in line with Rogers and Veraart (2013) and Glasserman and Young (2015), the likelihood of contagious default is increased when significant default costs are introduced. Furthermore our empirical study exhibits that the curvature of default probability increases with the shape parameter for banks affected by a contagious default.

The choice of the prior also influences the local properties of the network by enforcing a larger variability in the connectivity among banks of different assets size. In particular, we observe that the mean out-degree of small banks are the most affected by the choice of the prior by concentrating their exposure on fewer banks. On the contrary, large banks mean out-degrees increase on average with the shape parameter of the prior Gamma distribution.

Finally, we study the structure and the stability of a network consisting of 76 European banks by applying the same sampling and stress test methods. As discussed in Subsection 6.1.2, we significantly increase the thinning of the chain to ensure a good exploration of different regions. Our empirical results demonstrate that increasing the shape parameter of the prior distribution can lead to opposite effects on the default probability for banks of the same

size. The main characteristics of the German network structure carry over to the European network. In particular, the size of interbank asset is a determinant factor for the connectivity of each bank and the influence of the choice of the prior.

5.2 Model Setup

We consider an interbank market of $n \in \mathbb{N}$ banks with indices $\mathcal{N} = \{1, \dots, n\}$. Each bank is represented as a financial firm with a stylised balance sheet of assets and nominal liabilities. Our model is static by nature. In particular the maturity is the same for all liabilities and they are paid instantaneously. We also assume that all liabilities have the same seniority.

We model the market as a network where each node represents a bank. Each directed edge corresponds to the nominal liability of bank i to j and it is weighted by a non-negative cash amount,

$$L_{ij} \geq 0, \forall i \neq j \in \mathcal{N} \text{ and } L_{ii} = 0, \forall i \in \mathcal{N}.$$

The matrix $L = (L_{ij}) \in \mathbb{R}^{n \times n}$ is called the liability matrix. We also consider the binary adjacent matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ which specifies whether there is a liability between two banks,

$$A_{ij} = 1, \text{ if } L_{ij} > 0, \text{ and } A_{ij} = 0, \text{ if } L_{ij} = 0.$$

The total nominal interbank liabilities l_i of bank i is obtained by summing the liabilities of row i, and the total nominal interbank assets a_i is obtained by summing the elements of column i.

For two given vectors a, $l \in \mathbb{R}^n$, a matrix is called admissible if it is a liability matrix such that

$$\sum_{j=1}^{n} L_{ij} = l_i \text{ and } \sum_{j=1}^{n} L_{ji} = a_i, \forall i \in \mathcal{N}.$$

Gandy and Veraart (2015) show that the existence of such matrix is ensured under the necessary and sufficient condition that the total assets of any bank

does not exceed the sum of the total liabilities of all other banks,

$$a_i \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} l_j, \forall i \in \mathbb{N}.$$

Their proof is constructive and it provides an algorithm to build an admissible matrix for given a and l.

Following Gandy and Veraart (2015), we make the technical assumption that there are no isolated subnetworks,

$$\forall I, J, \subseteq \mathcal{N}, \left(\sum_{i \in I} l_i = \sum_{j \in J} a_j \Longrightarrow I = J = \mathcal{N} \text{ or } I = J = \emptyset \right). \tag{5.2.1}$$

In effect, this condition ensures that there cannot be two or more disjoints subnetworks in the main interbank network so that the Markov chain generated by the Gibbs sampler does not get trapped into such subnetworks.

We now describe the underlying probabilistic model. The adjacent matrix $A = (A_{ij})$ is built through a generalised version of the Erdős and Rényi (1959) model. Directed edges between two banks are drawn from independent Bernoulli random variables with off-diagonal probabilities $p_{ij} \in [0, 1]$, $\forall i \neq j \in \mathcal{N}$ and diagonal probabilities $p_{ii} = 0, \forall i \in \mathcal{N}$. Therefore, we have

$$P(A_{ij} = 1) = p_{ij} \text{ and } P(A_{ij} = 0) = 1 - p_{ij}. \tag{5.2.2}$$

If an edge exists, then we assume that the liability L_{ij} follows a Gamma distribution with probability density function with shape parameter α_{ij} and rate parameter λ_{ij},

$$f_{ij}(x) = \frac{\lambda_{ij}^{\alpha_{ij}}}{\Gamma(\alpha_{ij})} x^{\alpha_{ij}-1} e^{-\lambda_{ij} x}, \text{ with } \alpha_{ij} \geq 1 \text{ and } \lambda_{ij} > 0, \forall i \neq j \in \mathcal{N}. \tag{5.2.3}$$

For diagonal entries, all parameters are set to zero. In the special case $\alpha_{ij} \in \mathbb{N}$, the distribution is known as an Erlang distribution and it is of particular importance in Subsection 5.3.2. Note that the case $\alpha_{ij} = 1$ reduces to the exponential distribution.

For each liability, the moment generating function of the density is characterised by these two parameters. In particular, a positive mode exists for $\alpha_{ij} > 1$ and the asymptotic behaviour of the density around the origin is determined by the value of the shape parameter,

$$\text{mode}(f_{ij}) = \frac{\alpha_{ij} - 1}{\lambda_{ij}}, \alpha_{ij} \geq 1. \tag{5.2.4}$$

$$\lim_{x \to 0} f_{ij}(x) = \begin{cases} \lambda_{ij}, & \alpha_{ij} = 1, \\ 0, & \alpha_{ij} > 1. \end{cases}$$

$$\lim_{x \to 0} f'_{ij}(x) = \begin{cases} -\lambda_{ij}^2, & \alpha_{ij} = 1, \\ \infty, & 1 < \alpha_{ij} < 2, \\ 0, & \alpha_{ij} \geq 2. \end{cases}$$

Furthermore, the scaling property of the Gamma distribution allows us to sample indifferently either from L with λ or L/c with matrix $c\lambda$, where c is a constant. Hence, whenever necessary, one can normalise the matrix so that the total sum of interbank assets is equal to one, i.e. $\sum_{i=1}^{n} a_i = 1$.

By the law of total probabilities and the definition of the probabilistic model (5.2.2)-(5.2.3), the unconditional density of $(L_i)_{i \in I}$, with $I \subseteq \mathcal{N} \times \mathcal{N}$, is

$$f(l) = \prod_{i \in I} \left((1 - p_i) \mathbb{I}(l_i = 0) + p_i f_i(l_i) \mathbb{I}(l_i > 0) \right) \tag{5.2.5}$$

$$= \underbrace{\left(\prod_{i \in I_d(l)} (1 - p_i) \right)}_{\text{discrete part}} \underbrace{\left(\prod_{i \in I_c(l)} p_i f_i(l_i) \right)}_{\text{continuous part}}, \tag{5.2.6}$$

where $I_d(l) = \{i \in I | l_i = 0\}$ and $I_c(l) = \{i \in I | l_i > 0\}$ is a partition of I. Hence, the unconditional density has a discrete part for the elements equal to zero, i.e. when there is no liability, and a continuous part for the positive liabilities. As we will show in Section 5.3, the joint conditional distribution is proportional to the unconditional distribution and requires a numerical sampling method.

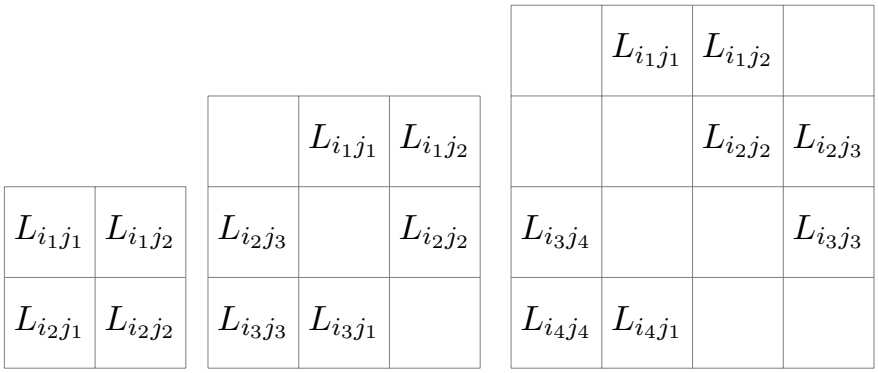


Figure 5.1: Illustration of submatrices for cycles of size $k = 2, 3, 4$; see Gandy and Veraart (2015) Figure 1.

5.3 The Gibbs Sampler

In this section, we build a Gibbs sampler to generate random samples from the joint distribution of L conditionally on the given sum of rows and columns. The Gibbs sampling approach consists in generating an ergodic Markov chain $(L^m)_{m\geq 1} = (L^m_{ij})_{m\geq 1}$ so that the expectation $\mathbb{E}(h(L)|a, l)$ can be approximated by the empirical average

$$\frac{1}{N}\sum_{m=1}^{N} h(L^m),$$

where N is the number of samples.

The key idea developed by Gandy and Veraart (2015) is to sample at each step from a subnetwork built by extracting k rows and k columns from the liability matrix and selecting two elements by row and by column. At each step t, the Markov chain is updated on a subnetwork, while the other elements of the matrix L^t remain unchanged. On such subnetworks, the construction imposed by the constraint of the sum of rows and columns on the subnetwork, leads to a cycle structure of length $2k$ as in Figure 5.1. As we discuss in Subsection 5.3.1, the cycle property reduces the multidimensional sampling problem to a one-dimensional sampling problem. Indeed,

the elements of the cycle are changed by an amount which corresponds to the realisation of a single random variable Δ. The support of Δ is built such that the sum of the rows and columns of the subnetworks remain unchanged, while keeping each entry non-negative.

The length of the cycle is chosen randomly by sampling discretely each value for the set $\{2, \ldots, n\}$ with probability $2^{n-k}/(2^{n-1} - 1)$. In particular, a cycle of size k is chosen approximately with probability $(1/2)^{k-1}$, while, with this choice, probabilities exactly sum up to 1. Indices of the cycle $(i_l, j_l)_{l=1,\ldots,k}$ are then chosen uniformly from the index set $\mathcal{N}$ without replacements.

The choice of the cycle length and indices have a strong influence on the performance of the Gibbs sampler. Indeed, using large cycles tends to change more entries of the liability matrix and thus increases the efficiency to explore different regions of the state space[3]. However, we want also to control for the length of the cycles as sampling from small cycles requires less computation time.

The choice of the indices also matters because the support of Δ is defined by the values of the liabilities in the cycle. For example, if a diagonal element belongs to the cycle, the Gibbs sampler will leave all the elements of the selected cycle unchanged, as the random variable Δ has to be set to zero. Even with only off-diagonal entries, the support of the random variable Δ may include or reduce to zero, leaving the matrix unchanged. If such cases are frequent, sampling efficiently from the target distribution requires a higher number of iterations.

5.3.1 Conditional Distribution

Consider a cycle of length k with indices

$$\eta = ((i_1, j_1), (i_1, j_2), (i_2, j_2), \ldots, (i_k, j_k), (i_k, j_1)).$$

To ensure that the sum of rows and columns of the new matrix $\tilde{L} = L^{m+1}$ and the matrix $L = L^m$ are equal for each step m, we have the following

[3]Gandy and Veraart (2015) show that, with such cycle moves, any configuration can be reached with the Gibbs sampler if the matrix of prior edge probabilities has at most one 0 in each row or column.

constraints on the elements of the cycle

$$\text{(5.3.1)} \qquad \begin{cases} \tilde{L}_{\eta_1} + \tilde{L}_{\eta_2} = L_{\eta_1} + L_{\eta_2}, \\ \tilde{L}_{\eta_2} + \tilde{L}_{\eta_3} = L_{\eta_2} + L_{\eta_3}, \\ \vdots \\ \tilde{L}_{\eta_{2k-1}} + \tilde{L}_{\eta_{2k}} = L_{\eta_{2k-1}} + L_{\eta_{2k}}, \\ \tilde{L}_{\eta_{2k}} + \tilde{L}_{\eta_1} = L_{\eta_{2k}} + L_{\eta_1}. \end{cases}$$

The only matrices $\tilde{L}$ verifying these constraints are of the form

$$\text{(5.3.2)} \qquad \tilde{L}_{\eta_i} = L_{\eta_i} + (-1)^{i+1}\Delta, i = 1, \dots, 2k.$$

Moreover, because of the non-negativity of the liabilities, the random variable Δ must take values in a restricted interval whose bounds are fixed by the minimum over odd and even indices,

$$\Delta \in [\Delta_{low}, \Delta_{up}] = [-\min_{i \text{ odd}} L_{\eta_i}, \min_{i \text{ even}} L_{\eta_i}].$$

Hence, at each step, we must sample from the random variable Δ with support in the interval $[\Delta_{low}, \Delta_{up}]$ while verifying constraints (5.3.1).

For ease of the notation, we now denote the indices by i instead of η_i for the fixed cycle η. Let us consider the mapping $g : \mathbb{R}^{2k} \to \mathbb{R}^{2k}$ defined by

$$g_1(x) = x_1 - L_1 \text{ and } g_i(x) = x_{i-1} + x_i, \text{ for } i = 2, \dots, 2k.$$

As Δ is simply the difference of $\tilde{L}_1$ and L_1, $\Delta = \tilde{L}_1 - L_1$, each constraint in (5.3.1) is defined recursively with

$$\text{(5.3.3)} \qquad g_i(L) = g_i(\tilde{L}), \text{ for } i = 2, \dots, 2k.$$

Thus, we are interested in sampling from the density of Δ conditionally on (5.3.3),

$$\text{(5.3.4)} \qquad f_\Delta(\cdot | g_i(\tilde{L}) = g_i(L), i = 2, \dots, 2k).$$

Denoting $C = \{g_i(\tilde{L}) = g_i(L), i = 2, \dots, 2k\}$, we write $f_\Delta(\cdot|C)$.

Let us define the number of elements which are on the boundary of the

support of Δ,

$$n^0_{up} = \#\{i \text{ even} : L_{\eta_i} = \Delta_{up}\} \text{ and } n^0_{low} = \#\{i \text{ odd} : L_{\eta_i} = \Delta_{low}\}.$$

To sample from the conditional law of Δ , we have four distinct settings

(a) $n^0_{low} \leq 1$ and $n^0_{up} \leq 1$,

(b) $n^0_{low} > 1$ and $n^0_{low} < n^0_{up}$,

(c) $n^0_{up} > 1$ and $n^0_{up} < n^0_{low}$,

(d) $n^0_{up} > 1$ and $n^0_{low} = n^0_{up}$.

We now show that the probability contribution of each boundary $\Delta \in \{\Delta_{low}, \Delta_{up}\}$ and the intermediate case $\Delta \in (\Delta_{low}, \Delta_{up})$ are of same order for case (a), and that the former dominates the latter in cases (b)-(d).

The arguments derived in Proposition 5.3.1 were first presented in Gandy and Veraart (2015) for exponential distributions. It relies on the asymptotic behaviour (5.3.9) of the cumulative distributions functions (CDF), which actually holds for any continuous distribution.Therefore, the dominance effect does not depend on the type of distributions considered.

Proposition 5.3.1. *Let*

$$l_i(\delta) = L_i + (-1)^{(i+1)}\delta, i = 1, \ldots, 2k.$$

Assume that Δ_{low} and Δ_{up} are different. If $\delta \in \{\Delta_{low}, \Delta_{up}\}$,

(5.3.5)
$$\mathbb{P}\left(\tilde{L}_i \in (l_i(\delta) - \epsilon, l_i(\delta) + \epsilon), i = 1, \ldots, 2k\right) = \mathcal{O}\left(\epsilon^{\#\{i:l_i(\delta)\neq 0\}}\right), \text{ as } \epsilon \to 0,$$

with

(5.3.6)
$$\#\{i : l_i(\Delta_{up}) \neq 0\} = 2k - n^0_{up} \text{ and } \#\{i : l_i(\Delta_{low}) \neq 0\} = 2k - n^0_{low},$$

If $\delta \in (\Delta_{low}, \Delta_{up})$,

(5.3.7)
$$\mathbb{P}\left(\tilde{L}_i = (l_i(\delta) - \epsilon, l_i(\delta) + \epsilon), i = 1, \ldots, 2k\right) = \mathcal{O}\left(\epsilon^{2k-1}\right), \text{ as } \epsilon \to 0.$$

Proof. By (5.2.5), the unconditional density of $\tilde{L}$ is

$$f(l) = \prod_{i=1}^{2k} \left((1-p_i)\mathbb{I}(l_i = 0) + p_i f_i(l_i)\mathbb{I}(l_i > 0)\right), \tag{5.3.8}$$

where f_i is the density function of liability $\tilde{L}_i$, $i = 1, \ldots, 2k$. Moreover, for each $i = 1, \ldots, 2k$, the marginal probability of being in ϵ-neighborhood of l_i for each $\tilde{L}_i$ is, by a first order approximation of the CDFs F_i,

$$P((l_i - \epsilon)^+ < \tilde{L}_i < l_i + \epsilon) = F_i(l_i + \epsilon) - F_i((l_i - \epsilon)^+) = \mathcal{O}(\epsilon), \text{ as } \epsilon \to 0. \tag{5.3.9}$$

We suppose that $\delta \in \{\Delta_{low}, \Delta_{up}\}$. Then at least one of the l_i is equal to 0. We now compute the joint probability of all $\tilde{L}_i$ being in the ϵ-neighborhood of l_i is given as

$$\begin{aligned}
&\mathbb{P}\left(\tilde{L}_i \in (l_i(\delta) - \epsilon, l_i(\delta) + \epsilon), i = 1, \ldots, 2k\right) \\
&= \prod_{i=1}^{2k} \left((1-p_i)\mathbb{I}(0 \in (l_i(\delta) - \epsilon, l_i(\delta) + \epsilon)) + p_i \int_{(l_i(\delta)-\epsilon)^+}^{l_i(\delta)+\epsilon} f_i(y)dy\right) \\
&= \prod_{i=1}^{2k} \left((1-p_i)\mathbb{I}(0 \in (l_i(\delta) - \epsilon, l_i(\delta) + \epsilon)) + p_i \left(F_i(l_i(\delta) + \epsilon) - F_i((l_i(\delta) - \epsilon)^+)\right)\right) \\
&= \mathcal{O}\left(\epsilon^{\#\{i: l_i(\delta) \neq 0\}}\right), \text{ as } \epsilon \to 0.
\end{aligned}$$

For $\delta \in (\Delta_{low}, \Delta_{up})$, all $l_i(\delta)$ are strictly positive since Δ_{low} and Δ_{up} are different, and

$$\mathbb{P}\left(\tilde{L}_i = (l_i(\delta) - \epsilon, l_i(\delta) + \epsilon), i = 1, \ldots, 2k\right) \tag{5.3.10}$$

$$= \mathbb{P}\left(\tilde{L}_1 = (L_1 + \Delta_{low}, L_1 + \Delta_{up}), \tilde{L}_i \in B_i(\delta, \epsilon), i = 2, \ldots, 2k\right), \tag{5.3.11}$$

where $B_i(\delta, \epsilon) = \left(\max(L_i + (-1)^{i+1}\delta - \epsilon, 0), L_i + (-1)^{i+1}\delta + \epsilon\right)$. Because $\Delta = \tilde{L}_1 - L_1$, the probability in (5.3.11) can be rewritten as an integral with

respect to continuous part of the density of $\tilde{L}_1$,

$$\begin{aligned}
&\int_{L_1+\Delta_{low}}^{L_1+\Delta_{up}} \left(\prod_{i=2}^{2k} \int_{B_i(u_1-L_1,\epsilon)} f_i(u_i) du_i \right) f_1(u_1) du_1, \\
&= \int_{L_1+\Delta_{low}}^{L_1+\Delta_{up}} \left(\prod_{i=2}^{2k} F_i(l_i(u_1-L_1)+\epsilon) - F_i(l_i(u_1-L_1)-\epsilon) \right) f_1(u_1) du_1, \\
&= \mathcal{O}\left(\epsilon^{2k-1}\right), \text{ as } \epsilon \to 0.
\end{aligned}$$

□

Comparing (5.3.5) and (5.3.7), it follows that the asymptotic behaviour of the intermediate case is dominated by the boundary case if there are two or more zeros in the matrix L, while they are of the same order otherwise. Hence, in setting (a), each case is sampled with respect to their probability of occurrence as discussed in the following subsection. For (b)-(c), the probability of being on the boundaries dominates and the boundary corresponding to the largest number of zeros is sampled with probability one. Finally, in setting (d), each boundary case is sampled with a probability proportional to the value of $f_\Delta(\delta|C)$, with $\delta \in \{\Delta_{low}, \Delta_{up}\}$.

5.3.2 Sampling from the Conditional Distribution

The conditional density of Δ, $f_\Delta(\delta|C)$, as defined in (5.3.4) is proportional to the joint density of $(\Delta, g_2(\tilde{L}), \ldots, g_{2k}(\tilde{L}))$ evaluated at the point $(\delta, g_2(L), \ldots, g_{2k}(L))$, which, by the change of variable formula for density functions, is given by

$$f_{(\Delta, g_2(\tilde{L}), \ldots, g_{2k}(\tilde{L}))}(g^{-1}(\delta, L_1+L_2, \ldots, L_{2k-1}+L_{2k})) = f\left(L + s\delta\right). \tag{5.3.12}$$

Let

$$\kappa = \int_{(\Delta_{low}, \Delta_{up})} f(L+s\delta) d\delta, \text{ and } \xi = \frac{1}{f(L+s\Delta_{low}) + \kappa + f(L+s\Delta_{up})}. \tag{5.3.13}$$

Then, the probability of hitting one of the boundaries is $\xi f(L+s\delta)$ for $\delta \in \{\Delta_{low}, \Delta_{up}\}$, while the probability of the intermediate case, $\xi\kappa$, is obtained by integrating the continuous part of density with respect to δ on the open

interval $(\Delta_{low}, \Delta_{up})$.

For each boundary case, sampling from the the conditional distribution of Δ is straightforward as the outcome reduces to a singleton. For the intermediate case on the interval $(\Delta_{low}, \Delta_{up})$, the conditional density is

$$f_\Delta(\delta|C) = \frac{1}{\kappa} \prod_{i=1}^{2k} p_i f_i(L_i + s_i\delta), \tag{5.3.14}$$

where the constant κ defined in (5.3.13) is the normalising constant.

To sample from the continuous part of the conditional density in (5.3.14), we use an inverse transform method; see Glasserman (2004) for background reading. By (5.3.14), the CDF, $F_\Delta(\cdot|C)$, is the integral of a product of density functions. Note that the constant κ also requires the integration of the density. For the exponential case, $\alpha_i = 1$, Gandy and Veraart (2015) show that the conditional density is proportional to an extended exponential in the sense that $\tilde{\lambda}$ may be negative,

(5.3.15)
$$f_\Delta(\delta|C) \propto \exp(-\tilde{\lambda}s\delta) \text{ and } F_\Delta(\delta|C) = \frac{1}{\kappa\tilde{\lambda}} \left(\exp(-\tilde{\lambda}\Delta_{low}) - \exp(-\tilde{\lambda}\delta)\right).$$

The inverse is then computed analytically[4],

$$F_\Delta^{-1}(u|C) = -\frac{1}{\tilde{\lambda}} \log\left(\exp(-\tilde{\lambda}\Delta_{low}) - \kappa\tilde{\lambda}u\right), \text{ for } \alpha_i = 1, i, \ldots, 2k.$$

Note this holds true for any values of $\tilde{\lambda}$.

As we show below, if all the liabilities follow an Erlang distribution with homogeneous rate parameters, the conditional CDF is known analytically. Otherwise a numerical integration of the density function $f_\Delta(\cdot|C)$ is required. The numerical integration is performed with the QAG algorithm of the QUADPACK package; see Piessens et al. (1983). The QAG algorithm is an adaptive quadrature based on 21 point Gauss-Kronrod integration rules. Next, we invert the CDF by solving the following equation,

$$F_\Delta(\delta|C) - u = 0,$$

where u is a sample of the uniform distribution on the interval (0,1). We solve the equation with a root finding Bisection-Newton method. Our hy-

brid algorithm computes the root with the Bisection method whenever the Newton algorithm does not converge in a given number of steps. Indeed, depending on the size on the cycle, the Newton algorithm may not converge for small or large of realisations of the random variable u, i.e $u \sim 0$ or $u \sim 1$. Since the conditional CDF is flat at the boundaries, the slope of its tangent is close to zero and the algorithm jumps from one boundary to the other. Because of the choice of the shape and rate parameters in (6.1.6), this situation happens more frequently as the shape parameter increases. Nevertheless, the implementation is eased by the fact that the random variable Δ takes values on an interval of finite length $(\Delta_{low}, \Delta_{up})$.

Let $\tilde{\lambda} = \sum_{i=1}^{2k} \lambda_i s_i$, the conditional density is proportional to

$$f_\Delta(\delta|C) \propto f(L_\eta + s\delta) \propto \prod_{i=1}^{2k} p_i (L_i + s_i\delta)^{\alpha_i - 1} \exp(-\lambda_i (L_i + s_i\delta)) \tag{5.3.16}$$

$$\propto \exp(-\tilde{\lambda}\delta) \prod_{i=1}^{2k} (L_i + s_i\delta)^{\alpha_i - 1}. \tag{5.3.17}$$

If the rate parameters are such that $\tilde{\lambda} = 0$, the conditional density is proportional to the product,

$$\prod_{i=1}^{2k} (L_i + s_i\delta)^{\alpha_i - 1}.$$

Therefore the density is characterised only by the shape parameter. In particular for the exponential case, this leads to sampling from the uniform distribution. From this aspect, using Gamma functions allows us to sample from uni-modals distributions even when the rate parameters are canceled out. Indeed, for natural shape parameters $\alpha_i \in \mathbb{N}, i = 1, \ldots, 2k$, the distribution is a polynomial of which the two roots closest to zero define the support of Δ.

This observation leads to the following proposition for Erlang type distributions.

Proposition 5.3.2. *Suppose that, on the cycle η of length $2k$, the liabilities follow an Erlang distribution with $\alpha_i - 1 = n_i \in \mathbb{N}$, $i = 1, \ldots, 2k$, and let the rate matrix be such that $\tilde{\lambda} = 0$. For each liability L_i in the the cycle, we*

define the factors

$$B_{q,i} = \binom{n_i}{q} L_i^{n_i - q} s_i^q,\ q = 0, \ldots, n_i.$$

Then, the conditional CDF of Δ is the polynomial[4]

$$F_\Delta(\delta|C) = \frac{1}{\kappa} \sum_{q=0}^{n_1+\ldots+n_{2k}} A_{q,2k}(\delta^{q+1} - \Delta_{low}^{q+1}), \tag{5.3.18}$$

where $A_{q,2k}$ is defined by recurrence as follows: For $i = 1$ and $q = 0, \ldots, n_1$, we set

$$A_{q,1} = B_{q,1}.$$

For $i = 2, \ldots, 2k$ and $q = 0, \ldots, n_1 + \ldots + n_i$, $A_{q,i}$ is defined through the product of two polynomials of degree $n_1 + \ldots + n_{i-1}$ and n_i,

$$\left(\sum_{q=0}^{n_1+\ldots+n_{i-1}} A_{q,i-1}\delta^q \right) \cdot \left(\sum_{r=0}^{n_i} B_{r,i}\delta^r \right) = \sum_{q=0}^{n_1+\ldots+n_i} \underbrace{\left(\sum_{r=0}^{q} A_{r,i-1}B_{q-r,i} \right)}_{=A_{q,i}} \delta^q. \tag{5.3.19}$$

Proof. By the Newton Binomial formula

$$\prod_{i=1}^{2k} (L_i + s_i\delta)^{n_i} = \prod_{i=1}^{2k} \sum_{q=0}^{n_i} \binom{n_i}{q} L_i^{n_i-q} (s_i\delta)^q = \prod_{i=1}^{2k} \sum_{q=0}^{n_i} B_{q,i}\delta^q.$$

The formula (5.3.19) is obtained by computing recursively the product of the polynomial obtained at step $i-1$ with the polynomial of power n_i corresponding to liability L_i. Finally, the expression of the CDF follows by integrating the polynomial on the interval (Δ_{low}, δ). □

[4] Because the density is computed by proportionality in (5.3.16), κ is not the normalising constant defined in (5.3.13). It is defined such that $F_\Delta(\Delta_{up}^-|C) - F_\Delta(\Delta_{down}^+|C) = 1$.

Chapter 6

Empirical Example

6.1 Stress Test Setup

6.1.1 Balance Sheet Contagion Model

The European Bank Authority (EBA) carried wide stress tests in 2009, 2010, 2011 and 2014. In Section 6.2 and 6.3, we use the balance sheet information of banks provided by the 2011 EBA report to build the initial liability matrix and choose the different parameters of the model.

The EBA report provides the Tier 1 capital, the total assets, and the exposure at default (EAD) of each bank i. Following Upper and Worms (2004), we close the interbank network by assuming that the EAD of each bank is held by other banks in the network. Therefore, by definition of the liability matrix, the EAD of bank i corresponds to the interbank assets a_i. The total assets are the sum of the interbank assets a_i and the external assets $a_i^{(e)}$, e.g., loans to non-financial companies and financial companies not included in the network.

The total interbank liabilities l_i are not observed. While a simple choice is to set $l_i = a_i$ for each bank as in Glasserman and Young (2015) and Chen et al. (2014), this assumption violates the condition (5.2.1) of no isolated subnetwork in our setting. As in Gandy and Veraart (2015), we consider interbank liabilities by altering total observed interbank assets with inde-

pendent perturbations,

$$l_i = \mathbf{round}\left(a_i + \epsilon_i \frac{\sum_{j=1}^{n} a_j}{\sum_{j=1}^{n}(a_j + \epsilon_j)}\right), i = 1, \ldots, n-1, \tag{6.1.1}$$

$$l_n = \sum_{j=1}^{n} a_j - \sum_{j=1}^{n-1} l_j, \tag{6.1.2}$$

where $\mathbf{round}(\cdot)$ is the rounding function to 1 decimal place and $(\epsilon_i)_{i=1}^{n-1}$ are identically independent distributed normal random variable with mean 0 and standard deviation 100.

Once we have obtained N samples of the liability matrix with the Gibbs sampler, we stress the interbank market by applying a deterministic shock to the external assets $a_i^{(e)}$ of every bank. We assume that the external assets fall by a fixed percentage for all banks, i.e the new external assets vector is $sa^{(e)}$ with $s < 1$. For example, one can think of this fall as the result of an external liquidity shock implying that borrowing firms, which are not part of the interbank system, fail to pay back fully their loans.

The book value of equity, or net worth, is then given by

$$w_i = a_i^{all} - l_i^{all},$$

where $a_i^{all} = s_i a_i^{(e)} + \sum_{j=1}^{n} L_{ji}$ and $l_i^{all} = \sum_{j=1}^{n} L_{ij} + l_i^{(e)}$ are the actual total assets and liabilities of bank i. If the net worth is non-negative, the bank is solvent, otherwise it is in default. When stressing the system, there are two types of default, namely fundamental and contagious. A fundamental default occurs as bank i net worth becomes negative because of the fall of its external assets even if all banks are able to pay their obligations. A contagious default occurs as the total interbank assets of another bank j are reduced because of the default of bank i, so that the net worth of bank j, w_j, becomes negative. In essence, contagious defaults reveal the presence of systemic risk in the interbank network.

The balance sheet contagion effect is modeled with the Eisenberg and Noe (2001) clearing system. In this model, a defaulting bank i pays each creditor j in the interbank system by a proportion Π_{ij} of the original nominal

liability L_{ij},

$$\Pi_{ij} = \begin{cases} L_{ij}/l_i^{all}, & \text{if } l_i^{all} > 0, \\ 0, & \text{if } l_i^{all} = 0. \end{cases}$$

To meet these obligations, the bank in default uses all its assets and equity, but it does not pay more than the total liabilities l_i^{all}. In effect, each bank pays the minimum between what it is able to pay, which is determined by the amount it receives from other banks, and what it is due to pay.

For a shock $s \in [0, 1]$, the vector of payments clearing the system, $p^*(s)$, is obtained as a fixed point of the mapping

(6.1.3)

$$\Phi(p(s))_i = \begin{cases} l_i^{all} & \text{if } l_i^{all} \leq \sum_{j=1}^n \Pi_{ji} p_j(s) + s_i a_i^{(e)}, \quad (1) \\ \sum_{j=1}^n \Pi_{ji} p_j(s) + d s a_i^{(e)} & \text{else }, \quad (2) \end{cases}$$

where $d \in (0, 1)$ is the discount price at which a bank in default is forced to sell its external assets [1]. Rogers and Veraart (2013) prove the existence of such clearing vectors. Note that clearing vectors are not necessarily unique.

With default costs, we compute the maximal clearing vector with the Greatest Clearing Vector Algorithm (GA) developed in Rogers and Veraart (2013). Starting from the nominal liability matrix, the GA algorithm builds a decreasing sequence of clearing vectors converging to the greatest clearing vector in at most n steps. At each step, the algorithm computes the set of solvent banks by assessing for which banks case (1) of (6.1.3) holds true. The clearing vectors of solvent banks are then set to their nominal obligations, while the clearing vectors of insolvent banks are computed by solving the linear system given in case (2) of (6.1.3). Checking condition (1), the algorithm stops when the set of insolvent banks remains identical between two consecutive steps. Therefore, the GA algorithm has the natural interpretation of starting from the nominal liability matrix and identifying, at each step, which banks become insolvent because of the default of other banks in the system.

In the absence of default costs, $d = 1$, Eisenberg and Noe (2001) show the fixed point problem of (6.1.3) is equivalent to any maximisation problem

[1] A second type of bankruptcy cost is that banks recover only a fraction of their interbank assets in case of default.

with a strictly increasing objective function under the constraint that the clearing vector does not exceed the total inflows, i.e., for each bank i, $p_i \leq \sum_{j=1}^{n} \Pi_{ji} p_j(s) + s_i a_i^{(e)}$. Therefore, we use a linear program to compute the clearing vectors.

For each bank i affected by contagious default, we estimate the probability of a contagious default with the average of the occurrence of default $\mathcal{D}_i^t \in \{0, 1\}$,

$$\hat{P}(\text{ Bank } i \text{ defaults}) = \frac{1}{N} \sum_{t=1}^{N} \mathcal{D}_i^t. \tag{6.1.4}$$

Note that fundamentally insolvent banks' defaults for each sample L^t of the liability matrix regardless of the individual values of L^t and the default costs. Moreover, for each bank i, we define the set of samples for which the bank defaults,

$$\mathbb{D}_i = \{t \in \{1, \ldots, N\} | \, \mathcal{D}_i^t = 1\}.$$

Then, for each defaulting bank, we compute the Mean Loss Given Default (MLGD),

$$\text{MLGD} = 1 - \frac{1}{\#\mathbb{D}_i} \sum_{t \in \mathbb{D}_i} \frac{p_i^t(s)}{l_i^{all}}, \tag{6.1.5}$$

where $\frac{p_i^t(s)}{l_i^{all}}$ is the recovery rate of bank i for sample t. For non-defaulting banks, the MLGD is zero, although their net worth is actually reduced as a result of other banks' defaults[2].

6.1.2 Choice of the Parameters

Based on the 2011 EBA report, we study the stability and the structure of two networks. In Section 6.2, the first network consists of $n = 11$ German banks with balance sheet information listed in Table D.1; see Appendix D.

[2]To measure the reduction of a bank equity, one should compute the difference (in net worth) between the original net worth assuming all banks are able to pay their nominal liabilities and the net worth resulting from the clearing vectors. Because of limited liability, the value of equity should be set to zero for banks in default; see Rogers and Veraart (2013), Definition 2.9. Therefore, the maximal loss is of 100% and this measure does not differentiate defaulting banks.

In Section 6.3, the second network consists of $n = 76$ European banks[3] with balance sheet information listed in Table E.1; see Appendix E.

For the German and the European networks, we apply on each system a fall in external assets by setting $s = 0.97$, respectively $s = 0.96$, and default costs with $d = 0.95$, respectively $d = 0.96$. To study the properties of the default probability estimator (6.1.4), the magnitude of the shock is chosen to avoid a binary situation where the default probabilities are either 0 or 1 for all banks.

We now discuss the choice of the parameters for the Gibbs sampler. We assume that both networks are homogeneous with parameters $p_{ij} = p$, $\alpha_{ij} = \alpha$, and $\lambda_{ij} = \lambda, \forall i \neq j \in \mathcal{N}$. As $2n - 1$ strictly positive constraints need to be verified, at least $2n-1$ non-diagonal entries of the liability matrix are required to be strictly positive. As such, we choose the probability of two banks being connected to be larger that the number of constraints divided by the number of non-diagonal entries, i.e., $p \geq (2n - 1)/(n^2 - n)$. The parameters α and λ are chosen such that the expected value of the sum of the entries equals to the observed total amount of interbank assets or equivalently to the total amount of interbank liabilities. Let $A = \sum_{i=i}^{n} a_i$, we have

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} L_{ij}\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\frac{\alpha_{ij}}{\lambda_{ij}} = n(n-1)p\frac{\alpha}{\lambda}.$$

Hence, for a given shape paramater α, we choose the rate parameter λ such that

$$\lambda = \frac{pn(n-1)}{A}\alpha. \tag{6.1.6}$$

Because of numerical considerations, simulations show that we are also constrained on the choice of the shape parameter. In the case of the European network, the normalisation constant of the conditional CDF is extremely small on certain cycles for large shape parameters, and, as computer precision is reached, it is rounded down to the value 0. In such cases, we simply skip the step in the chain and move forward. With the choice of our parame-

[3]Following Glasserman and Young (2015), we remove, the ten smallest banks, in terms of total assets, from the original EBA data set of 90 European banks. For some small banks, balance sheet data are note reliable as reported total assets are smaller than interbank assets. Countries with a single reported bank are also excluded.

ters, this happens very infrequently. For example, in the European network with $\alpha = 5$ and $p = 0.9$, we reject $3 \cdot 10^{-3}$ % of the samples requiring the normalisation constant to be computed. Nevertheless, the situation becomes more problematic as both the shape parameter and the edge probability increase. Although these numerical issues show the limitation of the inverse transform method introduced in Subsection 5.3.1, using a large shape parameter for the prior distribution is hardly justifiable. Indeed, for the choice of parameters (6.1.6), the variance of the priori decreases sharply to 0 as the shape parameter increases[4].

In our setting, we also need to carefully choose the number of samples computed with the Gibbs sampler. Indeed, samples between two steps in the chain are highly correlated because the Gibbs sampler updates smaller cycles with a higher probability. In the large European network, the proportion of entries changed at each step is small and autocorrelation is especially strong. Therefore, we want to go deep into the chain to ensure that the whole space is visited. As we cannot keep all the samples in memory, we need to thin the chain by keeping only every M-th sample. Moreover, it would be too time consuming to keep all the samples in order to compute the estimators (6.1.4) and (6.1.5). Although this method increases the variance of the empirical average, using thinning is justified because we do not lose too much information by throwing away strongly correlated samples.

Unlike a standard Monte-Carlo method, a central limit theorem for the chain generated by the Gibbs sampler is not established and we are not able to formally build confidence intervals for estimators (6.1.4) and (6.1.5). In particular, the exact rate of convergence of the algorithm is not known. The following parameters ensure consistent results when sampling from different seeds[5]. To control the dependency on the initial liability matrix, we throw away the first B samples. This is known as the burn-in period. Finally, the number of samples for the Gibbs sampler is chosen to be $N = 10000$, with a burning period $B = 10000$, respectively 50000, and we thin the chain by keeping every $M = 5000th$, respectively 20000th, sample for the German network, respectively for the European network[6].

[4]Let $\kappa = \frac{pn(n-1)}{A}$, the mean of the prior is $\frac{1}{\kappa}$ and its variance is equal to $\frac{1}{\kappa^2\alpha}$. Therefore the prior distribution degenerates to a Dirac measure localised at $\frac{1}{\kappa}$, as $\alpha \to \infty$.

[5]For a detailed discussion of convergence diagnosis for the Gibbs sampler; see Robert and Casella (2004).

[6]The initial matrix of the Markov Chain, verifying the given sum of the rows and

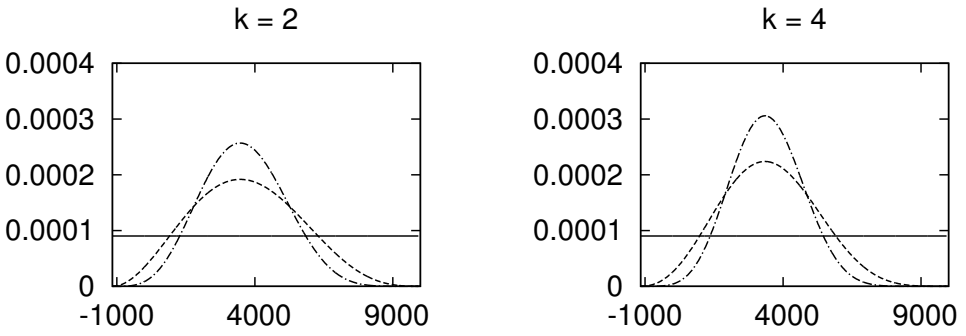


Figure 6.1: Plot of the continuous part of the density of the random variable Δ on two cycles of different sizes. The cycles are chosen such that the support of Δ is the same on both cycles. Parameters: Sample size = 10000, edge probability $p = 0.5$, shape parameter $\alpha = 1$ (full), $\alpha = 3$ (dashed) and $\alpha = 5$ (dashed/dotted).

6.2 Small Network

6.2.1 Properties of the Conditional Distribution

In this subsection, we investigate the properties of the conditional distribution of the random variable Δ and individual liabilities L_{ij} based on the German network data.

Figure 6.1 depicts the profile of the conditional density of Δ on two specific cycles of size $k = 2$ and $k = 4$. The cycles are chosen so that the support of Δ is the same on each cycle. The scale parameters are all set equal to $\alpha = 1$, $\alpha = 3$ or $\alpha = 5$ and and the rate parameters are defined as in (6.1.6). For the exponential case $\alpha = 1$, the assumption of homogeneity on the parameters implies that the random variable Δ is uniformly distributed. On the contrary, for homogeneous Erlang priors, the density is given as a polynomial whose degree increases with the shape parameter and the size of the cycle; see (5.3.18). Therefore, the conditional density becomes more peaked and its variance decreases. Similarly to the prior, the conditional density flattens out at the boundaries. By relation (5.3.2), sampling from a liability is tantamount to sampling from Δ. However, the law of Δ is different for each selected cycle and there is no direct interpretation of the properties of the conditional density of a liability through the density of Δ.

columns, is built with algorithm 2 in Gandy and Veraart (2015).

While each entry has the same unconditional prior distribution, the support of the conditional posterior density of each off-diagonal liability L_{ij} is the closed interval between 0 and the minimum of the total interbank liability l_i and the total interbank assets a_j. For the lower bound 0, we know that the value zero is sampled from the boundary case of $\Delta = \Delta_{low}$. Note that, at a specific step in the chain, the liability is not necessarily in the cycle. Still, it may be equal to zero because it was set to zero the last time it was part of a cycle. Nevertheless the frequency at which a liability equals zero measures the dominance of the discrete part of the posterior density.

In Table 6.1, we report properties of the distribution of nine liabilities. We consider three categories of banks classified by size of interbank assets a_i: the first tercile corresponds to small banks with interbank assets $a_i \leq 28000$, the second tercile to medium banks with interbank assets $28000 < a_i \leq 53000$, and the third tercile to large banks with $a_i > 53000$.

In our setting, the effect of the size is particularly strong across the three categories. The probability of hitting zero for a liability between two banks of the same category decreases with the size of interbank assets a_i. Liabilities of banks with small, medium and respectively large, interbank assets are equal to zero with probability 0.94, 0.45, and respectively 0.01. For liabilities of the type small-large and small-medium, the probability of liabilities being zero also dominates with value 0.88 and 0.82 respectively.

For liabilities between two banks of the same category, the distributions of the liability L_{ij} and L_{ji} are similar. This symmetry property between the distribution of liability L_{ij} and liability L_{ji} extends to liabilities between banks of different sizes. The distribution of liability L_{ij} shares the same characteristics as the distribution of L_{ji} both on their discrete and continuous parts, as by (6.1.1), the aggregate amounts of liability and assets are close to each other. Figure 6.2 illustrates the property of symmetry of the continuous part of the distribution between banks of different sizes, with L_{13} corresponding to the liability of bank DE017, with small interbank assets, to bank DE019, with large interbank assets. Figure 6.2 also confirms that the continuous part of the density become more peaked as the shape parameter α increases.

,

Liability	Size	Density function	Discrete Part	Continuous Part	
L_{ij}	Relation	Support	$\hat{P}(L_{ij}=0)$	Mean	Std. Dev. (%)
DE023-25	S-S	[0, 4645]	0.94	3259	44
DE017-18	M-M	[0, 46989]	0.45	8703	52
DE019-20	L-L	[0, 91314]	0.01	25360	41
DE017-25	M-S	[0, 4645]	0.88	3777	33
DE025-17	S-M	[0, 4841]	0.88	3923	33
DE019-25	L-S	[0, 4645]	0.82	3966	29
DE025-19	S-L	[0, 4841]	0.83	4107	30
DE017-19	M-L	[0, 46989]	0.20	12430	49
DE019-17	L-M	[0, 47102]	0.19	12380	50

Table 6.1: Characteristics of the liabilities distribution. The first column corresponds to the liability L_{ij}. The second columns indicates size of inter-bank assets for bank i and bank j, L = Large, M = Medium, S = Small. The third columns corresponds to the support of the distribution of L_{ij}. The fourth column reports the empirical probability of hitting the boundary case zero. The fourth and fifth columns report the mean and the relative standard deviation of the sample. Parameters: Sample size = 10000, edge probability $p = 0.5$, shape parameter $\alpha = 3$.

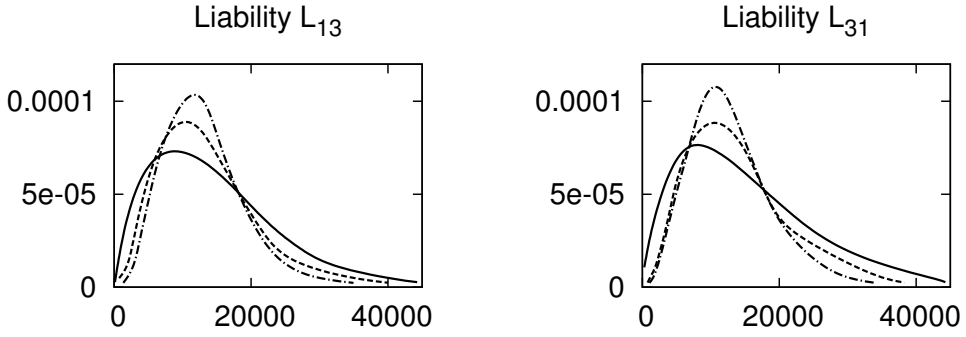


Figure 6.2: Plot of the continuous part of the density of liabilities L_{13} and L_{31} for different values of the shape parameter α.Parameters: Sample size = 10000, edge probability $p = 0.5$, $\alpha = 2$ (full) , $\alpha = 3.5$ (dashed), $\alpha = 5$. (dashed/dotted).

Bank Code	Bank Size	CTA Ratio(%)	Mean degree in	Mean degree out	Default without cost MLGD (%)	Default without cost Proba	Default with cost MLGD (%)	Default with cost Proba
Fundamental default								
DE017	M	1.59	5.05	5.02	1.36	1	6.23	1
DE022	M	1.74	5.29	5.28	0.60	1	5.04	1
DE023	S	1.69	2.83	2.84	1.27	1	6.15	1
DE024	S	2.20	4.14	4.13	0.46	1	5.17	1
Contagious default								
DE019	L	2.62	5.96	5.95	0	0	4.02	0.93
DE020	L	2.26	6.19	6.20	$1.13{\cdot}10^{-3}$	0.03	4.18	0.96
DE025	S	2.94	2.32	2.35	$4.64{\cdot}10^{-3}$	0.09	4.94	0.82
DE028	M	2.58	4.48	4.47	$3.90{\cdot}10^{-5}$	0.002	4.00	0.90
No default								
DE018	L	3.47	5.15	5.16				
DE021	L	3.64	5.55	5.57				
DE027	S	3.86	4.34	4.33				

Table 6.2: Characteristics of the banks in the German network. The third columns corresponds to the capital to total assets (CTA) ratio. The fourth and fifth columns represent the mean number of interbank liabilities and assets respectively. Parameters: Sample size = 10000, edge probability $p = 0.5$, shape parameter $\alpha = 1$, shock $s = 0.97$, default cost $d = 0.95$.

6.2.2 Stress Test

To understand the likelihood of default in the case of no default costs, we have to take into account the capital to total assets ratio (CTA) and the connectivity to the network of the bank. In Table 6.2, Bank DE025 has the largest CTA among defaulting banks. Nevertheless it has the largest default probability because it is exposed to 2.32 banks on average and its mean exposure to fundamentally insolvent banks represents 99% of its total exposure[7]. Furthermore, bank DE020 is well connected and it has the smallest CTA ratio among banks affected by contagious default. As such its equity buffer is not able to absorb external shocks. In particular, it has a higher probability of default compared to Bank DE019, which has a higher CTA, although both banks share the same characteristics in terms of connectivity and size of interbank assets. Hence, the connectivity helps to reduce the exposure to banks in default, as long as there is enough capital to absorb the reduction in expected payments from a single default.

When default costs are incorporated, the propagation of defaults becomes wider and deeper. The loss of 5% on the sell of external assets in-

[7]Note that if a bank is exposed mainly to defaulting banks, connectivity may actually increase its probability of default.

creases by one thousand fold the size of default and by ten fold the default probability of banks suffering from contagious default. Compared to the no cost case, these larger LGDs result into an additional bank exposed to a high default probability. As only banks with proportionally large equity buffer are able to cover their losses, the CTA is a good indicator of resilience. Fundamentally insolvent banks have a CTA between 1.59% and 2.20%. Banks affected by contagious default have a CTA between 2.26% and 2.94%, and non-defaulting banks have a CTA between 3.47% and 3.86%.

We now discuss the effect of the shape parameter on the default probability for three different shape parameters, $\alpha = 1, 3, 5$. Figure 6.3 shows that, with no default costs, the default probability decreases with the connectedness of the network. As the network becomes more connected, the losses are evenly spread through the system and bank net worths are large enough to absorb the external shock. Furthermore, bank DE025 has the largest default probability across all shape parameters. Similarly to its mean out-degrees, Bank DE025 probability of default becomes concave as the shape parameter increases. For $\alpha = 5$, the default probability is essentially flat on the interval $p \in [0.2, 0.5]$, it is then reduced significantly as the number of out-degrees, or equivalently of in-degrees, increases; see Figure 6.4.

For banks DE020 and DE028, the default probability becomes more sensitive to the edge probability as the shape parameter increases. Although every default probability converges to zero, bank DE028 is the most sensitive to p for all shape parameters. By Table D.1, bank DE028 has smallest net worth compared to bank DE020 and DE025. In a sparse network, it is affected by the default of the group of banks in Table 6.2 as losses due to fundamental defaults are shared only by a few banks. In effect, bank DE028 equity buffer (net worth) is not large enough to prevent it from defaulting. As the network becomes more connected, the losses are absorbed by more banks and, because of its relatively hight CTA, the likelihood of default is reduced efficiently.

As shown in Table 6.2, incorporating default costs changes fundamentally the outcome of the stress test. Figure 6.3 illustrates that best case scenarios with no default costs become worst case scenarios with default costs. Indeed, with no cost, bank DE025 has the highest probability of default among banks affected by contagious defaults, while it is the least likely bank to default in the presence of default costs. Moreover, the curvature of its default

probability as a function of the edge probability increases with the shape parameter.

6.2.3 Structure of the Small Network

Regarding the overall structure of the network, we find that mean out and in-degree are essentially the same; see Table 6.2. Because interbank liabilities are built as perturbations of interbank assets by (6.1.1), the structure of interbank assets and liabilities is symmetric for a given bank.

We now study the effect of the shape parameter on the structure of the network. As in the previous subsection, we choose three different values of the shape parameter, $\alpha = 1, 3, 5$, and their associated rate parameters. Figure 6.4 shows that large banks are the most connected with at least 4 banks with an edge probability $p = 0.2$ and the number of mean out-degrees increases at a steady rate for all shape parameters. For a fixed edge probability p, increasing the shape parameter significantly reduces, the connectivity of each small bank. In Figure 6.4, small banks face a reduction of 33% of mean out-degrees on average. For medium and large banks, the choice of the prior has less impact with a reduction of 11% for medium banks and an increase of 2% in mean out-degrees for large banks. Furthermore, the mean out-degree of each small bank shifts downwards for all values of p, while the difference in mean out-degrees may change sign as p increases for medium and large banks. Note that, in this example, banks suffering from contagious default are more sensitive to the choice of the prior.

For the minimal value of the edge probability $p = 0.2$, the network is strongly sparse regardless of the choice of the prior. As a result, the prior does not affect significantly the value of mean out-degrees. When the edge probability is in the interval $p \in [0.5, 0.7]$, the choice of the shape parameter results in a strong heterogeneous network. Such values of p allow for more variability across edge realisations and the choice of the prior changes the network structure at the local and global level. For example, at $p = 0.7$, there is a difference of 4 out-degrees between bank DE020 and DE025 for $\alpha = 1$ and a difference of 7 for $\alpha = 5$. Finally, all banks connect as the edge probability is set to $p = 1$.

While the effect of the prior of small banks connectivity by concentrating the exposure to fewer banks for the same prior edge probability, large banks connectivity is less sensitive to the choice of the prior. This effect can be

explained by the fact that the discrete part of the posterior distribution, which corresponds to the value 0, is dominant for small banks and any change on the discrete part is directly reflected on the number of mean out-degrees.

6.3 Large Network

6.3.1 Stress test

In terms of banks' interbank assets, the European network in Table E.1 has a larger variability than the German network. As such, our aim is to understand whether the structure and the stability of the network is affected by its size and its heterogeneity. Furthermore, our interest lies also in the verification of the performance of the Gibbs sampler on a large network.

Similarly to the German network, we classify banks by size of interbank assets a_i. In the first tercile, $a_i \leq 7602$, there are 26 banks, in second tercile, $7602 < a_i \leq 34983$, and third tercile, $a_i > 34983$, there are respectively 25 banks. On average, the CTA ratio decrease with the size group, with a CTA of 4.87, 4.20, and 3.54, for small, medium and large banks. Therefore, the external shock of 4%, $s = 0.96$ on the system results in more fundamentally insolvent medium and large banks, with 5 small banks, 9 medium banks, and 12 large banks. Furthermore, similarly to the German network, fundamental defaults have the largest MLGDs.

In Table 6.3, we report the characteristics of 9 banks, which are either fundamentally insolvent, affected by contagious default, or non-defaulting. As discussed above, fundamentally insolvent banks DE017 and IE037 default because of their low CTA. Although Bank ES069 has a larger CTA than banks DE017 and IE037, it is fundamentally insolvent because 97% of its exposure is on external assets. As a result, bank ES069 is more sensitive to a shock on its external assets than banks with similar CTA, such as the fundamentally solvent bank IE039 that has an exposure of 85% on its external assets.

Banks ES067, ES076, IE039 suffer from contagious default. In the stress test with shock $s = 0.96$ and with no cost, they have significantly higher default probabilities and MLGD's than similar banks in the stressed German network with a shock $s = 0.97$. As such, we see that, for the same edge probability $p = 0.5$, the large size of the network is not able to absorb

a stronger shock of 1%, although it should facilitate the spread of losses across banks. In this aspect, the behavior of the default probability of bank ES076 demonstrates that neither the size nor the connectivity of the network necessarily helps to reduce the probability of default. Indeed, Figure 6.5 shows that its probability of default increases with the edge probability p regardless of the choice of the prior in the no cost case. While its size of default is small compared to other defaulting banks, the likelihood is high as it has significantly more exposure to the interbank asset in terms of assets to total assets ratio, namely 63.65% compared to 40.20% and 13.11% for banks ES067 and IE039. When default costs are incorporated, the size of default is the same for both networks, since default costs cannot be shared through the network.

The choice of the prior has different effects on the default probability affected by contagious defaults. In the no cost case, increasing the shape parameter increases the slope of the default probability profile for bank ES076. For banks ES067 and IE039, the probability of default decrease with p, while starting from a slightly higher default probability at $p = 0.2$. With default costs, increasing the shape parameter has no significant impact on the default probability of bank ES067 and bank IE039, while it reduces the default probability of bank ES076.

Bank	Bank	CTA	Mean degree		Default without cost		Default with cost	
Code	Size	Ratio(%)	in	out	MLGD (%)	Proba	MLGD (%)	Proba
Fundamental default								
DE017	L	1.59	38.77	38.74	2.09	1	5.78	1
ES069	S	3.59	11.26	11.06	0.31	1	4.17	1
IE037	M	2.79	21.62	21.78	0.95	1	4.70	1
Contagious default								
ES067	S	3.89	12.87	13.09	0.39	0.002	3.40	0.88
ES076	S	3.76	10.50	10.34	0.07	0.49	3.81	0.98
IE039	S	3.60	17.35	17.50	0.05	0.007	3.54	0.97
No default								
DK009	S	5.19	15.56	15.91				
GB091	M	4.77	29.02	29.09				
IT040	L	4.53	37.12	37.02				

Table 6.3: Characteristics of the banks in the European network. The third columns corresponds to the capital to total assets (CTA) ratio. The fourth and fifth columns represent the mean number of interbank liabilities and assets respectively. Parameters. Sample size = 10000, edge probability $p = 0.5$, shape parameter $\alpha = 1$, shock $s = 0.96$, default cost $d = 0.96$.

6.3.2 Structure of the Large Network

The structure of the European network shares the same characteristics with the German network. Because of assumption (6.1.1) on the vector to total liabilities, the network is symmetric and the mean out and in-degrees are essentially the same. Similarly to the German network, Figure 6.6 shows that the choice of the prior has a strong influence on the connectivity of small banks. Reported banks suffering from contagious banks are all small banks and we observe a reduction of 50% in mean out-degrees from $\alpha = 1$ to $\alpha = 3$. Moreover, for large banks, we also find that the difference in mean out-degrees for two different priors may change sign as the edge probability p increases. For example, bank DE017 has 28% more out-degrees at $p = 0.2$ and 8% less out-degrees at $p = 0.9$. Hence, the assumption of homogeneous prior edge probability show that our sampling method which is determined by the size of interbank assets and liabilities reflects the expected tiered structure of interbank networks, with larger banks, being more connected. Finally, the dominance of the discrete part of the posterior distribution increases with the shape parameter, and, as a result, small banks are especially sensitive to the choice of the prior.

6.4 Conclusion

We have developed a new method for sampling from the joint distribution of liabilities conditionally on the given sum of the nominal interbank assets and liabilities of each bank. Our method allows for general Gamma density functions on the continuous part of the liabilities. At each step of the Gibbs sampler, the sampling of the joint distribution reduces to sample of a one dimensional random variable Δ by selecting an appropriate submatrix. Under the assumption of a homogeneous network and natural scale parameters, we obtain a closed form solution for the CDF and we sample from the law using an inverse transform method.

Based on a network of 11 banks, we show that, for the same prior, our sampling method results in a fully heterogeneous network at each sample. The continuous part of the posterior distributions have different means and variances and the size of the assets is determinant for the dominance of the discrete part over the continuous part of the density. Similarly, the properties of the default probability and connectivity of each bank are also

determined by the size of interbank assets. In particular, the structure of small banks is especially sensitive to the the choice of the prior.

Our empirical study of the European network of 76 banks shows that the Gibbs sampler also operates on large networks at the cost of using a large thinning. Our empirical results demonstrate that the profile of default probabilities is different for banks of the same size and we exhibit an example of a bank whose default probability always increases with the prior edge probability. This serves as empirical example that a more connected network does not necessarily reduce the spread of contagious defaults. Finally, we show that the structure of the large European network obtained with our sampling method is similar to the structure of the smaller German network.

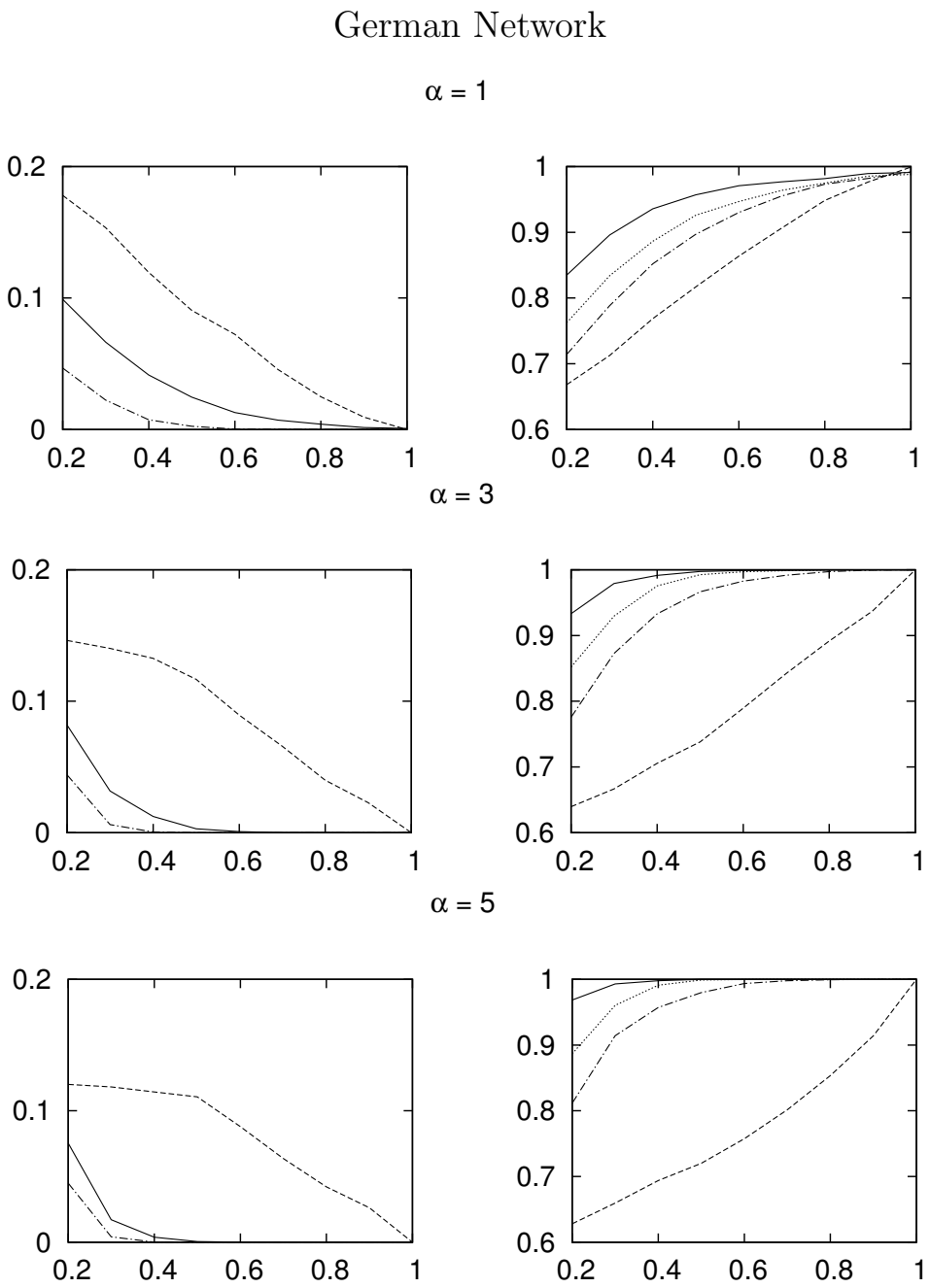
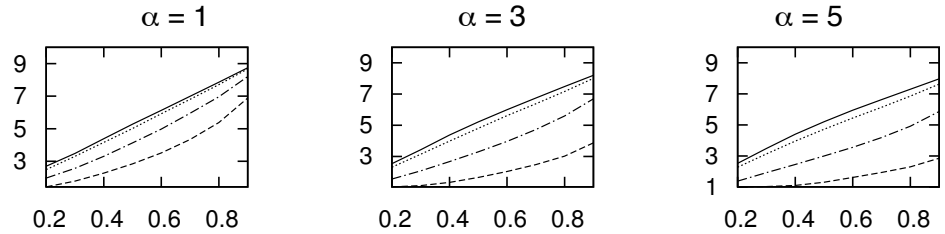
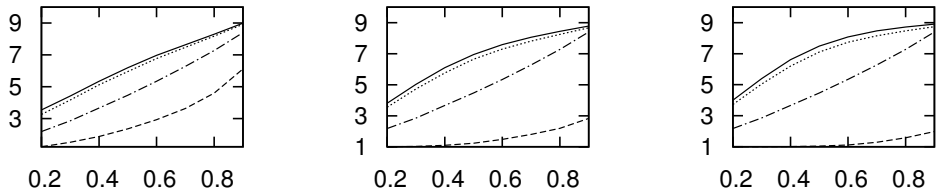


Figure 6.3: Plot of the probability of default, without cost (first column) and with default costs (second column), as a function of the edge probability p for different values of shape parameters. Bank DE019 (dotted), DE020 (full), DE025 (dashed), DE028 (dashed/dotted). Parameters: Sample size = 10000, external shock, $s = 0.97$, default costs $d = 0.95$.

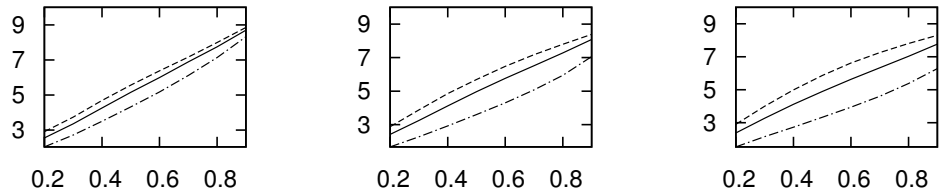
German Network



Banks in fundamental default: DE017 (dotted), DE022 (full), DE023 (dashed), Bank DE024 (dashed/dotted)



Banks in contagious default: DE019 (dotted), DE020 (full), DE025 (dashed), DE028 (dashed/dotted)



Non-Defaulting Banks: DE018 (full), Bank DE021 (dashed), DE027 (dashed/dotted)

Figure 6.4: Plot of the mean out-degree of bank j, $\mathbb{E}(\sum_j A_{ij}|a,l)$, as a function of the edge probability p for different values of shape parameters. Parameters: Sample size = 10000, external shock, $s = 0.97$, default costs $d = 0.95$.

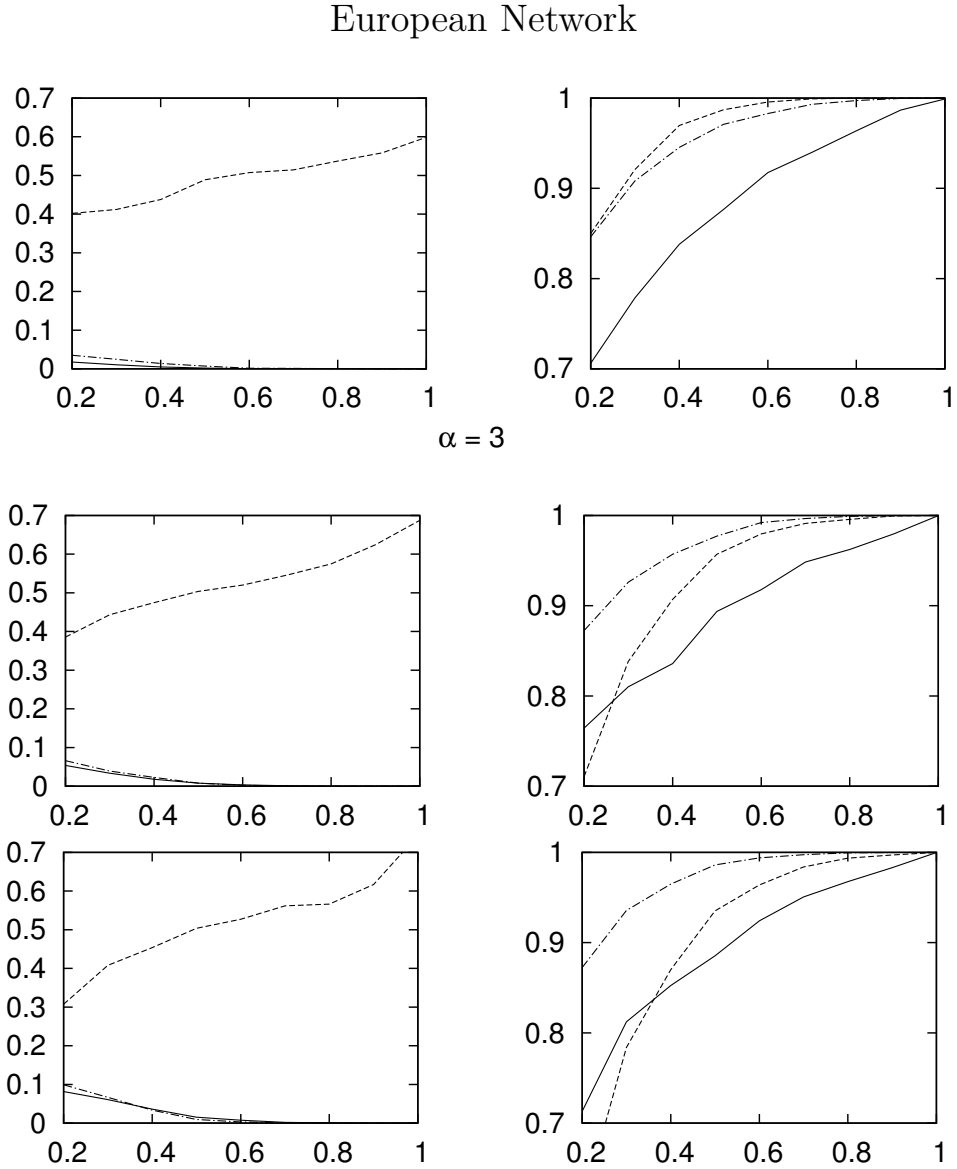
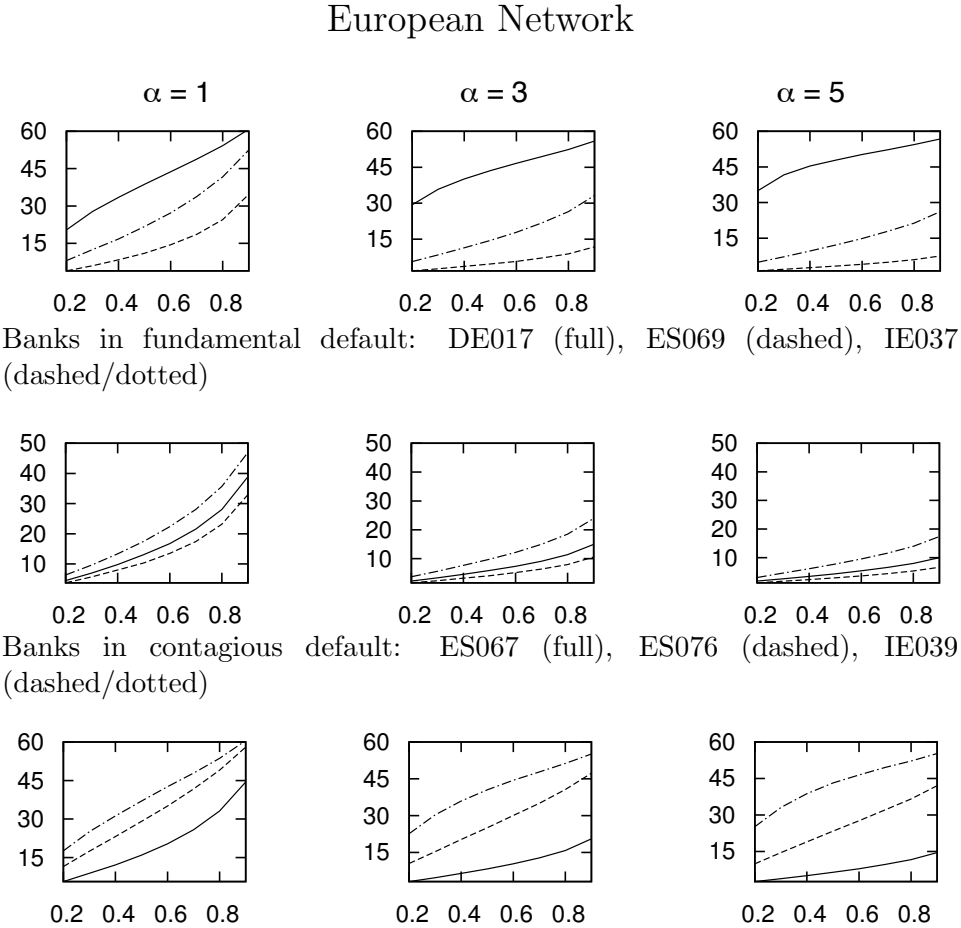


Figure 6.5: Plot of the probability of default, without cost (first column) and with default costs (second column), as a function of the edge probability p for different values of shape parameters. Parameters: Sample size = 10000, external shock, $s = 0.96$, default costs $d = 0.96$. Bank ES067 (full), ES076 (dashed), IE039 (dashed/dotted).



Banks in fundamental default: DE017 (full), ES069 (dashed), IE037 (dashed/dotted)

Banks in contagious default: ES067 (full), ES076 (dashed), IE039 (dashed/dotted)

Non-Defaulting Banks: Bank DK009 (full), GB091(dashed), IT040 (dashed/dotted)

Figure 6.6: Plot of the mean out-degree of bank j, $\mathbb{E}(\sum_j A_{ij}|a,l)$, as a function of the edge probability p for different values of shape parameters. Parameters: Sample size = 10000, external shock, $s = 0.96$, default costs $d = 0.96$.

Appendix D

Balance Sheet: German Network

Bank Code	Name	Total Assets $a + a^{(e)}$	Interbank Assets a	Tier 1 Capital w
DE017	DEUTSCHE BANK AG	1,905,630	47,102	30,361
DE018	COMMERZBANK AG	771,201	49,871	26,728
DE019	LANDESBANK BADEN-WURTTEMBERG	374,413	91,201	9,838
DE020	DZ BANK AG	323,578	100,099	7,299
DE021	BAYERISCHE LANDESBANK	316,354	66,535	11,501
DE022	NORDDEUTSCHE LANDESBANK -GZ-	228,586	54,921	3,974
DE023	HYPO REAL ESTATE HOLDING AG	328,119	7,956	5,539
DE024	WESTLB AG, DUSSELDORF	191,523	24,007	4,218
DE025	HSH NORDBANK AG, HAMBURG	150,930	4,645	4,434
DE027	LANDESBANK BERLIN AG	133,861	27,707	5,162
DE028	DEKABANK DEUTSCHE GIROZENTRAL	130,304	30,937	3,359

Table D.1: Balance sheet information of 11 German banks provided for the 2011 EBA stress test. All quantities are in million euros.

Appendix E

Balance Sheet: European Network

Bank Code	Name	Total Assets $a + a^{(e)}$	Interbank Assets a	Tier 1 Capital w
AT001	Erste Bank Group (EBG)	205938	25044	10507
AT002	Raiffeisen Bank International (RBI)	131173	30361	7641
AT003	Oesterreichische Volksbank AG	44745	10788	1765
BE004	DEXIA	548135	228211	17002
BE005	KBC BANK	276723	23871	11705
CY006	MARFIN POPULAR BANK PUBLIC CO LT	42580	7907	2015
CY007	BANK OF CYPRUS PUBLIC CO LTD	41996	7294	2134
DE017	DEUTSCHE BANK AG	1905630	194399	30361
DE018	COMMERZBANK AG	771201	138190	26728
DE019	Landesbank Baden-Wrttemberg	374413	133906	9838
DE020	DZ BANK AG Dt. Zentral-Genoss.	323578	135860	7299
DE021	Bayerische Landesbank	316354	97336	11501
DE022	Norddeutsche Landesbank -GZ	228586	91217	3974
DE023	Hypo Real Estate Holding AG	328119	29084	5539
DE024	WestLB AG Dsseldorf	191523	58128	4218
DE025	HSH Nordbank AG Hamburg	150930	9532	4434
DE027	Landesbank Berlin AG	133861	49253	5162
DE028	DekaBank Deutsche Girozentrale Franl	130304	41255	3359
DK008	DANSKE BANK	402555	75894	14576
DK011	Nykredit	175888	8597	6633
DK009	Jyske Bank	32752	4674	1699
DK010	Sydbank	20238	3670	1231
ES059	BANCO SANTANDER S.A.	1223267	51407	41998
ES060	BANCO BILBAO VIZCAYA ARGENTARIA	540936	110474	24939
ES061	BFA-BANKIA	327930	39517	13864
ES062	CAJA DE AHORROS Y PENSIONES	275856	5510	11109
ES064	BANCO POPULAR ESPANOL S.A.	129183	14810	6699
ES065	BANCO DE SABADELL S.A.	96703	3678	3507
ES066	CAIXA DESTALVIS DE CATALUNYA	76014	8219	3104
ES067	CAIXA DE AFORROS DE GALICIA	73319	2948	2849
ES083	CAJA DE AHORROS DEL MEDITERRANEO	72034	4981	1843
ES071	GRUPO BANCA CIVICA	71055	7419	3688
ES068	GRUPO BMN	69760	7660	3304
ES063	EFFIBANK	54523	4124	2656

(continued on next page)

Bank Code	Name	Total Assets $a + a^{(e)}$	Interbank Assets a	Tier 1 Capital w
ES069	BANKINTER S.A.	53476	2141	1920
ES070	CAJA ESPANA DE INVERSIONES SALAMANCA	45656	7235	2076
ES075	GRUPO BBK	44628	1924	2982
ES072	CAJA DE AHORROS Y M.P. DE ZARAGOZ	42716	1779	2299
ES073	MONTE DE PIEDAD Y CAJA DE AHORRO	34263	2599	2501
ES074	BANCO PASTOR S.A.	31135	1665	1395
ES076	CAIXA DESTALVIS UNIO DE CAIXES DE...	28310	1802	1065
ES077	CAJA DE AHORROS Y M.P. DE GIPUZKO	20786	285	1935
ES078	GRUPO CAJA3	20144	1926	1164
FR013	BNP PARIBAS	1998157	90328	55352
FR014	CREDIT AGRICOLE	1503621	83713	46277
FR016	SOCIETE GENERALE	1051323	100013	27824
FR015	BPCE	1000695	34983	31943
GB089	HSBC HOLDINGS plc	1783199	212092	86900
GB090	BARCLAYS plc	1725709	53873	46232
GB091	LLOYDS BANKING GROUP plc	1006082	29233	47984
GB088	ROYAL BANK OF SCOTLAND GROUP plc	607351	105506	58982
GR031	NATIONAL BANK OF GREECE	118832	8608	8153
GR030	EFG EUROBANK ERGASIAS S.A.	85885	3838	4296
GR032	ALPHA BANK	66798	3492	5275
GR033	PIRAEUS BANK GROUP	57680	1581	3039
GR034	AGRICULTURAL BANK OF GREECE S.A.	31221	1657	792
IE038	BANK OF IRELAND	156712	17254	7037
IE037	ALLIED IRISH BANKS PLC	131311	11277	3669
IE039	IRISH LIFE AND PERMANENT	46743	6127	1681
IT041	UNICREDIT S.p.A	929488	106707	35702
IT040	INTESA SANPAOLO S.p.A	576962	109909	26159
IT042	BANCA MONTE DEI PASCHI DI SIENA	244279	12074	6301
IT043	BANCO POPOLARE S.C.	140043	7602	5474
IT044	UNIONE DI BANCHE ITALIANE SCPA	130559	19793	6559
NL047	ING BANK NV	933073	111756	30895
NL048	RABOBANK NEDERLAND	607483	37538	27725
NL049	ABN AMRO BANK NV	379599	29196	11574
NL050	SNS BANK NV	78918	388	1782
PT053	CAIXA GERAL DE DEPSITOS SA	119318	14221	6510
PT054	BANCO COMERCIAL PORTUGUS SA	100010	7690	3521
PT055	ESPRITO SANTO FINANCIAL GROUP	85644	8690	4520
PT056	Banco BPI SA	43826	5463	2133
SE084	Nordea Bank AB	542853	61448	19103
SE085	Skandinaviska Enskilda Banken AB	212240	25955	9604
SE086	Svenska Handelsbanken AB	240202	20870	8209
SE087	Swedbank AB	191365	17358	7352

Figure E.1: Balance sheet information of 76 European banks system for the 2011 EBA stress test. The data is provided in Glasserman and Young (2015). All quantities are in million euros.

Bibliography

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Acharya, V. V. and Yorulmazer, T. (2008). Cash-in-the-market pricing and optimal resolution of bank failures. *Review of Financial Studies*, 21(6):2705–2742.

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

Bach, F., Ahipasaoglu, S. D., and d'Aspremont, A. (2010). Convex relaxations for subset selection. Technical report Arxiv 1006-3601.

Bäuerle, N., Urban, S. P., and Veraart, L. A. M. (2013). The relaxed investor with partial information. *SIAM Journal on Financial Mathematics*, 3(1):304–327.

Benoit, S., Colliard, J.-E., Hurlin, C., and Pérignon, C. (2015). Where the risks lie: a survey on systemic risk. Available at SSRN: http://ssrn.com/abstract=2577961.

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342.

Brandt, M. W. (2010). Portfolio choice problems. In *Hanbook of financial econometrics, volume 1: tools and techniques*, volume 1, 269–336. North Holland.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable Markowitz portfolios. *PNAS*, 106(30):12267–12272.

Brunnermeier, M. K. and Oehmke, M. (2013). *Bubbles, Financial Crises, and Systemic Risk*. Elsevier, Amsterdam.

Chan, L. K. C., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies*, 12(5):937–974.

Chen, N., Liu, X., and Yao, D. D. (2014). An optimization view of financial systemic risk modeling: The network effect and the market liquidity effect. Available at SSRN: http://ssrn.com/abstract=2463545 or http://dx.doi.org/10.2139/ssrn.2463545.

Chopra, V. K. (1993). Improving optimization. *The Journal of Investing*, 2(3):51–59.

Cocco, J. F., Gomes, F. J., and Martins, N. C. (2009). Lending relationships in the interbank market. *Journal of Financial Intermediation*, 18(1):24–48.

Cont, R., Moussa, A., and Santos, E. B. (2013). Network structure and systemic risk in banking systems. *Handbook on Systemic Risk*, 1:327–336.

Cvitanic, J. and Karatzas, I. (1992). Convex duality in constrained portfolio optimization. *The Annals of Applied Probability*, 2(4):767–818.

Davis, M. H. A. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):676–713.

De Bandt, O., Hartmann, P., and Peydró, J. L. (2009). Systemic risk in banking an update. *The Oxford Handbook of Banking*.

Degryse, H. and Nguyen, G. (2007). Interbank exposures: An empirical examination of contagion risk in the Belgian banking system. *International Journal of Central Banking*, 3(2):123–171.

DeMiguel, V., Garlappi, L., Nogales, F., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):15.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.

Donoho, L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics.*, 32(2):407–451.

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

El Karoui, N. (2010). High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *The Annals of Statistics*, 38(6):3487–3566.

Elsinger, H., Lehar, A., and Summer, M. (2006). Using market information for banking system risk assessment. *International Journal of Central Banking*, 2(1).

Elsinger, H., Lehar, A., and Summer, M. (2013). Network models and systemic risk assessment. *Handbook on Systemic Risk*, 1:287–305.

Erdős, P. and Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics*, 107:3–56.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.

Farhi, E. and Tirole, J. (2012). Collective moral hazard, maturity mismatch, and systemic bailouts. *American Economic Review*, 102(1):60–93.

Freixas, X., Parigi, B. M., and Rochet, J.-C. (2000). Systemic risk, interbank relations, and liquidity provision by the central bank. *Journal of Money, Credit and Banking*, 32(3):pp. 611–638.

Frost, P. A. and Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *The Journal of Portfolio Management*, 15(1):29–34.

Furfine, C. H. (2003). Interbank exposures: quantifying the risk of contagion. *Journal of money, credit and banking*, 111–128.

Gandy, A. and Veraart, L. A. M. (2013). The effect of estimation in high-dimensional portfolios. *Mathematical Finance*, 23(3):531–559.

Gandy, A. and Veraart, L. A. M. (2015). A Bayesian methodology for systemic risk assessment in financial networks. Available at SSRN: http://ssrn.com/abstract=2580869.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer, 1 edition.

Glasserman, P. and Young, H. P. (2015). How likely is contagion in financial networks? *Journal of Banking & Finance*, 50:383–399.

Goldstein, M. A., Irvine, P., Kandel, E., and Wiener, Z. (2009). Brokerage Commissions and Institutional Trading Patterns. *Review of Financial Studies*, 22(12):5175–5212.

Green, R. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well-diversified? *Journal of Finance*, 47(5):1785–1809.

Haldane, A. G. (2009). Rethinking the financial network,april 2009.speech delivered at the financial student association, amsterdam.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1684.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., Vol. 1 (Univ. of Calif. Press)*, 361–379.

Jobson, J. D. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

Jobson, J. D. and Korkie, B. (1981). Putting markowitz theory to work. *Journal of the American Statistical Association*, 7(4):70–74.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Karatzas, I. and Shreve, S. (1998). *Methods of mathematical finance*. New York: Springer-Verlag.

Keim, D. B. and Madhavan, A. (1998). The cost of institutional equity trades. *Financial Analysts Journal*, 54(4):50–69.

Kempf, A., Korn, O., and Sassning, S. (2015). Portfolio optimization using forward-looking information. *Review of Finance*, 19(1):467–490.

Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.

Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365 – 411.

Lee, G. M., Tam, N. N., and Yen, N. D. (2006). Continuity of the solution map in quadratic programs under linear perturbations. *Journal of Optimization Theory and Applications*, 129(3):415–423.

Levy, H. and Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review*, 69(3):308–317.

Luenberger, D. G. (1969). *Optimization by vector space methods*. John Wiley & Sons.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Merton, R. C. (1973). An inter temporal capital asset pricing model. *Econometrica*, 41:867–887.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5):1114–1127.

Moreau, J. J. (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la S.M.F*, 93:273–299.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234.

Nier, E., Yang, J., Yorulmazer, T., and Alentorn, A. (2007). Network models and financial stability. *Journal of Economic Dynamics and Control*, 31(6):2033–2060.

Osborne, M. R., Presnell, B., and Turlach, B. A. (2000). On the lasso and its dual. *Journal of Computational and Graphical Statistics*, 9(2):319–337.

Pham, H. (2009). *Continuous-time stochastic control and optimization with financial applications*. Springer.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C. W., and Kahaner, D. K. (1983). *QUADPACK: A subroutine package for automatic integration*. Springer, 1 edition.

Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer, 2 edition.

Rogers, L. C. G. (2001). The relaxed investor and parameter uncertainty. *Finance and Stochastics*, 5(2):131–154.

Rogers, L. C. G. and Veraart, L. A. M. (2013). Failure and rescue in an interbank network. *Management Science*, 59(4):882–898.

Roll, R. and Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, 35(5):1073–1103.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125.

Upper, C. and Worms, A. (2004). Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *European Economic Review*, 48(4):827–849.

Van, L. I. and Liedorp, F. (2006). Interbank contagion in the dutch banking sector: a sensitivity analysis. *International Journal of Central Banking*, 2(2):99–133.

Wells, S. J. (2004). Financial interlinkages in the United Kingdom's interbank market and the risk of contagion. Bank of England working paper.