UNIVERSITA DEGLI STUDI DI MILANO BICOCCA
Dip. di Economia, Metodi Quantitativi e Strategie di Impresa
Graduate school in Public Economics-DEFAP 29th cycle

Banks' resilience against economic shocks: theoretical and empirical essays

Mariya Pyatkova

Registration number: 787978

Supervisor: Vittoria Cerasi

Co-supervisor: Fabio Bellini

March 2017

Declaration of Authorship

I, Mariya Pyatkova, declare that this thesis titled, "Banks' resilience against economic shocks: theoretical and empirical essays" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

"It is no use going back to yesterday, because I was a different person then."

Lewis Carroll

UNIVERSITA DEGLI STUDI DI MILANO BICOCCA

Abstract

Dip. di Economia, Metodi Quantitativi e Strategie di Impresa

Graduate school in Public Economics-DEFAP 29th cycle

Doctor of Philosophy

Banks' resilience against economic shocks: theoretical and empirical essays

by Mariya Pyatkova

My Phd thesis consists of two independent studies on topics of interest for finance and banking. The first part is a joint work with V. Cerasi. It involves a theoretical model of interlocking directorates in banking. While the second part is a joint work with F. Bellini and I. Negri and it involves both theoretical and empirical studies about backtesting VaR and expectiles with realized scores.

Essay 1

Interlocking Directorates and Macroeconomic shocks

Interlocking directorates occur whenever an executive in one board sits also in the board of another organization. Usually interlocks are seen as a collusive device. Despite legal restrictions on interlocks for antitrust reasons, they remain quite common in the financial sector. In this paper the effect of interlocking directorates is investigated for the special case of banks. We delineate a theoretical model that illustrates cost and benefit of interlocking directorates and compare it to the case of independent boards. The setup of the model is that of imperfect information, since the quality of the board is imperfectly known to rivals. We study the conditions for a Perfect Bayesian equilibrium to exist with interlocking directorates. Furthermore, we show that two symmetric equilibria arise: one with interlocking directorates and one with independent boards. The equilibrium with interlocking may be preferable from a welfare point of view when coordination is prohibited by the antitrust law. This could be particularly relevant for banks located in independent markets but exposed to common macroeconomic shocks.

Keywords: Banks; Interlocking Directorates; Risk-taking.

Essay 2

Backtesting with realized scores

Several statistical functionals such as quantiles and expectiles arise naturally as the minimizers of the expected value of a scoring function, a property that is called elicitability (see Gneiting, 2011 and the references therein). The existence of such scoring functions gives a natural way to compare the accuracy of different forecasting models, and to test comparative hypotheses by means of the Diebold-Mariano test (see Ziegel and Nolde, 2016 and the references therein). In this paper we suggest a procedure to test the accuracy of a quantile or expectile forecasting model in an absolute sense, as in the original Basel I backtesting procedure of Value-at-Risk. To this aim, we study the asymptotic and finite-sample distributions of empirical scores for Normal and uniform i.i.d. samples. We compare on simulated data the empirical power of our procedure with alternative procedures based on empirical identification functions (i.e. in the case of VaR the number of violations) and we find an higher power in detecting at least misspecification in the mean. We conclude with a real data example where both backtesting procedures are applied to AR(1)-Garch(1,1) models fitted to SP500 logreturns for VaR and expectiles' forecast.

Keywords: Backtesting, scoring function, VaR, expectiles.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Prof. Vittoria Cerasi and Prof. Fabio Bellini for the continuous support of my Ph.D study and related research, for their patience, motivation and limitless knowledge. Their guidance helped me during all time of research and writing of this thesis. I could not have imagined having a better advisors and mentors for my Ph.D.

Additionally, I would like to thank my external readers Prof. Ombretta Main, Prof. Johanna Ziegel, for their helpful comments and suggestions and also, for the hard question which incented me to widen my research from various perspectives.

Besides of them, I would like to thank the rest of my thesis committee for their insightful comments, encouragement and suggestions.

I would like to thank the Graduate school in Public Economics- DEFAP for giving me the opportunity to carry out my doctoral research and for their financial support. It would have been impossible for me to even start my study had they not given me a scholarship.

Last but not the least, I would like to thank my family: my mom and my sister for supporting me spiritually throughout writing this thesis and my life in general.

Contents

List of Figures

List of Tables

Abbreviations

ID	**I**nterlocking **D**irectorates
VaR	**V**alue at **R**isk
ES	**E**xpected **S**hortfalls

To my dad, my sister and my mom

Chapter 1

Interlocking Directorates and Macroeconomic shocks

1.1 Introduction

Often we hear of an executive sitting in two or more boards at the same time, that is, of an interlocking directorate. Interlocking directorates (ID) concern either firms operating in different sectors or firms within the same sector. ID are unevenly spread across industries: 30% of all ID are in finance, 26% in energy, 22% in utilities, 14% in construction and 8% in consumer goods and retail (see Heemskerk, 2013).

We start observing that ID are more frequent in the financial industry relatively to other industries. Given that ID may act as a coordination device, they may be in contrast with competition laws and thus prohibited. Why are ID tolerated in the financial sector? Are there benefits in interlocking for the financial sector, that could compensate for their negative impact on competition?

In this chapter we focus on banks for two reasons. First of all, in the financial industry we see a greater percentage of ID. Second, in the banking industry more than in other industries we might see a critical trade-off between financial stability and competition (see Carletti, 2008). In fact banks are, more than other firms, vulnerable to macroeconomic shocks; this vulnerability might give rise to a systemic crisis, such as the global financial crisis we saw in the 2007, with all the disruptive effects for many economic activities. Financial stability could thus be, in extreme macroeconomic conditions, a far more desirable objective in terms of welfare, than preserving competition alone.

In this chapter, we address the analysis of ID within a theoretical model with incomplete information. The analysis provides the conditions for ID to arise as an equilibrium outcome in a non-cooperative game where each bank chooses whether to interlock or to stay independent.

Two banks, located in two separate regions, may be hit by two independent shocks. This leads to a systemic crisis when the two banks are hit at the same time. When a bank fails, the regulator intervenes by seizing and liquidating the assets in the bank portfolio and repaying depositors. The sale of the loans in the bank portfolio involves liquidation costs: these costs are more than proportional when two banks fail instead of one, due to fire sales (selling several assets at the same time). This assumption captures the disruption of a systemic crisis.

Ex-ante each banker might exert an effort to reduce the losses on the portfolio of loans, thus the probability of having to face ex-post liquidation costs; this effort has a cost which is private information for each banker. It is therefore important for the bankers to coordinate on the risk-reducing efforts. Not knowing the effort costs of the rival, this coordination is not possible unless bankers choose to interlock. A banker who sits in the rival bank's board, perfectly observes the cost of effort of the other banker. In this context ID allow to share information.

In this setting we compare the case of independent banks (two bankers sitting each in a separate board) with that of symmetric interlocking directorates (two bankers sitting in both boards). In both cases there is strategic interaction in the choice of efforts: however with ID there is sharing of information.

We search for conditions in which banks choose voluntarily to interlock rather than to stay independent. We show that ID is one of the possible equilibria outcome when the two bankers have symmetric private costs. We suggest that ID could be beneficial as a tool to reduce the costs of systemic risk, since it is the equilibrium in which the welfare losses are minimized. In addition we prove that there are conditions in which the equilibrium with ID arises also when bankers have asymmetric costs.

First of all this chapter contributes to the existing literature by suppling a novel framework to analyze ID. The model has two advantages, first of all it connects the degree of financial stability to ID. Second it distinguishes between ID and collusion: while collusion is the coordination of actions, ID involves just sharing of information without coordination of actions.

The remainder of the chapter is organized as follows. In Section 2 we provide the relation with the existing literature. Section 3 describes the basic model. Section 4 analyzes the equilibrium under the different settings (independent vs. interlocking banks). Section 5 studies the case of collusion and comparison analysis. Section 6 investigates the welfare analysis. Section 7 analyzes the conditions in which interlocking arises as an equilibrium within a non-cooperative game where banks decide whether to interlock or not. Section 8 concludes the chapter. All proofs are in Appendix A.

1.2 Relation to the literature.

Interlocking directorates influence corporate practices in several aspects. Mizruchi (1996) was one of the first to analyze this phenomenon. Academics study interlocking directorates from different perspectives. For instance, Kim et.al. (2011) suggest that ID may be a solution for complex economic and political questions. ID may affect not only politics through social networking, but also could affect the amount and type of funds the firms borrow (see Stearns & Mizruchi, 1993). Geletkanycz & Hambrick, (1997) and Useem(1982) study the effect of ID for the corporate strategy and general business environment.

Indeed, a large part of the literature is devoted to interlocking directorates as a way to collude. Haunschild (1993) provide evidence that firm managers are exposed to the acquisition activities of other firms when they sit on those firms' boards. While Croce and Grassi (2010) highlight the negative effect of ID on firm value. This suggests that controlling shareholders in Italy create a network to protect their private benefits of control. Along the same line, Drago et.al. (2011) find a negative relation between ID and company performance for Italian companies. Polo et al. (2009) focus on ID in Italian financial companies. Main and Creatini (2017) study the recent evolution of ID in the banking sector in Italy and the impact of regulatory limits on ID.

The most related papers are the following. Ambrus et.al. (2014) develop a model of informal risk sharing in social networks, where relationship between individuals can be used as social collateral to enforce insurance payments. Schoorman et.al. (1981) survey the existing literature and suggeste the idea that ID could be considered as a tool to reduce environmental uncertainty, although it never became popular among the practitioners. Moreover, Houston et.al., (2014) study the relation between different measures of centrality in network and its contribution to systemic risk in the banking system.

Here we review the contributions in the literature according to their emphasis to costs or benefits of ID.

Costs. From a competition policy perspective, firms have to take decisions independently to avoid anticompetitive behavior: ID instead may dwindle or eliminate competition and facilitate collusion Conzalez Diaz (2012) . ID are seen as an instrument to cartelize a market because sharing directors allows cartel to monitor directly activities by having an observer in place: this could reduce the cost of reaching and enforcement of cartel agreements. ID can be also a tool for systematic collusion among companies operating in different business sectors finalized at the expropriation of minority shareholders. A system based on direct interlocking directorates may thus potentially produce economic inefficiencies Carbonai & Bartolomeo (2006).

In addition, the indirect evidence on the relation between ID and industry concentration is mix. For instance Pennings (1980) found a positive association between industry concentration and horizontal ties, while Burt (1983) found an inverted U-shaped function: suggesting that with very high market concentration, few producers have little need to interlock to coordinate on prices.

Benefits. ID allow to share information and risks among the firms in the network, at the benefit of reducing uncertainty. The ability to control or at least reduce environmental uncertainty could be a positive side effect of ID. Directors have access to information related to trends, regulation rules, market conditions, new strategies, inter-bank transactions, and other crucial market data, which can flow across corporate boardroom network. Haunschild & Becman (1998) found that interlocking directorates could be a tool for channeling relevant information that may improve each single bank strategy. A well-connected board is one of the tool that is central to the network's aggregate flow of information and resources. Interlocking directorates may provide a better access to this additional information and comparative advantage in making strategic decisions (Mizruchi, 1996, Mol, 2001). The ID provide contacts that are sources of economics benefits and resource exchange (Mol, 2001, Nicholson et.al., 2004). Pfeffer and Salancik (1978) view ID as dyadic inter-organizational strategy, for example, by reducing "competitive uncertainty".

The most relevant paper for our objective is Houston et.al., (2014). They focus on the relation between financial stability and connected banks. They provide empirical evidence that central banks within a network contribute significantly to the systemic risk of the global banking system. The central banks in a network are defined as banks with the greatest number of connections. Bank connections are established through similar education paths, or past or present membership in a corporate board, government or medical institutions. They show that these connections are associated to a greater impact on systemic risk. We instead provide an argument as to why ID may arise as a reply to systemic risk in the financial sector.

Our contribution is a theoretical one. We are not aware of any model that analyses ID within a separate setting from collusion. Furthermore we focus on the banking industry to study the relation between financial stability and ID.

1.3 The basic model

Consider two banks $k = \{1, 2\}$ located in two independent regions. Each bank collects one unit of deposit to invest into a single illiquid loan. Each loan yields either $R > 1$ in case of success or 0 in case of failure at date $t = 2$. Project returns are i.i.d. across regions.

We model the activity of the bank as follows: each banker exerts an effort $\theta_k \in [0; 1]$, with associated effort cost proportional to c_k. The effort cost is quadratic. The banker's effort θ_k is not observable by third parties and, given that c_k is a private cost, there is a moral hazard problem.

There are two possible states of the world at date $t = 1$: state good, G, or state bad B.

The return of the loans are affected by the state of the world:

- in state G the loan returns always R;
- in state B the loan returns R with exogenous probability $p \in (0, 1)$, 0 otherwise.

Although with uncertainty, the banker's effort may improves the probability of success of the loan, by increasing the chances that state G occurs. In other words, the probability that state G occurs is $q(\theta_k) = \theta_k$. Hence the banker by exerting the effort may reduce the probability of ending up in the bad state of the world B where the probability of losses is positive, i.e. probability $1 - p$. The setup we have in mind is that B describes the case in which the region is hit by a negative macroeconomic shock. To avoid losses in the portfolio the banker needs to exert monitoring effort on the loans. In this way the probability of having losses is reduced.

Given that depositors are infinitely risk averse, there is a deposit insurance that acts in the interest of depositors: the deposit insurance refunds depositors when the bank fails to repay them. Deposit insurance, alike all third parties, perfectly observes the state of the world, but not the banker's effort level. Notice that although the state of the world $\{G, B\}$ is observable, the banker's effort can never be inferred from the state of the world. Conditional on the state of the world, the deposit insurance may decide to liquidate the loan in the bank's portfolio to avoid larger future losses.

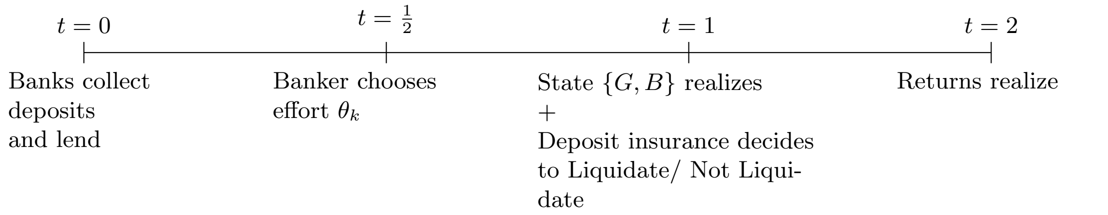

FIGURE 1.1: Timing of the model.

The timing of the model is summarized in Figure 1.1.

The game is solved by backward induction.

1.3.1 The role of deposit insurance

Let's focus at time $t = 1$ when the state of the world is observed by all third parties. When observing the state of the world, the deposit insurance may decide to liquidate the loan in the bank's portfolio in order to repay depositors and avoid larger future losses. Deposit insurance can decide to liquidate both banks (L_1, L_2), just one bank (L_1, NL_2) or (NL_1, L_2) or none of the banks (NL_1, NL_2).

Early liquidation of the loan portfolio can be expressed as a percentage recovery value of the return R, that is $w_s \in (0, 1)$. Assume that this percentage is:

$$w_s = \begin{cases} L & \text{if only one asset is liquidated;} \\ \ell & \text{if two assets are liquidated.} \end{cases}$$

This assumption means that the percentage of the recovery value depends upon the number of liquidated assets. We assume that $\ell < L < 1$, that is, there is a greater loss in recovery value when selling multiple assets at the same time ("fire sales"). This assumption captures our idea of systemic risk: when two banks are liquidated simultaneously there are larger losses in the economy compared to the case when just one bank is liquidated.

Assume that $RL < 1$ which means deadweight loss of liquidation. This assumption implies the idea of liquidation choice: when the deposit insurance decide to liquidate the bank it fully insures depositors and cashes the amount RL from the sold asset. In case the deposit insurance does not liquidate the assets in the portfolio of the bank instead, once depositors are repaid, the rest of the revenue remains to the banker.

Let's consider the total disbursement costs for the deposit insurance in each of the four possible states of the world according to the possible decisions about liquidating or not the two banks.

TABLE 1.1: Expected disbursement for the deposit insurance in each possible state of the world

State & Decision	(L_1, L_2)	(L_1, NL_2)	(NL_1, L_2)	(NL_1, NL_2)	
(B, B)	$2(1-\ell R)$	$(1-RL)+(1-p)$	$(1-p)+(1-RL)$	$2(1-p)$	$\ell < \frac{p}{R}$
(B, G)	$2(1-\ell R)$	$(1-RL)$	$(1-p)+(1-RL)$	$(1-p)$	$L > \frac{p}{R}$
(G, B)	$2(1-\ell R)$	$(1-RL)+(1-p)$	$(1-RL)$	$(1-p)$	$L > \frac{p}{R}$
(G, G)	$2(1-\ell R)$	$(1-RL)$	$(1-RL)$	0	always

At time $t = 1$ the deposit insurance anticipates that there is a probability $(1-p)$ of zero return on each loan portfolio. Therefore it can consider to liquidate the banks to avoid those losses and recover at least the certain amount LR by liquidating bank's assets. However in case deposit insurance liquidates the two banks the recovery rate is proportionally smaller, i.e. $\ell < L$, due to fire sales.

In Table 1.1 we report in each cell the total disbursement from the deposit insurance' s point of view according the liquidation choice and the realized couple of states of the world in the two regions.

The deposit insurance chooses the action (liquidation versus non liquidation) that minimizes its expected disbursement value in each of the possible states. We start by state (B, B). When (B,B) state occurs the decision (NL_1, NL_2) insures the minimum losses whenever $\ell < \frac{p}{R}$. In state (B,G) the minimum is reached with (L_1, NL_2) when $\ell < L < 1$. In state (G,G), (NL_1, NL_2) is the minimum under our assumptions.

Notice that there is no strategic interaction if $\ell = L$. Looking at table 1.1, if bank 1 is in the good state, than deposit insurance decide not to liquidate regardless bank 2 being in good or bad state and otherwise. This implies that information about rival's cost does not affect bank's choices.

At time $t = \frac{1}{2}$ banks maximize their profits. Table 1.2 presents the gross payoff (ignoring the costs of effort) for each banker in each different state of the world, anticipating the liquidation choice of the deposit insurance.

Indeed, banks need to put more effort to avoid losses in their portfolio.

1.4 Equilibrium profits with and without ID

Recall state $t = \frac{1}{2}$, at which banks maximize their profits, anticipating the liquidation choice of deposit insurance in each state of the world. Assume, that the cost of effort

TABLE 1.2: Bankers' payoffs under different states of the workd, taking into account liquidation choices

State & Bank's payoff	Prob. for bank 1	Prob. for bank 2	Bank 1's return	Bank 2's return
(B, B)	$(1-\theta_1)$	$(1-\theta_2)$	$p(R-1)$	$p(R-1)$
(B, G)	$(1-\theta_1)$	θ_2	0	$(R-1)$
(G, B)	θ_1	$(1-\theta_2)$	$(R-1)$	0
(G, G)	θ_1	θ_2	$(R-1)$	$(R-1)$

can assume only two values c_H or c_L. A banker with cost c_H is inefficient, while with cost $c_L < c_H$ is efficient. Remember that the cost is private information of the banker.

We consider two different cases.

1. **Independent bankers (IB,IB)**: each banker sits in one single board. Since the cost of effort is private information of the banker, this setup is one of symmetric and incomplete information.

   We define ID the case when one banker sits in the rival bank's board: in this case the interlocking director observes the private cost of the other banker.

2. **Symmetric interlocking directorates (IL,IL)**: two banks have both an executive in the other bank's board. The set up is one of symmetric and complete information.

Let's solve each case separately and compare the results.

1.4.1 Independent banks

Two banks act independently of each other and behave symmetrically. Each bank may exert an effort to maximize the expected profit anticipating the liquidation choice of the deposit insurance. Each banker's type is private information. Hence bankers don't know the type of the other banker; this requires solving a Static Bayesian Game with imperfect information.

We assume that each banker attaches a prior probability (belief) $\mu \in (0, 1)$ to the fact that the rival's cost of effort is high H and probability $(1-\mu)$ to the cost being low L. Notice that the probability μ is exogenous and belong to the open interval $\mu \in (0, 1)$.

Let's focus on banker 2's problem, whose expected profit is:

$$\Pi_2(c_H, \mu) = (1-\theta_2)[\mu(1-\theta_1(\mu, c_H)) + (1-\mu)(1-\theta_1(\mu, c_L))]p(R-1) + \theta_2(R-1) - c_H\frac{\theta^2}{2} \quad (1.1)$$

By maximizing Π_2 with respect to θ_2 we derive the reaction function for banker 2:

$$\theta_2^*(c_H,\mu) = \frac{(R-1)}{c_H}\{1 - p[\mu(1-\theta_1^*(\mu,c_H)) + (1-\mu)(1-\theta_1^*(\mu,c_L))]\}. \tag{1.2}$$

Proposition 1.1. *In the Bayesian Nash Equilibrium (BNE) the bankers' efforts are:*

$$\begin{cases} \theta_2^{IB,IB}(\mu,c_H) = \frac{(R-1)(1-p)}{c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}]}; \\ \theta_2^{IB,IB}(\mu,c_L) = \frac{c_L}{c_H}.\theta_1^{IB}(c_H,\mu); \\ \theta_1^{IB,IB}(c_H,\mu) = \theta_1^{IB}(c_H,\mu); \\ \theta_1^{IB,IB}(c_L,\mu) = \frac{c_L}{c_H}.\theta_2^{IB}(\mu,c_H). \end{cases}$$

where $p \in [0,1]$, $R > 1$ is the return of the project; $c_H, c_L > 0$ are the two possible values for the effort cost and $\ell < \frac{p}{R}$, $LR < 1$, $L > \frac{p}{R}$. In addition, all the equilibrium efforts, $\theta^{IB,IB}$, belong to the interval $(0,1)$.

Proof. See Appendix A. □

The (IB,IB) equilibrium profit can be derived by substituting the equilibrium efforts in the expression of the expected profit in (1.1). For instance, bank 2's equilibrium profit is:

$$\begin{aligned} \Pi_2^{IB,IB}(\mu,c_H) &= (1-\theta_2^{IB,IB}(c_H,\mu))[\mu(1-\theta_1^{IB,IB}(\mu,c_H)) + (1-\mu)(1-\theta_1^{IB,IB}(\mu,c_L))]p(R-1) \\ &+ \theta_2^{IB,IB}(R-1) - c_H\frac{(\theta_2^{IB,IB}(c_H,\mu))^2}{2} \end{aligned} \tag{1.3}$$

Similarly, we can compute the equilibrium profit for bank 1.

We can therefore compute the equilibrium profits $\Pi_2^{IB,IB}(\mu,c_2)$, $\Pi_1^{IB,IB}(c_1,\mu)$ for each possible combination of costs L and H for the two banks.

1.4.2 Interlocking Directorates

In the case of symmetric interlocking directorates (henceforth (IL,IL)), the two banks have each an executive in the rival bank's board. Therefore, each banker observes the type of the other banker. Each banker chooses her own effort to maximize the profit of the native bank. In this case information is symmetric and complete.

Bank 2's profit is:

$$\max_{\theta_2} \Pi_2 = (1-\theta_2)(1-\theta_1)p(R-1) + \theta_2(R-1) - c_1\frac{\theta_2^2}{2}, \tag{1.4}$$

where we write profit for a generic cost c_2 that stands either for c_H or c_L. By maximizing Π_2 with respect to θ_2 we derive the reaction function for banker 2:

$$\theta_2^{IL,IL} = \frac{(R-1)}{c_1}[1 - p(1 - \theta_1^{IL})]. \tag{1.5}$$

Proposition 1.2. *The equilibrium efforts in the Nash equilibrium are:*

$$\begin{cases} \theta_2^{IL,IL}(c_1, c_2) &= \frac{(1-p)(R-1)(c_1+p(R-1))}{c_1c_2-p^2(R-1)^2}. \\ \theta_1^{IL,IL}(c_1, c_2) &= \frac{(1-p)(R-1)(c_2+p(R-1))}{c_1c_2-p^2(R-1)^2}; \end{cases}$$

where $p \in [0, 1]$, $R > 1$ is the return of the project; $c_H, c_L > 0$ are the two possible values for the effort cost and $\ell < \frac{p}{R}$, $LR > 1$, $L > \frac{p}{R}$. In addition, all the equilibrium efforts $\theta^{IL,IL}$ belong to the interval $(0, 1)$.

We can substitute the efforts into the profit function and derive the L profits at equilibrium for each combination of types. Thus the equilibrium profits are $\Pi_2^{IL,IL}(c_1, c_2)$,$\Pi_1^{IL,IL}(c_1, c_2)$.

Proof. See Appendix A. □

The (IL,IL) equilibrium profits for bank 2 can be derived by substituting the equilibrium efforts in the expression of the expected profits in the equation (1.4):

$$\begin{aligned} \Pi_2^{IL,IL}(c_1, c_2) &= (1 - \theta_1^{IL,IL}(c_1, c_2))[1 - \theta_2^{IL,IL}(c_1, c_2)]p(R-1) + \theta_2^{IL,IL}(c_1, c_2)(R-1) \\ &- c_2\frac{(\theta_2^{IL,IL}(c_1, c_2))^2}{2} \end{aligned} \tag{1.6}$$

Similarly, we can derive the equilibrium profit for bank 1.

Notice that, we can provide close formula for the bank's profit if $c_1 = c_2 = c$ (see Appendix A).

1.5 Collusion.

In this section we compute the solution under collusion and we compare it with the equilibrium with interlocking directorates. In the antitrust literature we often find the two cases a bit overlapped. Collusive agreements are forbidden because of their anti-competitive effects. In this chapter we distinguish between coordination of actions (co-operative choice of efforts) with the sharing of information that comes from ID which does not necessarily imply coordinated choice of efforts. When banks collude (or merge) they maximize joint profits (or a weighted sum of profits) and set efforts in a coordinated manner. In this case there is symmetric information and coordination on effort choices.

$$\begin{aligned}\max_{\theta_1,\theta_2}(\Pi_1+\Pi_2) &= (1-\theta_1)(1-\theta_2)p(R-1)+\theta_1(R-1)-c_1\frac{\theta_1^2}{2} \\ &+ (1-\theta_2)(1-\theta_1)p(R-1)+\theta_2(R-1)-c_2\frac{\theta_2^2}{2} \qquad (1.7)\end{aligned}$$

where we derive profit for a generic cost c_1, which stands either for c_H or c_L. By maximizing $\Pi_1+\Pi_2$ with respect to θ_2 we derive the reaction function:

$$\theta_2^C = \frac{(R-1)}{c_1}[1-2p(1-\theta_1^C)]. \qquad (1.8)$$

Proposition 1.3. *At the optimum the efforts of the two banks are:*

$$\begin{cases}\theta_1^C = \frac{(R-1)(1-2p)[c_2+2(R-1)p]}{c_1c_2-4p^2(R-1)^2}; \\ \theta_2^C = \frac{(R-1)(1-2p)[c_1+2(R-1)p]}{c_1c_2-4p^2(R-1)^2}.\end{cases}$$

with $p \in [0.1]; R > 1; c_k > 0;\ \theta_k^C \in (0,1)$ and $\ell < \frac{p}{R}$; $RL < 1$; $L > \frac{p}{R}$.

Proof. See Appendix A. □

The equilibrium profits in collusion can be derived by substituting the equilibrium efforts in the expression of the expected profits. Notice that, we can provide close formula for the banker's profit if $c_1 = c_2 = c$ (see Appendix A1.)

Comparison analysis Inefficient banks $c_1 = c_2 = c_H$.

In the order to investigate the difference from the profits in symmetric interlocking case, collusion and independent case, we provide a numerical example for inefficient banks $c_1 = c_2 = c_H$ with fixed parameters $p = 0.7; R = 1.3; c_L = 0.3; c = 0.7$ (see Figure 2). The profits have been computed by inserting fixed parameters in the equilibrium profits in symmetric interlocking case (1.4), collusion (1.7) and independent case (1.1).

Figure 2 shows profits as a function of prior belief that rival's bank is inefficient. In collusion (yellow line) and symmetric interlocking cases (red line) profits do not depend

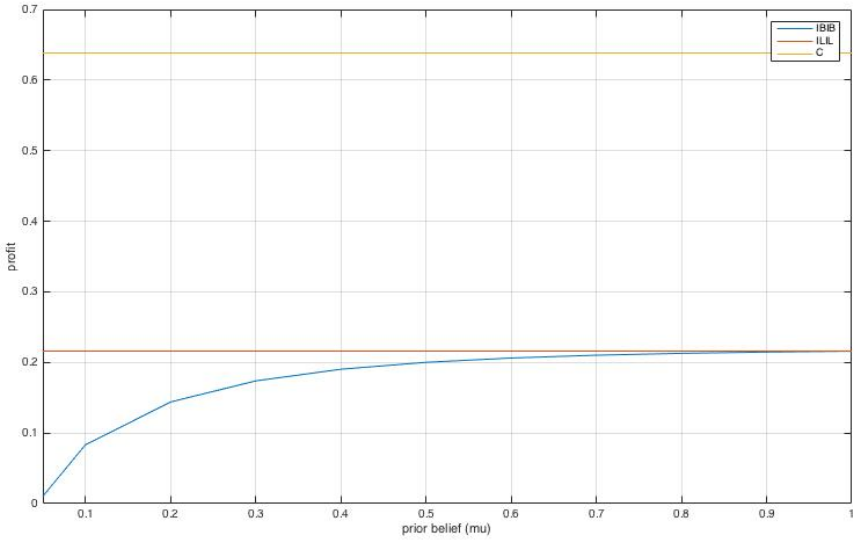

FIGURE 1.2: Bank's profit Π^{C}, $\Pi^{IL,IL}$ and $\Pi^{IB,IB}$ for $p = 0.7$; $R = 1.3$; $c_L = 0.3$; $c_H = 0.7$ and $\mu \in (0, 1)$.

on prior belief, since banks perfectly know the rival's type. While in independent case banks' profit (blue line) depends on it.

First of all, the difference of the profit for ID and independent case comes through the prior believe that rival's bank is inefficient. We consider the case when two banks are inefficient. In the independent case bank 1 may have a wrong belief that bank 2 is efficient which implies the lower profit for the banks (situation in the neighbourhood of 0) with respect to ID profit, since in ID case banks perfectly observe the rival's type. On the contrary, if bank 1 have correct belief that bank 2 is inefficient then the profits in independent and ID cases are equal to each other.

Indeed, one may see ID as an intermediate case between the case of independent banks and collusion. By choosing to interlock, the independent banks can establish a channel in which private information flows, as the interlocked director can effectively communicate reliable information in a timely manner between the two banks. The connection between banks and the access to private information can reduce uncertainty.

Collusion constitutes a way in which companies can increase their profits, while ID may enable flow of information. Banks use interlocks as strategic tools to access and exchange information and resources in order to reduce uncertainty. In today's complex business environment, information about rivals banks can be very valuable to banks. Mergers and acquisitions of shares reduce competitive uncertainty in more permanent ways than interlocks, but interlocks are more flexible and easier to implement.

Similarly one can provide the numerical example for efficient banks.

1.6 Welfare analysis

In this section we investigate the solution of the game from the deposit insurance perspective. The welfare objective function W in our model implies minimizing the expected disbursement costs for of deposit insurance, that is

$$W = -2(1-\theta)^2(1-p) - 2\theta(1-\theta)(1-RL),$$

is increasing function with respect to banker's effort.

$$\begin{aligned} \frac{\partial W}{\partial \theta} &= 2[(1-2p+RL) - 2\theta(RL-p)] \geq 0 \\ \theta &\leq \frac{(1-p)+(RL-p)}{2(RL-p)}, \end{aligned}$$

where RHS is greater than 1 given the deadweight loss of liquidation assumption ($RL < 1$). Let's compare the equilibrium efforts (IB, IB) and (IL, IL) cases.

$$\frac{(R-1)(1-p)}{c_H - p[R-1][\mu + (1-\mu)\frac{c_H}{c_L}]} > \frac{(1-p)(R-1)(c_1 + p(R-1))}{c_H c_1 - p^2(R-1)^2} \tag{1.9}$$

Lemma 1. *The condition for the (IL,IL) equilibrium to dominate in welfare terms is* $c_L > \frac{1-\mu}{\mu}p(R-1)$*. In this case* $\theta^{IL,IL} > \theta^{IB,IB}$ *for all* $\mu \in (0,1)$*.*

Proof. See Appendix A. □

The main intuition behind of this lemma lies in the same line with results of the game from bankers perspective. From the welfare point of view, we would consider only ID and independent case, because collusion is illegal. Lemma 1 shows condition under which (IL,IL) equilibrium dominates (IB,IB) from the deposit insurance perspective.

The differences between this two equilibrium comes through the prior belief that rival's bank is inefficient. If both banks believe that they are inefficient (μ close to 1) than the equilibrium efforts are greater in the equilibrium with ID. This means that the full information, which generates through interlocking directorates, about rival's bank cost is better whenever banks expect their rival to be very inefficient. Revealing this information allows to avoid setting a low effort in response to what it is believed an inefficient rival. Disclosing the true cost is better in this case.

In other words, deposit insurance will prefer to have interlocking directorates rather than independent. Growing evidence suggests that ID are an important consideration

Table 1.3: Banks' profits and the choice to interlock.

bank 1 & bank 2	IB	IL
IB	$\Pi_1^{IB,IB}(c_1,\mu);\Pi_2^{IB,IB}(\mu,c_2)$	$\Pi_1^{IB,IL}(c_1,\mu);\Pi_2^{IB,IL}(c_1,c_2)$
IL	$\Pi_1^{IL,IB}(c_1,c_2);\Pi_2^{IL,IB}(\mu,c_2)$	$\Pi_1^{IL,IL}(c_1,c_2);\Pi_2^{IL,IL}(c_1,c_2)$

in understanding corporate behaviour. ID have access to wide range of information, without inserting prior belief of rivals bank cost, which gives a higher equilibrium with respect to independent case.

1.7 Existence of the equilibrium with interlocking

Here we ask whether interlocking can arise as an equilibrium outcome in a non-cooperative game. At date $t = -1$ each bank decides whether to interlock without coordinating with the rival bank. Each bank can fully anticipate profits when interlocking or remaining independent banks.

Table 1.3 summarizes the payoff of the 2 players.

In order to be able to prove that an equilibrium with ID exists, we have to check for any possible deviations from the two equilibria. In other words, we have to solve the asymmetric case in which one bank interlocks, while the other remain independent, namely the asymmetric ID (henceforth (IB,IL)). One of the two banks is an independent bank, while the other bank has an executive in the board of the rival bank.

Notice that in one hand, our interpretation of ID here is that of a "unilateral" flow of information: the executive in the board delivers information from the "rival" board to the "native" board, not viceversa. On the other hand, ID does not mean that the banks coordinates actions. The difference between ID and collusion is that interlocking allows the sharing of private information but not coordination of actions. The exchange of such information may reduce environmental uncertainty. Cross-ownership of shares reduces uncertainty in a more permanent way compared to interlocks, although they are forbidden if anti-competitive. In contract to mergers, ID are not subject to the same antitrust scrutiny. The reason is that ID are not necessarily coordination of strategies.

1.7.1 Asymmetric interlocking directorates

The bank with ID has an advantage compared to the bank without. The set up is one of asymmetric and incomplete information. Asymmetric incomplite information implies that bank 2 observes both effort costs, while bank 1 knows only its own cost.

We solve this game by backward induction in two steps. First, we find the optimal effort for bank 2, who is perfectly informed about the rival's cost. The profit of the bank 2 is:

$$\max_{\theta_2} \Pi_2 \quad = \quad [1-\theta_1(c_1,\mu)](1-\theta_2)p(R-1)+\theta_2(R-1)-c_2\frac{\theta_2^2}{2}. \tag{1.10}$$

The reaction function for bank 2 is:

$$\theta_2^*(c_1,c_2)=\frac{(R-1)}{c_1}[1-p(1-\theta_1^*(c_1,\mu))] \tag{1.11}$$

Second, we find the optimal effort for bank 1. Bank 1 faces uncertainty about rival's cost of effort. We assume that bank 1 has a prior belief μ that rival's cost is high H and with probability $(1-\mu)$ rival's cost is low L:

$$E(\theta_2^*)=\mu\theta_2^*(c_1,c_H)+(1-\mu)\theta_2^*(c_1,c_L). \tag{1.12}$$

Bank 1's profit is:

$$\max_{\theta_1} \Pi_1 \quad = \quad (1-\theta_1)[(1-E(\theta_2^*))p+\theta_1](R-1)-c_1\frac{\theta_1^2}{2}. \tag{1.13}$$

The reaction function for bank 1 is:

$$\theta_1^*(c_1,\mu)=\frac{(R-1)}{c_1}[1-p(1-E(\theta_2^*))]. \tag{1.14}$$

Proposition 1.4. *The efforts in the Nash equilibrium are:*

$$\begin{cases} \theta_1^{IB,IL}(c_1,\mu)=\frac{(R-1)(1-p)[1+p(R-1)\bar{c}]}{c_1-p^2(R-1)^2\bar{c}} \\ \theta_2^{IB,IL}(c_1,c_H)=\frac{(R-1)(1-p)[c_1-2p^2(R-1)^2\bar{c}+p(R-1)]}{c_H[c_1-p^2(R-1)^2\bar{c}]}; \\ \theta_2^{IB,IL}(c_1,c_L)=\frac{(R-1)(1-p)[c_1-2p^2(R-1)^2\bar{c}+p(R-1)]}{c_L[c_1-p^2(R-1)^2\bar{c}]}, \end{cases}$$

where $\bar{c}=\frac{c_H+\mu(c_L-c_H)}{c_Lc_H}$.

Proof. See Appendix A. □

The (IB, IL) equilibrium profits can be derived by substituting the equilibrium efforts in the two expressions of expected profits (1.10) and (1.13). Bank 2's profit in equilibrium is:

$$\Pi_2^{IB,IL}(c_1, c_2) = [1 - \theta_1^{IB,IL}(c_1, \mu)](1 - \theta_2^{IB,IL})p(R-1) + \theta_2^{IB,IL}(R-1) - c_2\frac{(\theta_2^{IB,IL})^2}{2} \quad (1.15)$$

Notice that, we can provide close formula for the banker's profit if $c_1 = c_2 = c$ (see Appendix E.1.2).

1.7.2 Conditions for the existence of Subgame Perfect Equilibria

We are now able to compare the different equilibrium profits in Table 1.3 in order to find the Subgame Perfect equilibrium/a.

Given the complex analytical expression for equilibrium banks' profits, we solve the game only for the symmetric cost case, i.e. when $c_1 = c_2 = c$ and provide a numerical example to convince the reader. The cost can be either high c_H, meaning inefficient banks, or low c_L in this case banks are efficient. All the computations are reported in Appendix A.

Inefficient banks ($c = c_H$). We study the conditions for which there are no profitable deviations from (IL, IL). In this case we could prove that ID is a Subgame Perfect Equilibrium in the game where banks decide non-cooperatively to interlock.

The (IL, IL) is a Nash Equilibrium when none of the banks find it profitable to deviate, i.e.: $\Pi_1^{IL,IL}(c_1, c_2) > \Pi_1^{IB,IL}(c_1, \mu)$ and $\Pi_2^{IL,IL}(c_1, c_2) > \Pi_2^{IL,IB}(\mu, c_2)$.

Given the symmetry of the two conditions, we check just one the two, namely:

$$\Pi_2^{IL,IL}(c_1, c_2) - \Pi_2^{IL,IB}(\mu, c_2) > 0 \quad (1.16)$$

Condition (1.16) implies that bank 2 should find profitable to learn bank 1 cost, rather than not knowing it.

Recalling the definition of $\Pi_2^{IL,IL}(c_1, c_2)$ and $\Pi_2^{IL,IB}(\mu, c_2)$ from eq(1.13) and eq(1.4), we investigate the impact of this channel of information on banker's effort, i.e.:

$$\Pi_2^{IL,IL}(c_1, c_2) - \Pi_2^{IL,IB}(\mu, c_2) = (1-p)(R-1)(\theta_2^{IL,IL} - \theta_2^{IL,IB}) - p(R-1)(\theta_1^{IL,IL} - E(\theta_1^{IL,IB})) + p(R-1)(\theta_2^{IL,IL}\theta_1^{IL,IB} - \theta_2^{IL,IB}E(\theta_1^{IL,IB})) - \frac{c_2}{2}(\theta_2^{IL,IL} - \theta_2^{IL,IB})(\theta_2^{IL,IL} + \theta_2^{IL,IB}) > 0 \quad (1.17)$$

The intuition comes through two terms $(\theta_2^{IL,IL} - \theta_2^{IL,IB})$ and $(\theta_1^{IL,IL} - E(\theta_1^{IL,IB}))$.

If $c_H > p(R-1)$, then

- $(\theta_2^{IL,IL} - \theta_2^{IL,IB}) > 0$, if bank 2 perfectly observes the cost of bank 1, than she will put greater effort compared to the case in which she does not observe the cost of bank 1;
- $(\theta_1^{IL,IL} - E(\theta_1^{IL,IB})) > 0$, bank 1 will rather prefer bank 2 having ID than being independent.

If $c_H < p(R-1)$, then

- $(\theta_2^{IL,IL} - \theta_2^{IL,IB}) < 0$, if bank 2 perfectly observes the cost of bank 1, than she will put less effort than if she would not observes the cost of bank 1;
- $(\theta_1^{IL,IL} - E(\theta_1^{IL,IB})) > 0$, bank 1 will rather prefer bank 2 having ID than being independent.

There are two situation for ID being Subgame Perfect Equilibrium in inefficient case. It comes from the difference of the banker's effort in this two cases. We compared the banker's effort having symmetrical and asymmetrical interlocking directorates. Notice that bank 1 in both situation having interlocking directorates, while bank 2 may also be independent.

If banks are super inefficient, than the banks' 1 effort is always greater in the case of symmetrical ID then asymmetrical ID. Meaning that the full information, about rival?s bank cost is better of then incomplete information, generated by asymmetric ID.

If bank 2 perfectly observes the cost of bank 1, than she will put greater effort compared to the case in which she does not observe the cost of bank 1. In asymmetric ID case bank 2 may have a wrong belief that bank 1 is efficient which implies the lower effort for the banks with respect to ID.

If banks are less inefficient, than the bank 1 is still have greater effort in the case of symmetrical ID then asymmetrical ID. While bank 2 will put less effort than if she would not observes the cost of bank 1. One may think that if banks are less inefficient than bank 1 do not manage to motivate bank 2 to have interlocking directorates. In other words, bank 2 prefer to stay ignorant to have interlocking directorates, than paying cost to have interlocking directorates.

TABLE 1.4: Numerical example.

bank 1 & bank 2	IB	IL
IB	0.1832; 0.1832	0.1664; 0.1836
IL	0.1836; 0.1664	0.2025; 0.2025

Efficient banks ($c = c_L$). Given the similarity with the $c = c_H$ case, we report only the intuition. The intuition for (IL, IL) comes through two terms $(\theta_2^{IL,IL} - \theta_2^{IL,IB})$ and $(\theta_1^{IL,IL} - E(\theta_1^{IL,IB}))$.

If $c_L > p(R-1)$, then

- $(\theta_2^{IL,IL} - \theta_2^{IL,IB}) > 0$, bank 2 perfectly observes the cost of bank 1, than she will put higher effort than is she would not observes it;
- $(\theta_1^{IL,IL} - E(\theta_1^{IL,IB})) > 0$, bank 1 will rather prefer bank 2 having ID than being independent.

Indeed, in the case of efficient banks the banks' 1 effort is always greater in the case of symmetrical ID then asymmetrical ID. Meaning that the full information, about rival?s bank cost is better of then incomplete information, generated by asymmetric ID.

If bank 2 perfectly observes the cost of bank 1, than she will put greater effort compared to the case in which she does not observe the cost of bank 1. In asymmetric ID case bank 2 may have a wrong belief that bank 1 is efficient which implies the lower effort for the banks with respect to ID.

This result is also consistent with result of cooperative game, which is the full information is better off that incomplete. Disclosing the true cost is better in this case. Generally, this condition is considered to be an indication of bank 2 is motivated to have interlocking directorates.

Given the complex analytical expression for conditions of (IL, IL) is Subgame Perfect Equilibrium are not that trivial we provide a numerical example for inefficient banks $c_1 = c_2 = c_H$ with fixed parameters $p = 0.7; R = 1.3; c_L = 0.3; c = 0.7$ (see Table 1.4). We choose the same parameter as in section 5. There are conditions under which the (IL,IL) is Subgame Perfect Equilibrium.

1.8 Conclusion

This paper analyzes the effect of interlocking directorates in the financial sector especially for banks. Interlocking directorates involves a trade off in terms of positive and negative effects on the decision making set. In one hand,interlocks are seen as a collusion device. In fact, from the antitrust perspective there is legal restriction on interlocks to avoid it. One the other hand, the potential benefit of interlocking directorates steams from a flow of information. Particular, the ability to control or at least reduce environmental uncertainty could be an advantage for interlocking directorates.

We investigate the effect of interlocking directorates through a theoretical model. The cost and benefits of interlocking directorates compared to the case of independent banks. We model interlocking directorates in a set up of incomplete information, where the quality of the board is imperfectly known to rivals. We found conditionals where the Perfect Bayseian equilibrium outcome is improved by interlocking directorates when coordination is prohibited by the antitrust law. This could be particularly relevant for banks located in independent markets but exposed to common macroeconomics shocks.

Moreover, we show that in the case of symmetric effort cost there are two symmetric equilibria: one with interlocking directorates and one with independent banks. The equilibrium with interlocking may be preferable from a welfare point of view when banks are inefficient. This could be particular relevant for banks with common macroeconomic shocks such as financial crises.

Bibliography

[1] Ambrus, A., Mobius, M., Szeidl, A., Consumption Risk-Sharing in Social Networks, *American Economic Review.* 104(1), 149-182 (2014)

[2] Burt, R.S., Corporate Profits and Cooptation, *Ney York: Academic Press*(1983)

[3] Carbonai, D., Di Bartolomeo, G., Interlocking directorates as a trust substitute: The case of Italian non-life insurance industry, textitMPRA Paper 4420, University Library of Munich, Germany.(2006)

[4] Carletti, E., Competition and regulation in banking, *Handbook of financial intermediation and banking*, 126.8, 449-482 (2008)

[5] Conzalez Diaz, F.E., Minority shareholdings and interlocking directorship: The European Union Approach, *Antitrust Chronicle*,1 (2012)

[6] Croci, E., Grassi, R., The economic effects of interlocking directorates in Italy: New evidence using centrality measures. *Computational and Mathematical Organization Theory.* 1-24 (2010)

[7] Drago, C., Millo, F., Ricciuti, R., Santella, P., Corporate governance reforms, interlocking directorship networks and company value in Italy (1998-2007), *CESifo working paper, Industrial Organization* (2011)

[8] Geletkanycz, M.A., Hambrick, D.C., The external ties of top executives: Implications for strategic choice and performance, *Administrative Science Quarterly* 42, 2, 654-681 (1997)

[9] Haunschild, P.R., Interorganizational imitation: The impact of interlocks on corporate acquisition activity, *Administrative Science Quarterly* 38(3), 564-592 (1993)

[10] Haunschild, P.M., Becman, C.M., When do interlocks matter? Alternate sources of information and interlock influence, *Administrative Science Quarterly* 43 815-844 (1998)

[11] Heemskerk, E.K., The rise of the European corporate elite: evidence from the network of interlocking directorates in 2005 and 2010, *Economy and Society* 42(1) 74-101 (2013)

[12] Houston, J.F., Lee, J., Suntheim, F., Social Networks in the Global Banking Sector. *2013.SFS Finance Cavalcade* (2014)

[13] Kim, J., Kogut, B., Yang, J.S., Executive compensation, fat cats and best athletes, *Colunbia Business School Research Paper* (2011)

[14] Main, O., Creatini, R., Interlocking Directorates: un problema ricolto?, *Mercato Concorrenza Regole.* 3, 539-560 (2017).

[15] Mizruchi, M.S., What Do Interlocks Do? An Analysis, Critique, and Assessment of Research on Interlocking Directorates, *Annual Review of Sociology* 22 271-298 (1996)

[16] Mol, M., Creating wealth through working with others: Interorganizational relationships, *Academy of Mangement Executive.* 15, 150-152 (2001)

[17] Nicholson,G., Alexander, M., Kiel, G., Defining the social capital of the board of directors: an exploratory study, *Journal of Management and Organization.*, 10, 54-72 (2004)

[18] Pennings, J.M., Interlocking Directorates, *San francisco: Jossey-Bass.* (1980)

[19] Pfeffer,J., Salancik, G.R., A social information processing approach to job attitudes and task design, *Administrative science quarterly* 6(1), 224-253 (1978)

[20] Polo, A., Drago, C., Santella, P., The Italian chamber of lords sits on listed company boards An empirical analysis of Italian listed company boards from 1998 to 2006. *https://ssrn.com/abstract=1027947* (2009)

[21] Rao, H., Sivakumar, K., Institutional sources of boundary-spanning structures: The establishment of investor relations departments in the fortune 500 inductrials, *Organization Science* 10(1), 27-42 (1999)

[22] Stearns, L.B., Mizruchi, M.S., Board composition and corporate financing: The impact of financial institution representation on borrowing, *The Academy of Management Journal*, 36(3), 608-618 (1993)

[23] Schoorman, F. D., Bazerman, M. H., Atkin, R. S., Interlocking directorates: A strategy for reducing environmental uncertainty, *Academy of Management Review*, 6, 243 ? 251.(1981)

[24] Useem, U., Classwide rationality in the politics of managers and directors of large corporations in the United States and Great Britan, *Administrative Science Quarterly*, 27(2), 199-226 (1982)

Chapter 2

Backtesting with realized scores

2.1 Introduction

The notion of backtesting a risk measure has been the subject of a lot of research either in the academia and in the financial industry. Roughly, backtesting means developing a statistical test of a forecasting model, based on the sequence of realized portfolio losses and of risk measure forecasts. The aim of backtesting is to reject the forecasting model if the chosen test statistics, that should be related with the risk measure under scrutiny, is too far from its theoretical distribution.

Perhaps the first and most well-known example is the Basel I procedure for backtesting Value-at-Risk (VaR) by means of the sequence of violations. VaR at level α is simply the α-quantile of the forecasting distribution and a violation is said to occurr if the realized loss is bigger than VaR. Under the null hypothesis that the forecasting model is correct the sequence of violations is Bernoulli i.i.d. with probability α; classical backtesting procedures were based on counting the number of violations or measuring their tendence to cluster (see Kupiec, 1995, Christoffersen, 1998, Christoffersen and Pelletier, 2004, or Campbell, 2005 for a review).

One common critique to traditional backtesting procedures of VaR (see e.g. Campbell, 2005 or Wong, 2010) was that taking into account only the number of violations and not their magnitude lacks statistical power and may be misleading. For example, a risk manager could easily get a Bernoulli i.i.d. sequence of violations and thus pass a backtesting based on the number of violations by issuing extremely conservative VaR forecasts with probability $1-\alpha$ and extremely permissive VaR forecast with probability α in a random fashion. This argument has been further explored in Holzmann and Eulert (2014), that have shown that traditional backtesting of VaR is insensitive with respect to increasing information.

The idea of backtesting VaR by means of scoring functions already emerged in Lopez (1999a) and Lopez (1999b), but never became popular among the practitioners. Moreover, Lopez (1999a) and Lopez (1999b) used a non-consistent scoring function, as it was noted e.g. by Bellini and Figa'-Talamanca (2007). More recently, a similar approach has been pursued in Wong (2010), that used the mean of the magnitude of the violations as a test statistic for VaR.

After the seminal paper of Artzner et al. (1999), academic research recognized some shortcomings of VaR and suggested to move to Expected Shortfall (ES) as a proper risk measure for computing capital requirements; this lead to the problem of backtesting ES, that has been considered for example by MacNeil and Frey (2000), Kerkhof and Melenberg (2004) and Acerbi and Szekely (2014).

A paper that had a big influence on the financial community and initiated an active stream of new literature on backtesting was Gneiting (2011), that introduced in the financial community the ideas of consistent scoring function and elicitable functional. A scoring function is consistent for a given statistical functional if the functional can be defined as the minimizer of the expected value of the score; a functional is elicitable if it admits a consistent scoring function. Consistent scoring functions give a natural way to compare different forecasts of the same risk measure.

The most common elicitable risk measures are VaR, the expectiles, and the couple (VaR, ES). Indeed, being a quantile of the distribution of the future losses, VaR is the minimizer of the expected value of a suitable piecwise linear score; expectiles are by definition the minimizers of a suitable piecewise-quadratic score; and it has been recently established by Acerbi and Szekely (2014) and Fissler and Ziegel (2016) that the couple (VaR, ES) jointly minimizes the expectation of a suitable two-variable scoring function.

Consistent scoring functions have been used mainly for comparative backtesting (Fissler et al., 2016, Nolde and Ziegel, 2016), with a procedure that is based on the Diebold-Mariano test. Comparative backtesting refers to a situation in which the statistical hypothesis under examination regards the relative merit of two different models.

The aim of the present paper is to investigate the use of scoring function for traditional backtesting of VaR and expectiles, by refining the ideas introduced in Lopez (1999) and Wong (2010). To this aim, we derive the asymptotic distribution of the empirical scores in the case of a normal i.i.d. sample and in the case of a uniform i.i.d. sample; both are asymptotically normal, altough the finite sample distribution shows a considerable departure from normality. The same idea may be applied for the couple (VaR, ES), but we are not pursuing it in the present work.

In order to apply the results for the asymptotic distribution of the empirical scores to backtesting usual econometric models, we suggest to follow a procedure based on the Probability Integral Transform (also known as 'realized p-values'), as it has been suggested in the present context by Berkowitz (2001) for VaR and by Kerkhof and Melenberg (2004) for ES.

The chapter is structured as follows. In Section 2.2 we introduce the basic notations. In Section 2.3 we derive asymptotic distributions of realized scores and identification functions and study their small sample distribution by means of simulations. In Section 2.4 we provide a simple example on simulated data that shows that backtesting with a realized score may have more power than backtesting with a realized identification function. In Section 2.5 we provide an example of backtesting standard AR(1)-Garch(1,1) models, while Section 2.6 concludes.

2.2 Notations and preliminaries

We denote with Y the random future loss associated with a financial position. The α-Value at Risk of Y is simply the α-quantile

$$v_\alpha(Y) = \inf\{t : F_Y(t) \geq \alpha\},$$

where $F_Y(t) = P(Y \leq t)$ and typically $\alpha = 0.95$ or $\alpha = 0.99$. As it is well known from quantile regression (see e.g. Koenker, 2005),

$$v_\alpha(Y) = \arg\min_{x\in\mathbb{R}} E[S^{(v)}(x, Y)],$$

for the piecewise linear scoring function

$$S^{(v)}(x, y) = \alpha(y - x)_+ + (1 - \alpha)(y - x)_-. \tag{2.1}$$

Newey and Powell (1987) introduced the expectiles e_α as the minimizers of the expected value of an asymmetric quadratic scoring function: for $Y \in L^2$,

$$e_\alpha(Y) = \arg\min_{x\in\mathbb{R}} E[S^{(e)}(x, Y)],$$

with

$$S^{(e)}(x, y) = \alpha(y - x)_+^2 + (1 - \alpha)(y - x)_-^2. \tag{2.2}$$

Expectiles are thus an asymmetric genereralization of the mean, that arise when $\alpha = 1/2$. Expectiles are becoming increasingly popular in the financial literature since they are the only elicitable coherent risk measures; for this characterization and further properties of expectiles we refer to Weber (2006), Bellini (2012), Bellini et al. (2014), Bellini and Bignozzi (2015), Bellini and Di Bernardino (2015), Delbaen et al. (2016). It has to be stressed that the scoring functions $S^{(v)}$ and $S^{(e)}$ are by no means the only ones whose expected values are minimized respectively in the quantiles and in the expectiles. Under mild technical conditions, Thomson (1979) and Saerens (2000) proved that the more general scoring function consistent with quantiles is of the form

$$S(x,y) = (1_{x\leq y} - \alpha)\,(g(x) - g(y)) \tag{2.3}$$

for some nondecreasing $g\colon \mathbb{R} \to \mathbb{R}$.

Gneiting (2011) proved that the more general scoring function consistent with expectiles is of the form

$$S(x,y) = |1_{x\leq y} - \alpha|\,\big(\phi(y) - \phi(x) - \phi'(x)(y-x)\big)\,, \tag{2.4}$$

where $\phi\colon \mathbb{R} \to \mathbb{R}$ is a convex function with right derivative ϕ'. Newey and Powell (1987) showed that for $Y \in L^1$ expectiles are the unique solutions of the equation

$$E[I^{(e)}(x,Y)] = 0,$$

where the identification function $I^{(e)}$ is given by

$$I^{(e)}(x,y) = \alpha(y-x)_+ - (1-\alpha)(y-x)_-. \tag{2.5}$$

By analogy with (2.5), we denote the identification function of quantiles with

$$I^{(v)}(x,y) = \alpha 1_{\{y>x\}} - (1-\alpha)1_{\{y<x\}}. \tag{2.6}$$

Let now $Y_k, k = 1,\dots,n$ be an i.i.d. sample. The empirical versions of the expected scores and expected scores are called respectively realized scores and realized identification functions and defined as follows:

$$\begin{aligned} \widehat{S}_n^{(v)}(x) &= \frac{1}{n}\sum_{k=1}^n S^{(v)}(x,Y_k), \\ \widehat{S}_n^{(e)}(x) &= \frac{1}{n}\sum_{k=1}^n S^{(e)}(x,Y_k), \\ \widehat{I}_n^{(v)}(x) &= \frac{1}{n}\sum_{k=1}^n I^{(v)}(x,Y_k), \\ \widehat{I}_n^{(e)}(x) &= \frac{1}{n}\sum_{k=1}^n I^{(e)}(x,Y_k). \end{aligned} \tag{2.7}$$

Table 2.1: Mean and variance of the realized score $S^{(v)(v_\alpha,Y)}$ in the normal (upper) and in the uniform (lower) case for different values of α.

α	0.5	0.05	0.01	0.001
v_α	0	-1.6449	-2.3263	-3.0902
$E[S^{(v)}]$	3.9894×10^{-1}	1.0314×10^{-1}	2.6652×10^{-2}	3.3671×10^{-3}
$Var[S^{(v)}]$	9.0845×10^{-2}	1.2698×10^{-2}	2.0053×10^{-3}	1.4337×10^{-4}
v_α	0.5	0.05	0.01	0.001
$E[S^{(v)}]$	1.2500×10^{-1}	2.3750×10^{-2}	4.9500×10^{-3}	4.9950×10^{-4}
$Var[S^{(v)}]$	5.2083×10^{-3}	1.8802×10^{-4}	8.1675×10^{-6}	8.3167×10^{-8}

In the present paper we focus for simplicity only on $S^{(v)}$ and $S^{(e)}$, but similar computations may be done for other consistent scoring functions of the families 2.3 and 2.4. Alternative choices or the simultaneous use of multiple scoring functions in analogy with the idea of the so called Murphy diagrams in Ehm et al. (2016) will be investigated in future work.

2.3 Asymptotic distributions of realized scores and identification functions

In this section we study the asymptotic distribution of the realized scores and identification functions (2.7) for normal and uniform i.i.d. samples. All the computations related to the present Section are reported in Appendix. Let $Y \sim N(0,1)$, and let ϕ, Φ and z_α be respectively its density, distribution function and α-quantile. Recalling the definition of $S^{(v)}$ and $S^{(e)}$ from (2.1) and (2.2), we have

$$
\begin{aligned}
E[S^{(v)}(v_\alpha, Y)] &= \phi(v_\alpha) \\
Var[S^{(v)}(v_\alpha, Y)] &= \alpha(1-\alpha)(1+v_\alpha^2) + (1-2\alpha)v_\alpha\phi(v_\alpha) - \phi^2(v_\alpha)
\end{aligned}
$$

If $Y \sim U[0,1]$, we get

$$
\begin{aligned}
E[S^{(v)}(v_\alpha, Y)] &= \frac{\alpha(1-\alpha)}{2} \\
Var[S^{(v)}(v_\alpha, Y)] &= \frac{\alpha^2(1-\alpha)^2}{12}
\end{aligned}
$$

We report some values of $E[S^{(v)}(v_\alpha, Y)]$ and $Var[S^{(v)}(v_\alpha, Y)]$ in Table 2.1.
We now consider the expectile scoring function (2.2). In the normal case we get

TABLE 2.2: Mean and variance of the realized score $S^{(e)}$ in the normal (upper) and in the uniform (lower) case for different values of α.

α	0.5	0.05	0.01	0.001
e_α	0	-1.1402	-1.7173	-2.4265
$E[S^{(e)}]$	0.5	1.6440×10^{-1}	5.2091×10^{-2}	8.4147×10^{-3}
$Var[S^{(e)}]$	0.5	8.6607×10^{-2}	1.6176×10^{-2}	1.4115×10^{-3}
e_α	0.5	1.8661×10^{-1}	9.1325×10^{-2}	3.0668×10^{-2}
$ES^{(e)}]$	4.1667×10^{-2}	1.1027×10^{-2}	2.7523×10^{-3}	3.1320×10^{-4}
$Var[S^{(e)}]$	1.3889×10^{-3}	9.7273×10^{-5}	6.0601×10^{-6}	7.8476×10^{-8}

$$\begin{aligned} E[S^{(e)}(e_\alpha, Y)] &= (1-2\alpha)e_\alpha\phi(e_\alpha) + (1+e_\alpha^2)[\alpha\overline{\Phi}(e_\alpha) + (1-\alpha)\Phi(e_\alpha)] \\ Var[S^{(e)}(e_\alpha, Y)] &= \left[3 + 6e_\alpha^2 + e_\alpha^4\right]\left\{\alpha^2 - 2\alpha\Phi(e_\alpha) + \Phi(e_\alpha)\right\} + \\ &\quad + (1-2\alpha)\left(5e_\alpha\phi(e_\alpha) + e_\alpha^3\phi(e_\alpha)\right) - E^2[S^{(e)}(e_\alpha, Y)]. \end{aligned}$$

In the uniform case the expectile is given by

$$e_\alpha = \begin{cases} \frac{\alpha - \sqrt{\alpha - \alpha^2}}{2\alpha - 1} & \text{if } \alpha \neq 1/2 \\ 1/2 & \text{if } \alpha = 1/2. \end{cases} \tag{2.8}$$

For $\alpha \neq 1/2$ we get

$$\begin{aligned} E[S^{(e)}(e_\alpha, Y)] &= \frac{\alpha(1-\alpha)(1-2\sqrt{\alpha(1-\alpha)})}{3\,(2\alpha-1)^2} \\ Var[S^{(e)}(e_\alpha, Y)] &= \frac{4\alpha^2\,(1-\alpha)^2\left(1-2\sqrt{\alpha(1-\alpha)}\right)^2}{45\,(2\alpha-1)^4}. \end{aligned}$$

If $\alpha = 1/2$ the expectile coincides with the mean, and it is easy to compute directly $E\left[S^{(e)}(e_{1/2}, Y)\right] = \frac{1}{24}$ and $Var\left[S^{(e)}(e_{1/2}, Y)\right] = \frac{1}{720}$.
Notice that for each $\alpha \in (0,1)$

$$Std[S^{(e)}(e_\alpha, Y)] = \frac{2}{\sqrt{5}} E[S^{(e)}(e_\alpha, Y)],$$

that shows that for a $U(0,1)$ i.i.d. sample the coefficient of variation of the score does not depend on α. We report some values of $E[S^{(e)}(e_\alpha, Y)]$ and $Var[S^{(e)}(e_\alpha, Y)]$ in Table 2.2. Under our assumptions, realized scores are time averages of i.i.d. variables, so from the law of large numbers and the central limit theorem they have an asymptotically normal distribution:

$$\widehat{S}_n^{(v)}(x) \to E[S^{(v)}(x, Y)] \text{ a.s.},$$

$$\sqrt{n}\left(\widehat{S}_n^v(x) - E[S^v(x, Y)]\right) \xrightarrow{d} N(0, Var[S^v(x, Y)]),$$

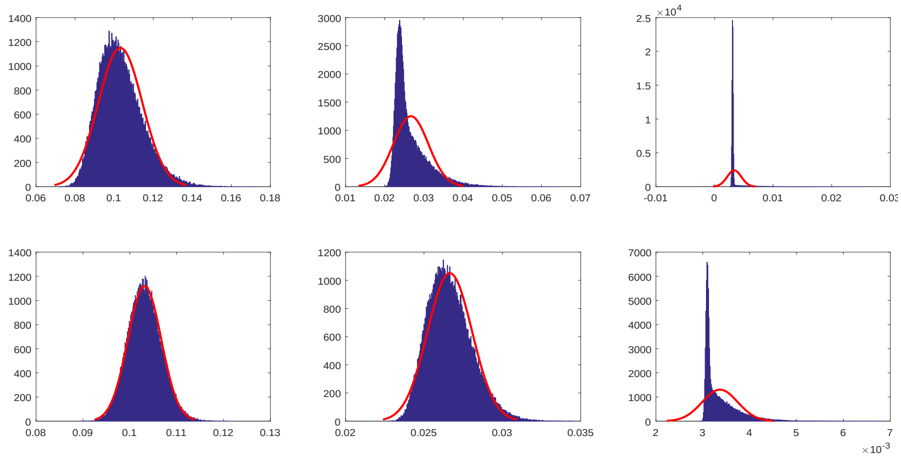

FIGURE 2.1: Distributions of 100000 realizations of the realized score $\widehat{S}_n^{(v)}$ for a normal i.i.d. sample with $n = 100$ (upper part) and $n = 1000$ (lower part), for $\alpha = 0.05, 0.01, 0.001$.

and similarly for $S^{(e)}$. Of course, asymptotic normality holds under much more general hypotheses, related to the ergodicity of the process Y_n. The whole point for constructing a backtesting procedure by means of realized scores is to be able to compute explicitly asymptotic variances, as we did in Tables 2.1 and 2.2.

In order to investigate the departure from normality of the distribution of $\widehat{S}_n^{(v)}$ and $\widehat{S}_n^{(e)}$, we resimulated $N = 100000$ trajectories of lenghts $n = 100$ and $n = 1000$, for $\alpha = 0.05, 0.01, 0.001$ (see Figures 2.1, 2.2 for the normal case and Figures 2.3, 2.4 for the uniform case). The normal expectiles have been computed by means of the R function `enorm`, while uniform expectiles by formula (2.8).

In the final part of the section we derive the asymptotic distributions of the realized identification functions $\widehat{I}_n^{(v)}$ and $\widehat{I}_n^{(e)}(x)$ under the same hypotheses of normal and uniform i.i.d samples. From the very definition of $I^{(v)}$ and $I^{(e)}$ in (2.5) and (2.6), it holds that

$$E[I^{(v)}(Y, v_\alpha)] = 0$$
$$E[I^{(e)}(Y, v_\alpha)] = 0$$

so $E[\widehat{I}_n^{(v)}] = 0$ and $E[\widehat{I}_n^{(e)}] = 0$. Moreover, notice that

$$\widehat{I}_n^{(v)}(v_\alpha) = \frac{1}{n}\sum_{k=1}^{n} I^{(v)}(v_\alpha, Y_k) = \frac{1}{n}\sum_{k=1}^{n} \alpha 1_{\{Y_k > v_\alpha\}} - (1-\alpha)1_{\{Y_k < v_\alpha\}} =$$
$$= \frac{[\alpha(n - N_V) - (1-\alpha)N_V]}{n} = \alpha - \frac{N_V}{n},$$

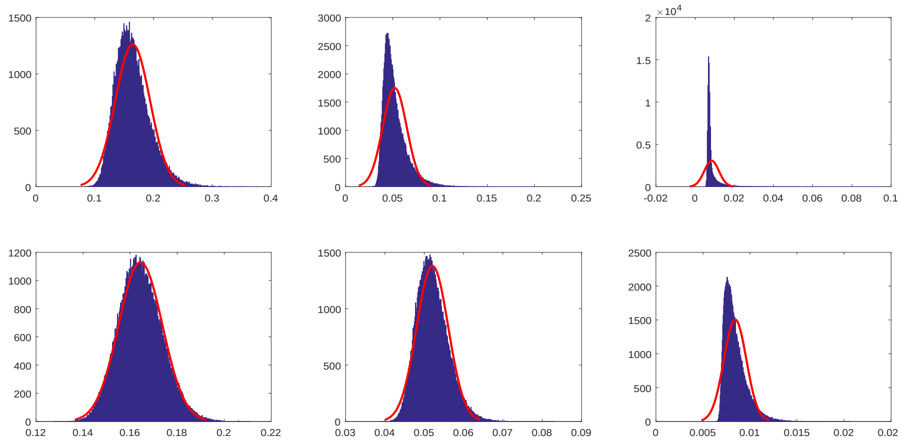

FIGURE 2.2: Distributions of 100000 realizations of the realized score $\widehat{S}_n^{(e)}$ for a normal i.i.d. sample with $n = 100$ (upper part) and $n = 1000$ (lower part), for $\alpha = 0.05, 0.01, 0.001$.

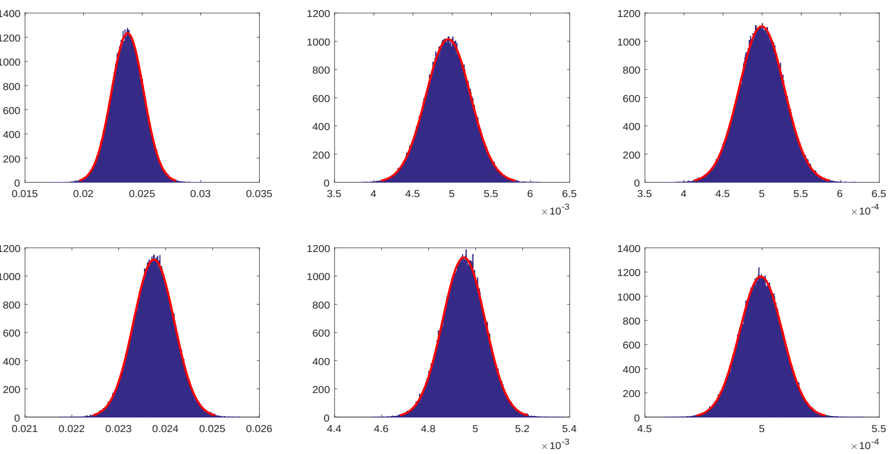

FIGURE 2.3: Distributions of 100000 realizations of the realized score $\widehat{S}_n^{(v)}$ for a uniform i.i.d. sample with $n = 100$ (upper part) and $n = 1000$ (lower part), for $\alpha = 0.05, 0.01, 0.001$.

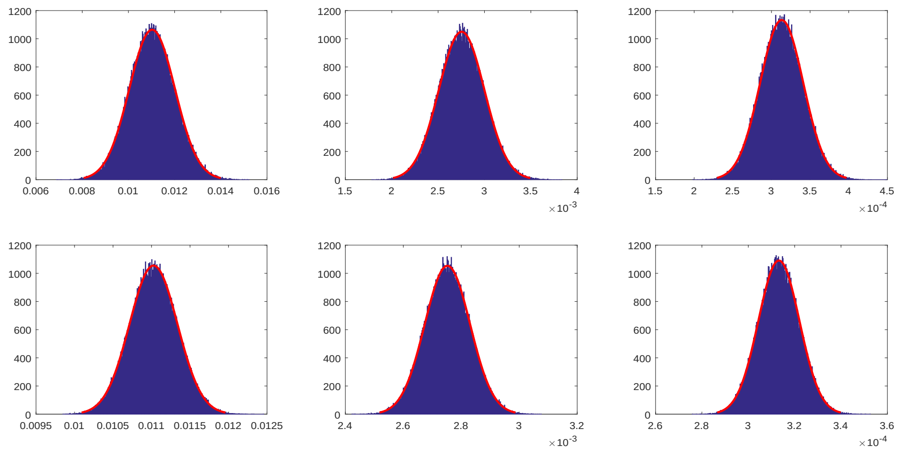

FIGURE 2.4: Distributions of 100000 realizations of the realized score $\widehat{S}_n^{(e)}$ for a uniform i.i.d. sample with $n = 100$ (upper part) and $n = 1000$ (lower part), for $\alpha = 0.05, 0.01, 0.001$.

where

$$N_V = \sum_{k=1}^{n} 1_{\{Y_k < v_\alpha\}}$$

is the number of violations. So in the case of VaR the realized identification function is simply equal to the number of violations, up to a linear transformations. Hence the exact finite sample distribution of the realized identification function is binomial; in the VaR case backtesting with the realized identification function is the traditional backtesting with the number of violations.

In the case of expectiles, if $Y \sim N(0, 1)$

$$Var[I^{(e)}(Y, e_\alpha)] = (1 + e_\alpha^2) \left[\alpha^2 + \Phi(e_\alpha)(1 - 2\alpha)\right] + e_\alpha \phi(e_\alpha) \left[1 - 2\alpha\right],$$

while if $Y \sim U(0, 1)$

$$Var[I^{(e)}(Y, e_\alpha)] = \frac{\alpha^2 (1 - e_\alpha)^3}{3} + \frac{(1 - \alpha)^2 e_\alpha^3}{3}.$$

Some numerical values are reported in Table 2.3. As before, we investigate the finite sample distribution of $\widehat{I}_n^{(e)}$ by means of risimulation for $n = 100$, 1000 and $\alpha = 0.05, 0.01, 0.001$ in the normal (Fig. 2.5) and in the uniform case (Fig. 2.6).

TABLE 2.3: Variance of the realized score $I^{(e)}$ in the normal (upper) and in the uniform (lower) case for different values of α.

α	0.5	0.05	0.01	0.001
e_α	0	-1.1402	-1.7173	-2.4265
$Var\left[I^{(e)}\right]$	2.5000×10^{-1}	5.5147×10^{-2}	1.2995×10^{-2}	1.5337×10^{-3}
e_α	0.5	1.8661×10^{-1}	9.1325×10^{-2}	3.0668×10^{-2}
$Var\left[I^{(e)}\right]$	-	2.4032×10^{-3}	2.7385×10^{-4}	9.8993×10^{-6}

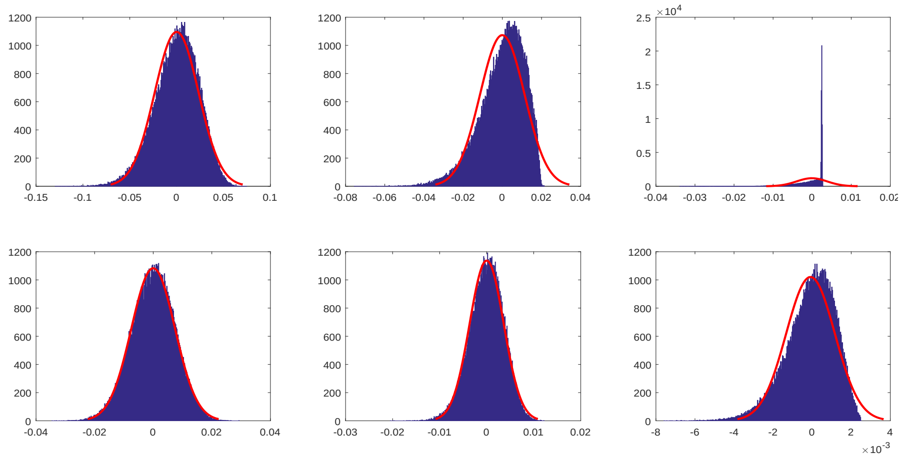

FIGURE 2.5: Distributions of 100000 realizations of the realized identification function $\widehat{I}_n^{(e)}$ for a normal i.i.d. sample with $n = 100$ (upper part) and $n = 1000$ (lower part), for $\alpha = 0.05, 0.01, 0.001$.

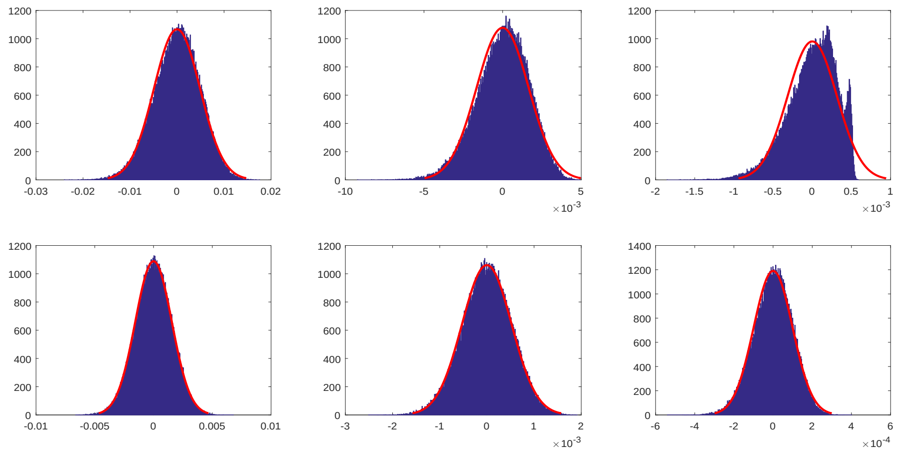

FIGURE 2.6: Distributions of 100000 realizations of the realized identification function $\widehat{I}_n^{(e)}$ for a uniform i.i.d. sample with $n = 100$ (upper part) and $n = 1000$ (lower part), for $\alpha = 0.05, 0.01, 0.001$.

2.4 Backtesting with realized scores and identification functions

The aim of this section is to provide a first example of backtesting with scoring functions on simulated data. The idea is very simple; the model used for forecasting VaR or expectiles does not pass the backtest if the realized score is too big. In order to find the rejection level of the test we will use the asymptotic distributions of the realized scores derived in the previous sections, and in particular the parameters in Table 2.2 and Table 2.3.

The simulation setting is the following: we assume that the true data generating process is known and given by $Y_i = N(\mu_i, 1)$, with Y_i independent, for a fixed vector of means μ_i, $i = 1, ..., 100$. The time-varying means μ_i have been randomly chosen at the beginning of the experiment by simulating from a zero mean normal distribution with three possible values of the standard deviation: 0.1, 0.5 and 1. This originated three vector of means, which we call respectively μ^A, μ^B and μ^C, that are kept fixed through the experiment and plotted in Fig. 2.7. We will refer to these as model A, B and C respectively.

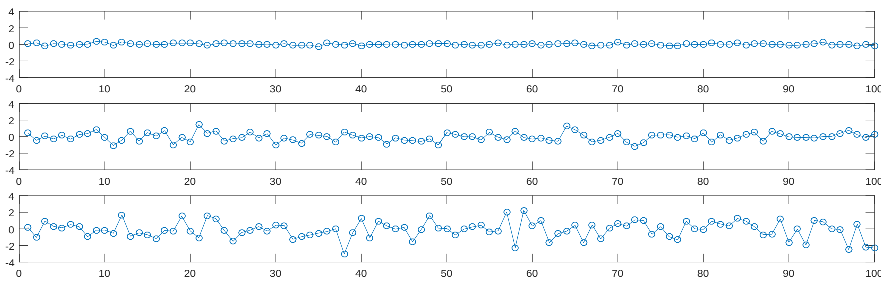

FIGURE 2.7: Vector of 100 time-varying means μ^A, μ^B and μ^C.

We consider a risk manager who wrongly believes that the data generating process is $Y_i = N(0,1)$, with Y_i i.i.d.; his forecasting model is not able to capture the real dynamic of the conditional mean. We would check two different backtesting procedures: the first one is based on counting the number of violations, the second one is based on considering the realized score and comparing it with its asymptotic distribution under the null hypothesis of a normal i.i.d. sample. Of course this example is highly stylized, but it can be made much more realistic by just interpreting the sequence Y_i as the residuals of a more complicate econometric model. Since in the example the forecasting distribution is stationary, the VaR forecasts are constant and equal to $z_{0.05} = -1.6449$. Under the null hypothesis that the forecasting model is correct, the number of violations N_V has a binomial distribution with $p = 0.05$ and $n = 100$. Since the number of violations is an integer variable, we cannot fix a threshold such that the level of the test is exactly 5%, as it would be costumary. So, we reject if $N_V \leq 1$ or $N_V \geq 10$, that corresponds to a level of the binomial test

$$P(N_V \leq 1) + P(N_V \geq 10) = 0.03708 + 0.02819 = 6,527\%.$$

Notice that we use a bilateral test, rejecting either if the number of violations is too low, or if the number of violation is too big. We penalize a model if it is too permissive (many violations), but also if it is too restrictive (too few violations). We believe that the aim of accurate risk management should not to avoid all violations (an objective that could require excessivingly high margins, perhaps restricting too severely banking activity) but to keep the number of violations under control from a statistical point of view. The threshold $\overline{S}$ of the test based on the realized score $\widehat{S}^{(v)}_{100}$ is chosen in order to satisfy

$$P(\widehat{S}^{(v)}_{100} > \overline{S}) = 6,527\%.$$

Using the asymptotic distribution derived in Section 2.3, we get

$$\overline{S} = 0.1202.$$

However, since as we saw the distribution of $\widehat{S}^{(v)}_{100,0.05}$ shows considerable departure from normality, we prefer to use a risimulated finite-sample distribution, that gives

$$\overline{S} = 0.1216.$$

In order to compare the empirical power of the test based on the number of violations with the test based on the realized score, we risimulate $N = 100000$ times models A, B and C, and as a check also the model with $\mu_i = 0\ \forall i$. We report in Table 2.4 the number of rejections of the forecasting model based on the number of violations (R1)

TABLE 2.4: Fraction of rejections of the correct model $\mu = 0$ and of the wrong models $\mu = \mu_A$, $\mu = \mu_B$, $\mu = \mu_C$ when backtesting with the number of violations (R1) and when backtesting with scoring functions (R2)

	$\mu = 0$	$\mu = \mu_A$	$\mu = \mu_B$	$\mu = \mu_C$
$R1$	$6,56\%$	$6,44\%$	$17,71\%$	$80,56\%$
$R2$	$6,66\%$	$6,98\%$	$41,76\%$	$97,60\%$

and the number of rejections based on the empirical score (R2). The results shows that backtesting with realized score has more power than backtesting with the number of violations in detecting the misspecification in the conditional mean. This seems quite intuitive, since the realized score contains also information on the magnitude of the violation, not only on the fact that a violation occurred. We repeat the same experiment in the case of a risk manager interested in forecasting the 5%-expectile. The true data generating processes are as in the preceding example, while the forecasted risk measure is now $e_{0.05}(Y) = -1.1402$. Our aim is now to compare backtesting with the empirical identification function $\widehat{I}_n^{(v)}$ with backtesting with the empirical score $\widehat{S}_n^{(e)}$. Since now both statistics have a continuous distribution, we fix a 5% level of the test. For the identification function we consider a bilateral test with tresholds I_L and I_U determined by

$$P(\widehat{I}_{100}^{(e)} < I_L) = 2,5\% \text{ and } P(\widehat{I}_{100}^{(e)} > I_U) = 2,5\%.$$

For the realized score we find $\overline{S}$ such that

$$P(\widehat{S}_{100}^{(e)} > \overline{S}) = 5\%.$$

As before, since the distribution of $\widehat{I}_n^{(v)}$ and $\widehat{S}_n^{(e)}$ shows a serious departure from normality, we determine the values of I_L, I_U and $\overline{S}$ by risimulation, obtaining

$$I_L = -0.0505, I_U = 0.0411, \overline{S} = 0.2184.$$

The empirical powers of the two different tests are reported in Table 2.5. As in the case of VaR, backtesting with scoring functions seem to have more power against misspecification in the mean then backtesting with identification functions. We stress that this is just a preliminary example that may be enlarged in several directions, but it seems to indicate a general principle that backtesting with scoring function is more sensitive than backtesting with identification functions, at least with regard to wrong conditional means.

Table 2.5: Fraction of rejections of the correct model $\mu = 0$ and of the wrong models $\mu = \mu_A$, $\mu = \mu_B$, $\mu = \mu_C$ when backtesting with realized identification function (R1) and when backtesting with realized score (R2)

	$\mu = 0$	$\mu = \mu_A$	$\mu = \mu_B$	$\mu = \mu_C$
$R1$	$5,00\%$	$4,98\%$	$23,11\%$	$98,71\%$
$R2$	$5,01\%$	$5,59\%$	$37,00\%$	$99,87\%$

2.5 A real data example

In the previous section we showed that it is possible to backtest a simple forecasting model by means of realized scores; moreover, the test based on realized score seemed to have more empirical power in detecting a misspecified mean. In this section we consider backtesting of a more realistic econometric model of Garch type, with normal or Student t innovations. We suggest two different approaches; in the first one, we determine the empirical distribution of the realized score under the null hypothesis that the forecasting model is correct by means of risimulations, as in Bellini and Di Bernardino (2017).

In the second approach we first perform a Probability Integral Transform on the realized logreturns

$$U_i = F_{Y_i}(Y_i),$$

and then test that the resulting sequence U_i is is i.i.d. with a $U(0, 1)$ distribution by computing the realized score and by comparing it with its asymptotic distribution for uniform i.i.d. variables that has been derived in Section 2.3. A similar idea have been suggested by Kerkhof and Melenberg (2004) for backtesting Expected Shortfall.

We consider the closing values of the SP500 Index from 03/01/2007 to 14/12/2012, and we compute the corresponding series of 1500 daily logreturns. On this series we estimate two models: an AR(1)-Garch(1,1) with normal innovations and an AR(1)-Garch(1,1) with t innovations. Estimations are performed on rolling windows of 500 logreturns, so we shave 1000 distributional forecast, related to days $501 - 1500$. We used the Econometrics Toolbox in Matlab; we noticed some numerical instabilities in the estimation of the degrees of freedom of the t innovations that have been overcome by occasionally switching the default settings to the interior-point algorithm when convergence problems arised. We forecasted VaR at the 5% level and expectiles at the 1% level.

The forecasted mean and standard deviations of the two models are reported in Fig. 2.8 and 2.9, the forecasted VaR and expectiles in Fig. 2.10 and 2.11, the time series and the distribution of the PIT in Fig. 2.12 and 2.13.

In order to derive the distribution of the realized scores $\widehat{S}^{(v)}_{1000}$ and $\widehat{S}^{(e)}_{1000}$ under the null hypothesis that the AR(1)-Garch(1,1) model with normal innovation is correct, we risimulated $N = 100000$ trajectories of length $n = 1000$. A similar approach has been followed for the model with t innovations. The results are reported in Table 2.6. We see that in the normal case the realized scores $\widehat{S}^{(v)}_{1000}$ and $\widehat{S}^{(e)}_{1000}$ are quite high with respect to their theoretical distribution, leading to very low p-values (respectively 0.00072 and 0.02164). On the contrary the model with t innovations is not rejected, since the p-values in the left part of Table 2.6 are respectively equal to 0.52471 and 0.6136. For the sake of clarity, we stres that this is not a comparative backtesting as in Nolde and Ziegel (2016), based on a Diebold-Mariano test on difference of realized scores, but we a separate backtest of the two models.

In Table 2.7 we show the results of the second backtesting approach, based on first applying the PIT. In the first two lines we report the realized values of the four statistics N_V, $\widehat{S}^{(v)}_{1000}$, $\widehat{I}^{(e)}_{1000}$, and $\widehat{S}^{(v)}_{1000}$ in the two models. In the last two lines we report the mean and the standard deviation of their asymptotic distribution.

As before, for VaR forecast (left part) and for expectile forecast (right part), the model with t innovations performs better, as expected. Indeed, it has in both cases lower realized scores (2.3801×10^{-2} vs 2.4497×10^{-2} and 2.6088×10^{-3} vs 2.7603×10^{-3}) and realized identification functions closer to their theoretical mean (the number of violations is 52 vs 60 and the expectiles identification function is -2.1496×10^{-4} vs -8.7402×10^{-4}).

We see from Table 2.7 that backtesting with only the number of violations would not reject the normal model at 5% level. On the contrary, backtesting with the realized score (second column) would reject the normal model, since the realized score 2.4497×10^{-2} is too high with respect to its theoretical mean. In the case of expectiles (third and forth column) the opposite situation arise: the normal model would be rejected by backtesting with the realized identification function but would pass the test based on the realized score. Indeed, the t-model performs much better in terms of realized score, but the performance of the normal model is not as bad as to produce a rejection.

2.6 Conclusions and directions for further research

The provided simulated and real data examples show that in principle it may be sensible to implement a backtesting procedure based on the realized value of the scoring function, with the idea of rejecting the model if the realized score is too big. Apparently, this approach captures the idea of backtesting by keeping into account also the magnitude

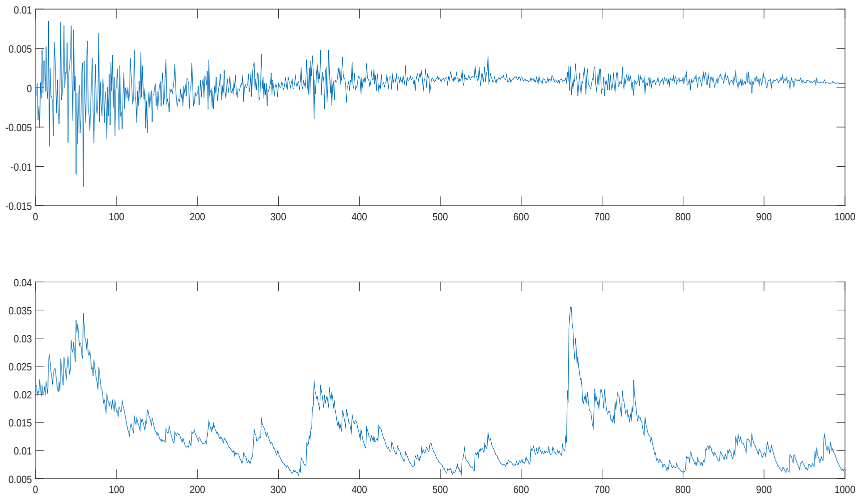

FIGURE 2.8: Forecasts of the mean (upper part) and standard deviations (lower part) of the AR(1)-Garch(1,1) model with normal innovations

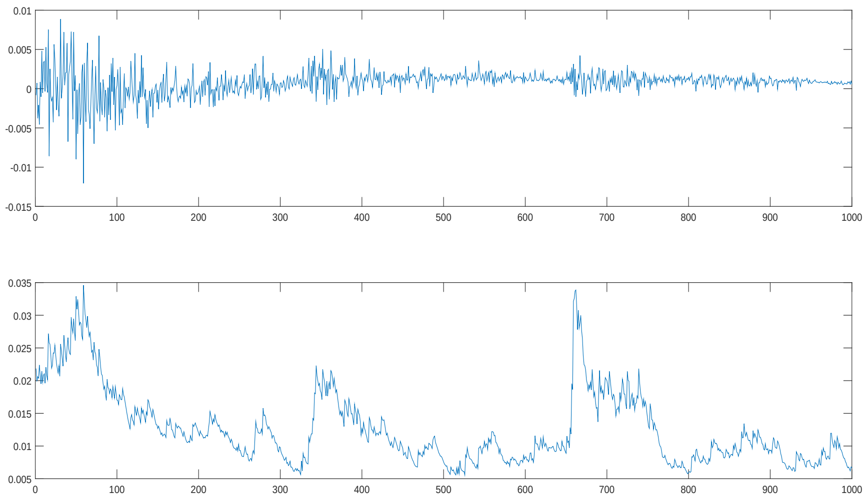

FIGURE 2.9: Forecasts of the mean (upper part) and standard deviations (lower part) of the AR(1)-Garch(1,1) model with t innovations

TABLE 2.6: Realized values of $\widehat{S}^{(v)}_{1000}$ and $\widehat{S}^{(e)}_{1000}$ in the normal (left) and t (right) models. Mean and standard deviation of $\widehat{S}^{(v)}_{1000}$ and $\widehat{S}^{(e)}_{1000}$ computed by risimulations in the normal (left) and t (right) models. In the last line we report the corresponding p-values.

	$\widehat{S}^{(v)}_{1000}$	$\widehat{S}^{(e)}_{1000}$	$\widehat{S}^{(v)}_{1000}$	$\widehat{S}^{(e)}_{1000}$
realized score	1.5044×10^{-3}	1.3790×10^{-5}	1.4883×10^{-3}	1.2826×10^{-5}
risimulated mean	1.3081×10^{-3}	1.0047×10^{-5}	1.5157×10^{-3}	1.5802×10^{-5}
risimulated std	4.9517×10^{-5}	1.0969×10^{-6}	8.0590×10^{-5}	1.0063×10^{-5}
p-value	0.00072	0.02164	0.52471	0.6136

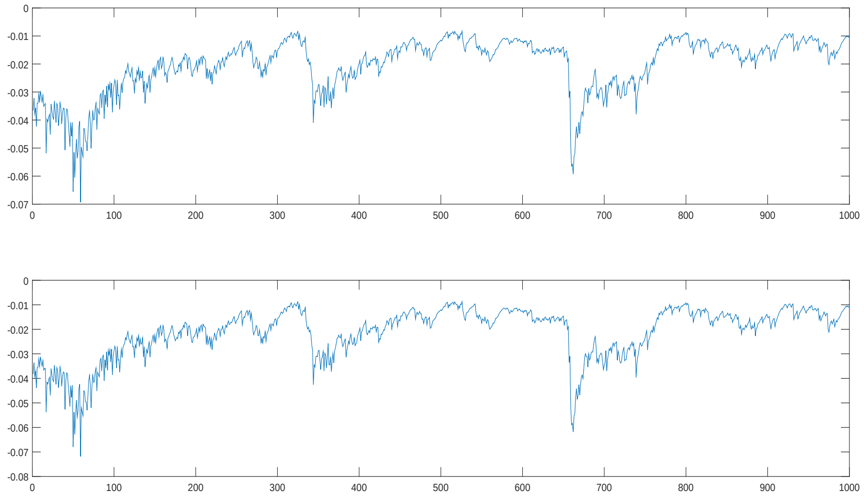

FIGURE 2.10: Forecasts of the VaR (upper part) and expectiles (lower part) of the AR(1)-Garch(1,1) model with normal innovations

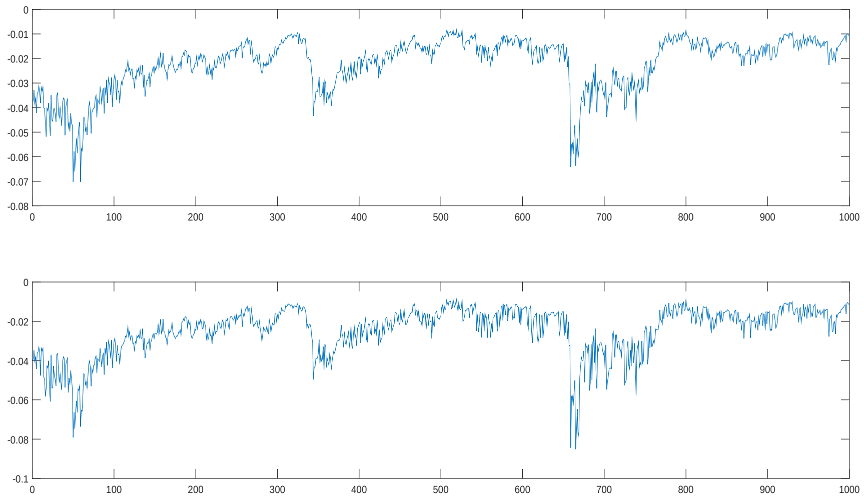

FIGURE 2.11: Forecasts of the VaR (upper part) and expectiles (lower part) of the AR(1)-Garch(1,1) model with t innovations

TABLE 2.7: Realized values of the four statistics N_V, $\widehat{S}^{(v)}_{1000}$, $\widehat{I}^{(e)}_{1000}$, $\widehat{S}^{(v)}_{1000}$ after performing the PIT in the normal and t model (upper part). Theoretical mean and standard deviation of the corresponding asymptotic distributions and p-values in the normal and t model (lower part).

	N_V	$\widehat{S}^{(v)}_{1000}$	$\widehat{I}^{(e)}_{1000}$	$\widehat{S}^{(e)}_{1000}$
normal	60	2.4497×10^{-2}	-8.7402×10^{-4}	2.7603×10^{-3}
t	52	2.3801×10^{-2}	-2.1496×10^{-4}	2.6088×10^{-3}
theoretical mean	50	2.3750×10^{-2}	0	2.7523×10^{-3}
theoretical std	6.8920	4.3361×10^{-4}	5.2331×10^{-4}	7.7847×10^{-5}
p-value normal	0.0734	0.0425	0.04744	0.45907
p-value t	0.3858	0.4532	0.34062	0.96736

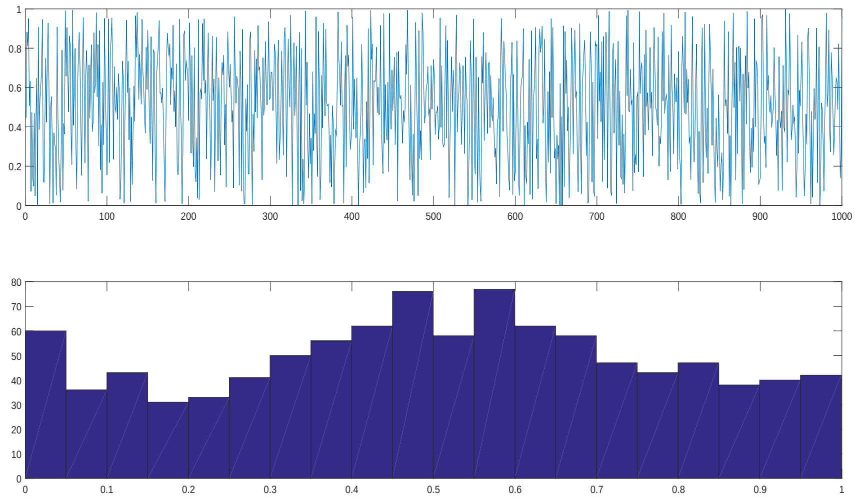

FIGURE 2.12: Time series (upper part) and distribution (lower part) of the Probability Integral Transform in the AR(1)-Garch(1,1) model with normal innovations

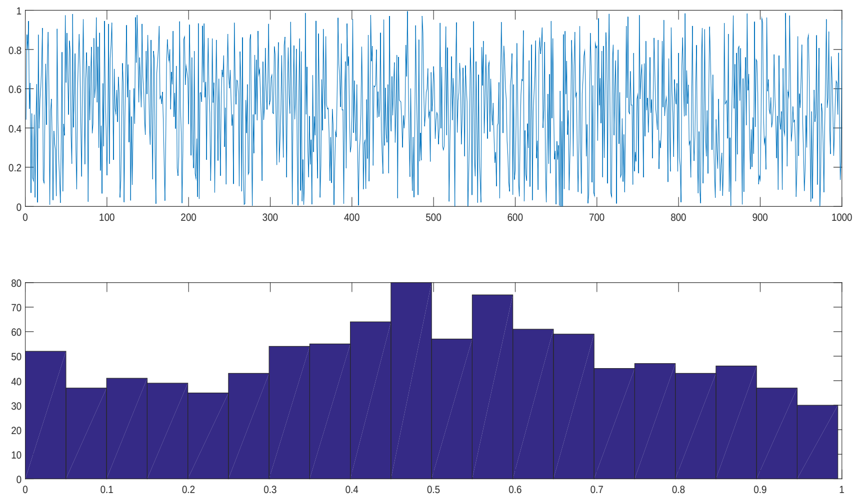

FIGURE 2.13: Time series (upper part) and distribution (lower part) of the Probability Integral Transform in the AR(1)-Garch(1,1) model with t innovations

of violations and not only their number; it may have more power than the traditional one, at least in simple examples.

Our use of the scoring function is different from Fissler et al. (2016) and Nolde and Ziegel (2016), since we are not testing hypotheses realtive to the comparative merit of two models, but we are backtesting a single model in isolation, as in the traditional Basel I framework. From a statistical point of view our approach requires the determination of the distribution of the realized scoring function under the null hypothesis that the forecasting model is correct; we compute the asymptotic distribution for i.i.d. normal and uniform samples, and we use risimulations for more complicated econometric models.

As directions for further research, we believe it would be important to find a more accurate approximation for finite-sample distributions of realized scores. Finally, the same ideas may be applied to the traditional backtesting of the couple (VaR, ES) by means of the consistent scoring functions studied in Acerbi and Szekely (2014) and Fissler and Ziegel (2016).

Bibliography

[1] Acerbi, C., Szekely., B., Backtesting expected shortfall, *Risk magazine*, 27, 76-81 (2014)

[2] Bao, Y., Lee, T., Saltoglu, B., Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check. *Journal of Forecasting*, 25, 101-128 (2006)

[3] Bellini, F., Figa'-Talamanca, G., Conditional tail behaviour and Value at Risk. *Quantitative Finance*, 7, 599-607 (2007)

[4] Bellini, F. Isotonicity results for geneneralized quantiles. *Statistics and Probability Letters*, 82, 2017-2024 (2012)

[5] Bellini, F., Klar, Bernard, Muller, A., Rosazza-Gianin, E. Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54, 41-48 (2014)

[6] Bellini, F., Di Bernardino, E. Risk management with expectiles, *The European Journal of Finance*, 23(6) 487-506 (2017)

[7] Delbaen, F., Bellini, F. Bignozzi, V., Ziegel, J. Risk measures with the CxLS property. *Finance and Stochastics*, 20(2), 433-453 (2016)

[8] Bellini, F., Bignozzi, V., On elicitable risk measures. *Quantitative Finance*, 15, 725-733 (2014)

[9] Berkowitz, J. Testing the accuracy of density forecasts, applications to risk management. *Journal of Business and Economics Statistics* 19, 465-474, (2001)

[10] Christoffersen, P., Backtesting, in *Enciclopedia of Quantitative Finance*, Wiley Online Library (2010)

[11] Campbell, S. D. A review of backtesting and backtesting procedures. *Finance and Economics Discussion Series*, Federal Reserve Board, 21, (2005)

[12] Christoffersen, P.F. Evaluating interval forecasts. *International Economic Review*, 39, 841-862 (1998)

[13] Christoffersen, P.F., Pelletier, D., Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2, 84-108 (2004)

[14] Corbetta, J., Peri, I. A new approach to backtesting and risk model selection. Preprint, *ssrn.com/abstract=2796253* (2016)

[15] Davis, M., Verification of internal risk measure estimates. *Statistics and Risk Modeling*, 33(3-4), 67-93 (2016)

[16] Delbaen,F., Bellini,F., Bignozzi,V., Ziegel, J., Risk measures with the CxLS property. *Finance and Stochastics*, 20, 2, 433-453 (2016)

[17] Ehm, W., Gneiting, T., Jordan, A., Krüger, F., Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society Series B*, 78, 3, 505-562 (2016)

[18] Fissler, T., Ziegel, J.F. Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4), 1680-1707 (2016).

[19] Fissler, T., Ziegel, J.F., Gneiting, T., Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting. *Risk Magazine*, January, 58-611 (2016)

[20] Gneiting, T. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106, 746-762 (2011)

[21] Holzmann, H., Eulert, M., The role of the information set for forecasting - with applications to risk management. *The Annals of Applied Statistics*, 8(1), 595-621 (2014)

[22] Kerkhof, J., Melenberg, B., Backtesting for Risk-Based Regulatory Capital. *Journal of Banking and Finance*, 28, 1845-1865 (2004)

[23] Kupiec, P. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3, 73-84 (1995)

[24] Lopez, J.A. Evaluation of Value-at-Risk Models, *Journal of Risk*, 1, 37-64 (1999)

[25] Lopez, J.A. Methods for Evaluating Value-at-Risk Models, *Federal Reserve Bank of San Francisco Economic Review*, 2, 3-17 (1999)

[26] Macneil, A.J, Frey, R. Estimation of tail-related risk measures for heteroschedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, 271-300 (2000)

[27] Nolde, N., Ziegel, J., Elicitability and backtesting. Preprint, *arxiv.org/abs/1608.05498* (2016)

[28] Saerens, M., Building cost functions minimizing to some summary statistics *IEEE Trans. Neural Network* 11, 1263-1271 (2000)

[29] Thomson, W., Eliciting production possibilities from a well-informed manager. *Journal of Economic Theory* 20(3), 360-380 (1979)

[30] Weber, S. Distribution-invariant risk measures, information and dynamic consistency. *Mathematical Finance* 16, 419-441 (2006)

[31] Wong, W.K. Backtesting value-at-risk based on tail losses. *Journal of Empirical Finance* 17, 526-538 (2010)

[32] Ziegel, J., Coherence and elicitability. *Mathematical Finance* 26(4), 901-918 (2016)

Appendix A

Mathematical calculation for Interlocking Directorates

We report in this Appendix the solution for all cases of profit mazimization problem and choice of interlock for symmetric and asymmetric cases.

Profit maximization problem for independent banks. Let's focus on bank 2 profit maximization problem for independent banks.

$$\Pi_2 = (1-\theta_2)[\mu(1-\theta_1(c_H,\mu)) + (1-\mu)(1-\theta_1(c_L,\mu))]p(R-1) + \theta_2(R-1) - c_H\frac{\theta_2^2}{2}.$$

A reaction function of bank 1 is:

$$\theta_2^*(\mu, c_H) = \frac{(R-1)}{c_H}\{1 - p[\mu(1-\theta_1^*(c_H,\mu)) + (1-\mu)(1-\theta_1^*(c_L,\mu))]\}$$

Given that the bankers' problems are symmetric, the system of reaction functions, that offends the optimal level of banker's effort.

$$\begin{cases} \theta_2^*(\mu,c_H) &= \frac{(R-1)}{c_H}\{1-p[\mu(1-\theta_1^*(c_H,\mu)) + (1-\mu)(1-\theta_1^*(c_L,\mu))]\}; \\ \theta_2^*(\mu,c_L) &= \frac{(R-1)}{c_L}\{1-p[\mu(1-\theta_1^*(c_H,\mu)) + (1-\mu)(1-\theta_1^*(c_L,\mu))]\}; \\ \theta_1^*(c_H,\mu) &= \frac{(R-1)}{c_H}\{1-p[\mu(1-\theta_2^*(\mu,c_H)) + (1-\mu)(1-\theta_2^*(\mu,c_L))]\}; \\ \theta_1^*(c_L,\mu) &= \frac{(R-1)}{c_L}\{1-p[\mu(1-\theta_2^*(\mu,c_H)) + (1-\mu)(1-\theta_2^*(\mu,c_L))]\}. \end{cases}$$

From the system and the problem symmetry, i.e. $\begin{cases} \theta_2^*(\mu,c_H) = \theta_1^*(c_H,\mu); \\ \theta_2^*(\mu,c_L) = \theta_1^*(c_L,\mu). \end{cases}$

$$\begin{aligned}
\frac{\theta_2^*(\mu, c_H)}{\theta_2^*(\mu, c_L)} &= \frac{c_L}{c_H}; \\
\theta_2^*(\mu, c_H) &= \theta_2^*(\mu, c_L)\frac{c_L}{c_H}; \\
\theta_2^*(\mu, c_H) &= \frac{[R-1](1-p)}{c_H - p[R-1][\mu + (1-\mu)\frac{c_H}{c_L}]}.
\end{aligned}$$

So the Static Bayesian equilibrium is

$$\begin{cases}
\theta_2^{IB,IB}(\mu, c_H) = \frac{[R-1](1-p)}{c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}]}; \\
\theta_2^{IB,IB}(\mu, c_L) = \frac{c_L[R-1](1-p)}{c_H(c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}])}; \\
\theta_1^{IB,IB}(c_H, \mu) = \frac{[R-1](1-p)}{c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}]}; \\
\theta_1^{IB,IB}(c_L, \mu) = \frac{c_L[R-1](1-p)}{c_H(c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}])}.
\end{cases}$$

with $p \in [0.1]; R > 1; c_k > 0$ and $\theta_k^{IB,IB} \in [0,1]$, $\forall k = \{1,2\}$ and $\ell < \frac{p}{R}; RL > 1; L > \frac{p}{R}$. Solution is $(\theta_1^{IB,IB}, \theta_2^{IB,IB})$ for each couple of types.

Profit maximization problem for Symmetric ID.

General case Let's focus on bank 1 profit maximization problem for Symmetric ID.

$$\max_{\theta_2} \Pi_2 = (1-\theta_1)(1-\theta_2)p(R-1) + \theta_2(R-1) - c_1\frac{\theta_2^2}{2}.$$

Given that the bankers' problems are symmetric, the system of reaction functions, that offends the optimal level of banker's effort.

$$\begin{cases}
\theta_2^{IL,IL} = \frac{(R-1)}{c_1}[1 - p(1-\theta_1^{IL,IL})]; \\
\theta_1^{IL,IL} = \frac{(R-1)}{c_2}[1 - p(1-\theta_2^{IL,IL})].
\end{cases}$$

$$\begin{aligned}
\theta_2^{IL,IL} &= \frac{(R-1)}{c_1}(1-p) + \frac{(R-1)^2}{c_1c_2}p[1-p(1-\theta_2^{IL,IL})]; \\
\theta_2^{IL,IL} &= \frac{(R-1)}{c_1}(1-p) + \frac{(R-1)^2}{c_1c_2}p(1-p) + \frac{(R-1)^2}{c_1c_2}p^2\theta_2^{IL,IL}; \\
\theta_2^{IL,IL} &= \frac{(R-1)(1-p)(c_2 + (R-1)(1-p))}{c_1c_2 - (R-1)^2p^2}.
\end{aligned}$$

The Nash equilibrium conditional on the observed effort cost, is

$$\begin{cases} \theta_2^{IL,IL}(c_1,c_2) &= \frac{(1-p)(R-1)(c_1+p(R-1))}{c_1c_2-p^2(R-1)^2}; \\ \theta_1^{IL,IL}(c_1,c_2) &= \frac{(1-p)(R-1)(c_2+p(R-1))}{c_1c_2-p^2(R-1)^2}. \end{cases}$$

where $p \in [0,1]$ is probability of being on the "good" state of the world, $R > 1$ is positive return of the unit; $c_1, c_2 > 0$ is the cost of the effort and $\ell < \frac{p}{R}$; $RL > 1$; $L > \frac{p}{R}$. In addition, notice that all the efforts $\theta^{IL,IL}$ are between 0 and 1.

Symmetric cost case Assume $c_1 = c_2 = c$, when c can be either $c = c_H$ or $c = c_L$.

The equilibrium efforts are symmetric $\theta_2^{IL,IL} = \theta_1^{IL,IL} = \theta^{IL,IL}$ and given by:

$$\theta^{IL,IL} = \frac{(1-p)(R-1)}{c-p(R-1)};$$
$$1-\theta^{IL,IL} = \frac{c-(R-1)}{c-p(R-1)}.$$

The equilibrium profit in this case is:

$$\begin{aligned}\Pi^{IL,IL}(c,c) &= \frac{(c-(R-1))^2}{(c-p(R-1))^2}p(R-1) + \frac{(R-1)^2(1-p)}{c-p(R-1)} - \frac{c}{2}\frac{(c-(R-1))^2}{(c-p(R-1))^2} \\ &= \frac{1}{(c-p(R-1))^2}[c^3p(R-1) - 0.5c^3 + p^2(R-1)^3 + c(R-1)^2(0.5-3p)]\end{aligned}$$

Collusion When the banks are collude (or merge) the banker maximizes joint profits and chooses both efforts. Symmetric information and coordination on efforts.

$$\begin{aligned}\max_{\theta_1,\theta_2}(\Pi_1+\Pi_2) &= (1-\theta_1)(1-\theta_2)p(R-1) + \theta_1(R-1) - c_1\frac{\theta_1^2}{2} \\ &+ (1-\theta_2)(1-\theta_1)p(R-1) + \theta_2(R-1) - c_2\frac{\theta_2^2}{2}\end{aligned} \tag{A.1}$$

where we derive profit for a generic cost c_1. It stands either c_H or c_L. by maximazing $\Pi_1 + \Pi_2$ with respect to θ_2 we derive the reaction function:

$$\theta_2^C = \frac{(R-1)}{c_1}[1-2p(1-\theta_1^C)]. \tag{A.2}$$

At the optimum the efforts of both banks are

$$\begin{cases} \theta_2^C = \frac{(R-1)(1-2p)[c_2+2(R-1)p]}{c_1c_2-4p^2(R-1)^2}; \\ \theta_1^C = \frac{(R-1)(1-2p)[c_1+2(R-1)p]}{c_1c_2-4p^2(R-1)^2}. \end{cases}$$

with $p \in [0.1]; R > 1; c_k > 0$ and $\theta_k^C \in [0,1]$, $\forall k = \{1,2\}$ and $\ell < \frac{p}{R}; L > \frac{p}{R}$; $RL > 1$.

Symmetric cost case. Assume $c_1 = c_2 = c$ the equilibrium effort for the bank 1 is given by:

$$\theta^C = \frac{(R-1)(1-2p)}{c-2p(R-1)} \tag{A.3}$$

$$1-\theta^C = \frac{c-(R-1)}{c-2p(R-1)} \tag{A.4}$$

The profit in this case is:

$$\Pi_2^C = \frac{(R-1)[2pc^2 - 6pc(R-1) - 2p(R-1)^2 + c(R-1) + 8p^2(R-1)^2 - 4p^2c(R-1)]}{2(c-2p(R-1))^2}$$

The welfare implications In this section we provide proof for Lemma 3.

Proof. Recall

$$\theta_1^{IB}(c_H,\mu) = \frac{(R-1)(1-p)}{c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}]}$$

$$\theta_1^{IL}(c_1,c_2) = \frac{(1-p)(R-1)(c_2+p(R-1))}{c_1c_2-p^2(R-1)^2}$$

Note that $c_1 = c_H$ by construction.

$$\frac{(R-1)(1-p)}{c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}]} > \frac{(1-p)(R-1)(c_2+p(R-1))}{c_Hc_2-p^2(R-1)^2}$$

$$\frac{1}{c_H - p[R-1][\mu+(1-\mu)\frac{c_H}{c_L}]} > \frac{(c_2+p(R-1))}{c_Hc_2-p^2(R-1)^2}$$

$$-p^2(R-1)^2 > c_Hp(R-1) - c_2p(R-1)[\mu+(1-\mu)\frac{c_H}{c_L}] - p^2(R-1)^2[\mu+(1-$$

$$p(R-1)[-c_L(1-\mu)+(1-\mu)c_H] > [c_Hc_L - \mu c_Lc_2 - (1-\mu)c_Hc_2]$$

Since in (IL, IL) case, each banker observes the type of the banker. We will proceed with two cases. Case 1: $c_2 = c_H$

$$
\begin{aligned}
p(R-1)(1-\mu)(c_H - c_L) &> [c_H c_L - \mu c_L c_H - (1-\mu)c_H^2] \\
p(R-1) &> -c_H
\end{aligned}
$$

It's always true, $c_2 = c_H$ implies $\theta^{IB,IB} > \theta^{IL,IL}$ $\forall$ $c_H, c_L \in [0.1]$, $p \in [0.1]$, $R > 1$. Case 1: $c_2 = c_L$

$$
\begin{aligned}
p(R-1)(1-\mu)(c_H - c_L) &> [c_H c_L - \mu c_L^2 - (1-\mu)c_H c_L] \\
p(R-1)(1-\mu) &> \mu c_L
\end{aligned}
$$

So, If $c_2 = c_L$ then $\theta^{IB,IB} > \theta^{IL,IL}$ $\forall c_L < \frac{1-\mu}{\mu} p(R-1)$ $p \in [0.1]$, $R > 1$, $\mu \in (0.1)$. The same analysis can be done for $\theta_2^{IB,IB}$ and $\theta_2^{IL,IL}$ and it will give the same result, because banks are symmetric. □

The equilibrium with interlocking

Asymmetric interlocking directorates

General case Assume that bank 1 is an independent bank (IB), while bank 2 interlocks (IL), i.e. the executive in bank 2 sits in the board of bank 1, but the reverse is not true.

Bank 2, who has full information, chooses its effort:

$$
\max_{\theta_2} \Pi_2 = [1 - \theta_1(c_1, \mu)](1 - \theta_2)p(R-1) + \theta_2(R-1) - c_2 \frac{\theta_2^2}{2}
$$

$$
\theta_2^*(c_1, c_2) = \frac{(R-1)}{c_2}[1 - p(1 - \theta_1^*(c_1, \mu))]
$$

Bank 1, not having info on bank 2 cost, assumes $E(\theta_2^*) = \mu\theta_2^*(c_1, c_H) + (1-\mu)\theta_2^*(c_1, c_L)$.

$$
\max_{\theta_1} \Pi_1 = (1 - \theta_1)[(1 - E(\theta_2^*))p + \theta_1](R-1) - c_1 \frac{\theta_1^2}{2}
$$

$$
\theta_1^*(c_1, \mu) = \frac{(R-1)}{c_1}[1 - p(1 - E(\theta_2^*))]
$$

At the Bayes Nash asymmetric equilibrium the efforts of both bankers are given by the solution to the following system of equations:

$$\begin{aligned}
\theta_2^*(c_1, c_H) &= \frac{(R-1)}{c_H}[1 - p(1 - \theta_1^*(c_1, \mu))]; \\
\theta_2^*(c_1, c_L) &= \frac{(R-1)}{c_L}[1 - p(1 - \theta_1^*(c_1, \mu))]; \\
\theta_1^*(c_1, \mu) &= \frac{(R-1)}{c_1}[1 - p(1 - E(\theta_2^*))].
\end{aligned}$$

where $E(\theta_2^*) = \mu\theta_2^*(c_1, c_H) + (1-\mu)\theta_2^*(c_1, c_L)$.

$$E(\theta_2^*) = (R-1)(\frac{c_H + \mu(c_L - c_H)}{c_L c_H})[1 - p(1 - \theta_1^*(c_1, \mu))]$$

Then

$$\begin{aligned}
\theta_1^{IB,IL}(c_1, \mu) &= \frac{(R-1)(1-p)[1 + p(R-1)\bar{c}]}{c_1 - p^2(R-1)^2\bar{c}} \\
\theta_2^{IB,IL}(c_1, c_H) &= \frac{(R-1)(1-p)[c_1 - 2p^2(R-1)^2\bar{c} + p(R-1)]}{c_H[c_1 - p^2(R-1)^2\bar{c}]}; \\
\theta_2^{IB,IL}(c_1, c_L) &= \frac{(R-1)(1-p)[c_1 - 2p^2(R-1)^2\bar{c} + p(R-1)]}{c_L[c_1 - p^2(R-1)^2\bar{c}]},
\end{aligned}$$

where $\bar{c} = \frac{c_H + \mu(c_L - c_H)}{c_L c_H}$, $p \in [0,1]$ is probability of being on the "good" state of the world, $R > 1$ is positive return of the unit; $c_1, c_2 > 0$ is the cost of the effort and $\ell < \frac{p}{R}$, $L > \frac{p}{R}$. In addition, notice that all the efforts $\theta^{IB,IL}$ are between 0 and 1.

Symmetric cost case Assume $c_1 = c_2 = c$, when c can be either $c = c_H$ or $c = c_L$.

The equilibrium effort for bank 1 is given by:

$$\begin{aligned}
\theta_1^{IB,IL}(c, \mu) &= \frac{(1-p)(R-1)[1 + p(R-1)\frac{\bar{c}}{c_H c_L}]}{c - p^2(R-1)^2\frac{\bar{c}}{c_H c_L}} \\
1 - \theta_1^{IB,IL}(c, \mu) &= \frac{c - p^2(R-1)^2\frac{\bar{c}}{c_H c_L} - (1-p)(R-1)[1 + p(R-1)\frac{\bar{c}}{c_H c_L}]}{c - p^2(R-1)^2\frac{\bar{c}}{c_H c_L}}
\end{aligned}$$

where $\bar{c} = \mu c_L + (1-\mu)c_H$.

The equilibrium effort of bank1 must be substituted into the two possible efforts of bank

2 given by

$$\theta_2^{IB,IL}(c,c_H) = \frac{(R-1)}{c_H}[1-p(1-\theta_1^{IB,IL}(c,\mu))]$$
$$\theta_2^{IB,IL}(c,c_L) = \frac{(R-1)}{c_L}[1-p(1-\theta_1^{IB,IL}(c,\mu))]$$

where $E(\theta_2^{IB,IL}) = \mu\theta_2^{IB,IL}(c,c_H) + (1-\mu)\theta_2^{IB,IL}(c,c_L)$.

$$E(\theta_2^{IB,IL}) = \frac{(R-1)(1-p)\frac{\bar{c}}{c_Lc_H}(c+(R-1)p)}{c-p^2(R-1)^2\frac{\bar{c}}{c_Lc_H}}$$
$$1-E(\theta_2^{IB,IL}) = \frac{c-p^2(R-1)^2\frac{\bar{c}}{c_Lc_H}-(R-1)(1-p)\frac{\bar{c}}{c_Lc_H}(c+(R-1)p)}{c-p^2(R-1)^2\frac{\bar{c}}{c_Lc_H}}$$

Then we can substitute these equilibrium efforts into the equilibrium profit of bank 1 given by:

$$\Pi_1^{IB,IL}(c,\mu) = (1-\theta_1^{IB,IL})(1-E(\theta_2^{IB,IL}))p(R-1) + \theta_1^{IB,IL}(R-1) - c_1\frac{(\theta_1^{IB,IL})^2}{2}$$

$$\begin{aligned}\Pi_1^{IB,IL}(c,\mu) &= \frac{c-p^2(R-1)^2\frac{\bar{c}}{c_Hc_L}-(1-p)(R-1)[1+p(R-1)\frac{\bar{c}}{c_Hc_L}]}{(c-p^2(R-1)^2\frac{\bar{c}}{c_Hc_L})^2}\\ &\times\ [c-p^2(R-1)^2\frac{\bar{c}}{c_Lc_H}-(R-1)(1-p)\frac{\bar{c}}{c_Lc_H}(c+(R-1)p)]p(R-1)\\ &+\ \frac{c(R-1)-p^2(R-1)^3\frac{\bar{c}}{c_Hc_L}-(1-p)(R-1)^2[1+p(R-1)\frac{\bar{c}}{c_Hc_L}]}{(c-p^2(R-1)^2\frac{\bar{c}}{c_Hc_L})}\\ &-\ \frac{c}{2}\frac{(c-p^2(R-1)^2\frac{\bar{c}}{c_Hc_L}-(1-p)(R-1)[1+p(R-1)\frac{\bar{c}}{c_Hc_L}])^2}{(c-p^2(R-1)^2\frac{\bar{c}}{c_Hc_L})^2}\end{aligned}$$

Condition for the existence of Subgame Perfect Equilibria (IL,IL) is a Nash equilibrium iff:
$\Pi_1^{IL,IL}(c_1,c_2) > \Pi_1^{IB,IL}(c_1,\mu)$ and $\Pi_2^{IL,IL}(c_1,c_2) > \Pi_2^{IL,IB}(\mu,c_2)$. Let's find condition under which the inequality $\Pi_2^{IL,IL}(c_1,c_2) > \Pi_2^{IL,IB}\mu,c_2)$ holds, given the symmetry of the other condition

This equivalent to

$$\begin{aligned}\Pi_2^{IL,IL}(c_1,c_2) - \Pi_2^{IL,IB}(\mu,c_2) &= (1-p)(R-1)(\theta_2^{IL,IL}-\theta_2^{IL,IB}) - p(R-1)(\theta_1^{IL,IL}-E(\theta_1^{IL,IB}))\\ &+p(R-1)(\theta_2^{IL,IL}\theta_1^{IL,IB}-\theta_2^{IL,IB}E(\theta_1^{IL,IB})) - \frac{c_2}{2}(\theta_2^{IL,IL}-\theta_2^{IL,IB})(\theta_2^{IL,IL}+\theta_2^{IL,IB}) > 0\end{aligned}$$

We calculate separately

- $\theta_2^{IL,IL} - \theta_2^{IL,IB} = \frac{(1-p)(R-1)^2 p(1-c_2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))(p(R-1)+c_1)}{(c_1c_2-p^2(R-1)^2)(c_1-p^2(R-1)^2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))}$
- $\theta_1^{IL,IL} - E(\theta_1^{IL,IB}) = \frac{c_1(1-p)(R-1)(c_1+p(R-1))(1-c_2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))}{(c_1c_2-p^2(R-1)^2)(c_1-(R-1)^2p^2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))}$
- $\theta_2^{IL,IL}\theta_1^{Il,IL} - \theta_2^{IL,IB}E(\theta_1^{IL,IB}) = \frac{(1-p)^2(R-1)^2(c_1+p(R-1))(1-c_2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))}{(c_1c_2-p^2(R-1)^2)^2(c_1-p^2(R-1)^2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))^2} \times (c_1^2c_2 - p^4(R-1)^4(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}) + c_1^2p(R-1)(1+c_2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L})) - 2c_1p^3(R-1)^3(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))$
- $(\theta_2^{IL,IL})^2 - (\theta_2^{IL,IB})^2 = \frac{(1-p)^3(R-1)^3(1-c_2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L})))}{(c_1c_2-p^2(R-1)^2)^2(c_1-p^2(R-1)^2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))^2} \times [2c_1c_2(c_1 + p(R-1)) + p(R-1)(1+c_2(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L}))(c_1^2 - p^2(R-1)^2) - 2p^3(R-1)^3(c_1 - p(R-1)(\mu\frac{1}{c_H}+(1-\mu)\frac{1}{c_L})]$

Appendix B

Mathematical calculation for Backtesting with realized scores

B.1 Appendix

We report in this Appendix the explicit computation of the expected values and variances of piecewise linear and quadratic scores and identification function in the normal and uniform cases. That is, we compute the quantities $E[S^{(v)}(x,Y)]$, $E[S^{(e)}(x,Y)]$, $E[I^{(v)}(x,Y)]$, $E[I^{(e)}(x,Y)]$ and the corresponding variances when $Y \sim N(0,1)$ and $Y \sim U(0,1)$. We denote with $\phi(x)$, $\Phi(x)$ and $\overline{\Phi}(x)$ the density, the cumulative and the retro-cumulative function of a standard Normal r.v.
The following identities will be repeatedly used:

$$\int_{-\infty}^{x} y\phi(y)dy = -\phi(x), \int_{x}^{+\infty} y\phi(y)dy = \phi(x)$$
$$\int_{-\infty}^{x} y^2\phi(y)dy = \Phi(x) - x\phi(x), \int_{x}^{+\infty} y^2\phi(y)dy = \overline{\Phi}(x) + x\phi(x)$$
$$\int_{-\infty}^{x} y^3\phi(y)dy = -x^2\phi(x) - 2\phi(x), \int_{x}^{+\infty} y^3\phi(y)dy = x^2\phi(x) + 2\phi(x)$$
$$\int_{-\infty}^{x} y^4\phi(y)dy = -x^3\phi(x) + 3\left[\Phi(x) - x\phi(x)\right]$$
$$\int_{x}^{+\infty} y^4\phi(y)dy = x^3\phi(x) + 3\left[\overline{\Phi}(x) + x\phi(x)\right]$$

Piecewise linear score, normal case

$$\begin{aligned}
E[S^{(v)}(x,Y)] &= \alpha\int_x^{+\infty}(y-x)\phi(y)dy+(1-\alpha)\int_{-\infty}^x(x-y)\phi(y)dy\\
&= \alpha[\phi(x)-x\overline{\Phi}(x)]+(1-\alpha)[x\Phi(x)+\phi(x)]\\
&= \phi(x)+x[\Phi(x)-\alpha].
\end{aligned}$$

When $x=z_\alpha$, we get

$$E[S^{(v)}(z_\alpha,Y)]=\phi(z_\alpha).$$

$$\begin{aligned}
E[S^{(v)}(x,Y)^2] &= \int_{-\infty}^{+\infty}\left[\alpha(y-x)_+ +(1-\alpha)(y-x)_-\right]^2\phi(y)dy\\
&= \alpha^2\int_x^{+\infty}(y-x)^2\phi(y)dy+(1-\alpha)^2\int_{-\infty}^x(y-x)^2\phi(y)dy\\
&= \alpha^2\left\{\overline{\Phi}(x)-x\phi(x)+x^2\overline{\Phi}(x)\right\}+\\
&\quad +(1-\alpha)^2\left\{\Phi(x)+x\phi(x)+x^2\Phi(x)\right\}\\
&= \left[1+x^2\right]\left\{\alpha^2+(1-2\alpha)\Phi(x)\right\}+(1-2\alpha)x\phi(x).
\end{aligned}$$

When $x=z_\alpha$, we get

$$\begin{aligned}
E[S^{(v)}(x,Y)^2] &= \left[1+z_\alpha^2\right]\left\{\alpha-\alpha^2\right\}+(1-2\alpha)z_\alpha\phi(z_\alpha)\\
Var[(S^{(v)}(z_\alpha,Y)] &= E[S^{(v)}(x,Y)^2]-\phi^2(z_\alpha)\\
&= \alpha(1-\alpha)(1+z_\alpha^2)+(1-2\alpha)z_\alpha\phi(z_\alpha)-\phi^2(z_\alpha).
\end{aligned}$$

Piecewise quadratic score, normal case

$$\begin{aligned}
E[S^{(e)}(x,Y)] &= \alpha\int_x^{+\infty}(y-x)^2\phi(y)dy+(1-\alpha)\int_{-\infty}^x(x-y)^2\phi(y)dy\\
&= \alpha[\overline{\Phi}(x)+x\phi(x)+x^2\overline{\Phi}(x)-2x\phi(x)]+\\
&+(1-\alpha)[\Phi(x)-x\phi(x)+x^2\Phi(x)+2x\phi(x)]\\
&= (1-2\alpha)x\phi(x)+(1+x^2)[\alpha\overline{\Phi}(x)+(1-\alpha)\Phi(x)].\\
E[S^{(e)}(x,Y)^2] &= \int_{-\infty}^{+\infty}\left[\alpha(y-x)_+^2+(1-\alpha)(y-x)_-^2\right]^2\phi(y)dy\\
&= \int_x^{+\infty}\alpha^2(y-x)^4\phi(y)dy+\int_{-\infty}^x(1-\alpha)^2(y-x)^4\phi(y)dy.
\end{aligned}$$

Now we have for the first term:

$$\begin{aligned}\int_x^{+\infty}(y-x)^4\phi(y)dy &= x^3\phi(x)+3\left[\overline{\Phi}(x)+x\phi(x)\right]-4x\left[x^2\phi(x)+2\phi(x)\right]+\\ &\quad +6x^2\left[\overline{\Phi}(x)+x\phi(x)\right]-4x^3\phi(x)+x^4\overline{\Phi}(x)\\ &= \overline{\Phi}(x)\left[3+6x^2+x^4\right]-5x\phi(x)-x^3\phi(x).\end{aligned}$$

Similarly for the second term:

$$\begin{aligned}\int_{-\infty}^{x}(y-x)^4\phi(y)dy. &= -x^3\phi(x)+3\left[\Phi(x)-x\phi(x)\right]-4x\left[-x^2\phi(x)-2\phi(x)\right]+\\ &\quad +6x^2\left[\Phi(x)-x\phi(x)\right]-4x^3\left[-\phi(x)\right]+x^4\Phi(x)\\ &= \Phi(x)\left[3+6x^2+x^4\right]+x^3\phi(x)+5x\phi(x).\end{aligned}$$

Summing up, we get

$$E[S^{(e)}(x,Y)^2]=(3+6x^2+x^4)\left(\alpha^2-2\alpha\Phi(x)+\Phi(x)\right)+(1-2\alpha)\left(5x\phi(x)+x^3\phi(x)\right)$$

When $x=e_\alpha$, we get

$$\begin{aligned}Var[S^{(e)}(e_\alpha,Y)] &= (3+6e_\alpha^2+e_\alpha^4)\left(\alpha^2-2\alpha\Phi(e_\alpha)+\Phi(e_\alpha)\right)+\\ &\quad +(1-2\alpha)\left(5e_\alpha\phi(e_\alpha)+e_\alpha^3\phi(e_\alpha)\right)+\\ &\quad -\left((1-2\alpha)e_\alpha\phi(e_\alpha)+(1+e_\alpha^2)\left(\alpha\overline{\Phi}(e_\alpha)+(1-\alpha)\Phi(e_\alpha)\right)\right)^2.\end{aligned}$$

Identification function, Normal case. Recall that

$$I^{(e)}(Y,e)=\alpha(Y-e)_+-(1-\alpha)(Y-e)_-.$$

Since $E[I^{(e)}(Y,e_\alpha)]=0$, we get

$$\begin{aligned}Var[I^{(e)}(Y,e_\alpha)] &= \alpha^2\int_{\varepsilon}^{+\infty}(y-e_\alpha)^2\phi(y)dy+(1-\alpha)^2\int_{-\infty}^{e_\alpha}(e_\alpha-y)^2\phi(y)dy\\ &= \alpha^2\left[\overline{\Phi}(e_\alpha)+e_\alpha\phi(e_\alpha)-2e_\alpha\phi(e_\alpha)+e_\alpha^2(1-\Phi(e_\alpha))\right]+\\ &\quad +(1-\alpha)^2\left[\Phi(e_\alpha)-e_\alpha\phi(e_\alpha)+2e_\alpha\phi(e_\alpha)+e_\alpha^2\Phi(e_\alpha)\right]\\ &= (1+e_\alpha^2)\left[\alpha^2+\Phi(e_\alpha)(1-2\alpha)\right]+e_\alpha\phi(e_\alpha)\left[1-2\alpha\right].\end{aligned}$$

Let now $Y\sim U(0,1)$, $\phi(y)=1_{[0,1]}$, $\Phi(y)=y1_{[0,1]}$ and $z_\alpha=\alpha$.

Piecewise linear score, uniform case.

$$\begin{aligned}
E[S^{(v)}(z_\alpha, Y)] &= \alpha \int_\alpha^1 (y-\alpha)\phi(y)dy + (1-\alpha)\int_0^1 (\alpha - y)\phi(y)dy \\
&= \frac{\alpha(1-\alpha)}{2}. \\
E[S^{(v)}(z_\alpha, Y)^2] &= \alpha^2 \int_\alpha^1 (y-\alpha)^2 dy + (1-\alpha)^2 \int_0^\alpha (\alpha - y)^2 dy \\
Var[S^{(v)}(z_\alpha, Y)] &= E[S^{(v)}(z_\alpha, Y]^2) - [E[S^{(v)}(z_\alpha, Y)]]^2 \\
&= \frac{1}{12}\alpha^2 (\alpha - 1)^2
\end{aligned}$$

Piecewise quadratic score, uniform case. Recall that if $Y \sim U(0,1)$

$$e_\alpha = \frac{\alpha - \sqrt{\alpha - \alpha^2}}{2\alpha - 1}. \tag{B.1}$$

$$\begin{aligned}
E[S^{(e)}(e_\alpha, Y)] &= \int_0^1 \left[\alpha(y-e_\alpha)_+^2 + (1-\alpha)(y-e_\alpha)_-^2\right] dy \\
&= \alpha \int_{e_\alpha}^1 (y-e_\alpha)^2 dy + (1-\alpha)\int_0^{e_\alpha} (e_\alpha - y)^2 dy = \\
&= \frac{1}{3} e_\alpha^3 (1-\alpha) + \alpha\left(e_\alpha^2 - e_\alpha - \frac{1}{3}e_\alpha^3 + \frac{1}{3}\right)
\end{aligned}$$

Substituting (B.1) we get

$$E[S^{(e)}(e_\alpha, Y)] = \frac{1}{3}\alpha(1-\alpha)\frac{1-2\sqrt{\alpha(1-\alpha)}}{(2\alpha-1)^2}.$$

$$\begin{aligned}
E[S^{(e)}(e_\alpha, Y)^2] &= \int_0^1 \left[\alpha(y-e_\alpha)_+^2 + (1-\alpha)(y-e_\alpha)_-^2\right]^2 dy \\
&= \alpha^2 \int_{e_\alpha}^1 (y-e_\alpha)^4 dy + (1-\alpha)^2 \int_0^{e_\alpha} (e_\alpha - y)^4 dy \\
&= \frac{1}{5} e_\alpha^5 (1-\alpha)^2 + \alpha^2\left(2e_\alpha^2 - e_\alpha - 2e_\alpha^3 + e_\alpha^4 - \frac{1}{5}e_\alpha^5 + \frac{1}{5}\right)
\end{aligned}$$

$$\begin{aligned}
E[S^{(e)}(e_\alpha, Y)^2] &= -\frac{11\alpha^4 - 2\alpha^3 - \alpha^2 - 12\alpha^5 + 4\alpha^6 - \left(\alpha - \alpha^2\right)^{\frac{5}{2}}}{5\left(2\alpha - 1\right)^4} + \\
&\quad - \frac{5\alpha^2\sqrt{\alpha - \alpha^2} - 10\alpha^3\sqrt{\alpha - \alpha^2} + 5\alpha^4\sqrt{\alpha - \alpha^2}}{5\left(2\alpha - 1\right)^4} \\
&= -\frac{1}{5}\alpha^2\left(-4\alpha + 4\alpha^2 - 1\right)\frac{\left(\alpha - 1\right)^2}{\left(2\alpha - 1\right)^4} - \frac{4}{5}\frac{\left(\alpha - \alpha^2\right)^{\frac{5}{2}}}{\left(2\alpha - 1\right)^4} \\
&= \frac{\alpha^2\left(\alpha - 1\right)^2}{5\left(2\alpha - 1\right)^4}\left(4\alpha - 4\alpha^2 + 1 - 4\sqrt{\alpha - \alpha^2}\right) \\
&= \frac{\alpha^2\left(1 - \alpha\right)^2}{5\left(2\alpha - 1\right)^4}\left(1 - 2\sqrt{\alpha(1 - \alpha)}\right)^2.
\end{aligned}$$

$$\begin{aligned}
Var[S^e(e_\alpha, Y)] &= \frac{\alpha^2\left(1 - \alpha\right)^2}{5\left(2\alpha - 1\right)^4}\left(1 - 2\sqrt{\alpha(1 - \alpha)}\right)^2 + \\
&\quad - \left(\frac{1}{3}\alpha(1 - \alpha)\frac{1 - 2\sqrt{\alpha(1 - \alpha)}}{\left(2\alpha - 1\right)^2}\right)^2 \\
&= \frac{4}{45}\frac{\alpha^2\left(1 - \alpha\right)^2}{\left(2\alpha - 1\right)^4}\left(1 - 2\sqrt{\alpha(1 - \alpha)}\right)^2
\end{aligned}$$

Note the relationship

$$Std[S^e(e_\alpha, Y)] = \frac{2}{3\sqrt{5}}\frac{\alpha\left(1 - \alpha\right)}{\left(2\alpha - 1\right)^2}\left(1 - 2\sqrt{\alpha(1 - \alpha)}\right) = \frac{2}{\sqrt{5}}E[S^e(e_\alpha, Y)].$$

Identification function, Uniform case.

$$\begin{aligned}
Var[I^e(Y, e_\alpha)] &= \alpha^2\int_{e_\alpha}^{1}(y - e_\alpha)^2 dy + (1 - \alpha)^2\int_0^{e_\alpha}(y - e_\alpha)^2 dy \\
&= \frac{\alpha^2}{3}\left(3e_\alpha^2 - 3e_\alpha - e_\alpha^3 + 1\right) + (1 - \alpha)^2\frac{1}{3}e_\alpha^3 \\
&= \frac{\alpha^2}{3}\left(1 - e_\alpha\right)^3 + \frac{1}{3}(1 - \alpha)^2 e_\alpha^3.
\end{aligned}$$