On the Relation between Firm Characteristics and Volatility Dynamics with an Application to the 2007-2009 Financial Crisis

Christian T. Brownlees*

This Draft: October, 2010
Preliminary and Incomplete

Abstract

Despite the extensive literature on the analysis of firm equity volatility, relatively little is known about the relation between firm characteristics and volatility dynamics. This is partly due to the lack of an appropriate modelling framework in which these research questions can be addressed adequately. This work proposes a Hierarchical Factor GARCH model for multivariate volatility analysis in large panels of assets. The novelty consists of augmenting the dynamic specification with equations that link the volatility dynamics parameters of each firm to observed and unobserved characteristics. The hierarchical approach has features that are useful for both economic and forecasting applications. It permits one to investigate how variation of firm variables explains variation in volatility dynamics. Moreover, it allows for a parsimonious parameterization of a multivariate system that is independent of its dimension, yet capable of retaining flexibility in the individual series dynamics thanks to the random effect structure. The model is estimated via Maximum Likelihood. The proposed methodology is used to analyse the volatility dynamics of top U.S. financial institutions during the 2007-2009 crisis using a financial index as common factor. Dynamics are a function of firm size, leverage, distance to default and liquidity before the beginning of the credit crunch. Results show that leverage is the most influential variable, and firms with high leverage have high factor exposure, high idiosyncratic volatility as well as high sensitivity to temporary idiosyncratic volatility shocks. Factor exposure in the crisis is also high for firms that are large, have a small distance to default and are illiquid. Overall, the model captures a substantial portion of cross sectional variation in volatility dynamics.

Keywords: Hierarchical Nonlinear Dynamic Modelling, GARCH, Volatility, Systemic Risk

JEL: C31, C32, C33

*Department of Finance, Stern School of Business, New York University, e-mail: `ctb@stern.nyu.edu` web page: `http://pages.stern.nyu.edu/~cbrownle/`.
Acknowledgements. I am grateful to Edward Altman, Matteo Barigozzi, Massimiliano Bonacchi, Giorgio Calzolari, Giuseppe Corvasce, Sara De Masi, Robert Engle, Farhang Farazmand, Giampiero Gallo, William Greene, Dennis Kristensen, Claudia Moise, Seda Oz, Stephen Ryan, Donatella Taurasi, Sjoerd Van Bekkum, Marina Vannucci and David Veredas for useful comments and discussions. All mistakes are mine.

1 Introduction

The literature on financial volatility measurement, modelling and forecasting has witnessed a steady growth over the last three decades. Methodological advances and empirical findings in the different areas of the discipline are surveyed in Bollerslev et al. (1994), Ghysels et al. (1996), Poon and Granger (2003, 2005) Andersen et al. (2006), Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009), to name a few. In the study of firm equity volatility few contributions have investigated the relation between firm characteristics and volatility dynamics. For instance, one might be interested in assessing whether levered companies are more prone to abrupt volatility shocks or if illiquid firms have higher volatility persistence. More generally, there is a conspicuous amount of information on firm characteristics coming from, among other sources, financial statements. This information can provide important economic insights on the nature of firm's volatility but is seldom used in this literature.

Understanding the relationship between firm characteristics and volatility dynamics is also of particular interest in the analysis of the 2007–2009 financial crisis. The crisis has harshly demonstrated that the collapse of a number of large financial institutions can generate vast negative externalities to the rest of the economy. These failures are especially harmful because they are not independent, as they are influenced by common economic determinants, such as market, industry or regional factors. The study of these issues goes under the header of *systemic risk* and the topic already has a significant body of contributions (cf. Adrian and Brunnermeier (2009), Acharya, Pedersen, Philippe and Richardson (2010) and Brownlees and Engle (2010)). Current U.S. regulation produced in the aftermath of the crisis to promote financial stability defines a number of criteria to identify risky firms on the basis of their characteristics (cf. Acharya, Brownlees, Engle, Farazmand and Richardson (2010)). This calls for an empirical assessment of which features relate to higher volatility and lower diversification in periods of distress, as well as understanding to which extent they are able to fully explain a firm's exposure to adverse economic conditions.

This work proposes to frame these research questions with a *hierarchical* model. I introduce a Hierarchical Factor GARCH model for the analysis of multivariate volatility dynamics in large panels of assets. The model assumes that the return on an asset is a linear function of an observed systematic factor and an idiosyncratic shock, where both the systematic and idiosyncratic components are assumed to be GARCH. The parameterization employed characterises the dynamics of each asset with four coefficients: a factor loading, long run idiosyncratic variance, persistence of idiosyncratic variance shocks and smoothness of the conditional idiosyncratic variance path. The novelty with respect to standard modelling approach consists of augmenting the specification with equations that link the coefficients to a set of observed characteristics and a random effect. The attribute hierarchical stems from the fact that the specification can be decomposed in two levels: the, so called, Level I, which models returns using a vector of unknown coefficients, and the Level II, which models the Level I coefficients. The methodology has features that are useful in both economic and forecasting applications. It permits one to straightforwardly test for a significant relation between a firm characteristic and a facet of the dynamics of volatility of interest. Moreover, it allows for a parsimonious parameterization of a multivariate system that is independent of its dimension, yet capable of retaining flexibility in the individual series dynamics thanks to the random effect structure. Inference on the model is based on Maximum Likelihood (ML) estimation. The log–likelihood involves the computation of analytically intractable integrals that arise because of the latent random effects. The integrals are low dimensional and standard quadrature rules are applied to approximate them. I resort to Sparse Grid Integration (SGI) methods (Smolyak (1963), Heiss and Winschel (2008), Winschel and Krätzig (2010)) to overcome this hurdle, which are numerically accurate, computationally efficient and easy to implement quadrature rules. Properties of the estimator follow from standard arguments used in the literature, which are analogous to those used to establish the properties of Maximum Simulated Likelihood estimators (cf. Hajivassiliou and Ruud (1994)).

The Hierarchical Factor GARCH is applied to a panel of top U.S. financial firms in the financial crisis period (defined as July 2007 to July 2009) using a financial index as the common factor. The sample of companies is the same analysed in Acharya, Pedersen, Philippe and Richardson (2010). The set of firm characteristics considered are size, leverage, distance to default and liquidity measured using the data available as of the end of June 2007. Of particular interest in this context is the factor loading coefficient that captures the dependence of a firm to the industry factor. *Ceteris paribus*, a high factor loading denotes high exposure to a large fall in the common factor. Results show that the coefficient increases with size and leverage, and decreases with distance to default and liquidity, with leverage being the most relevant variable. The relation between size and the factor loading is somehow surprising as it is typically believed that larger firms are more diversified hence less exposed to common shocks. Results point out that this need not to be the case in a systemic crisis. This finding is also related to the issue of *interconnectedness* (cf. Acharya, Brownlees, Engle, Farazmand and Richardson (2010)), which is usually considered one of the harnessing factors of the crisis. Everything being equal, it is sensible to assume that a large firm ought to be more interconnected than a small one, hence the size variable could be capturing this effect. This finding is also broadly consistent with the evidence of Verde et al. (2005), who document that in high default periods (like a recession) the default rate of large–cap is higher than the one of mid–cap companies. The proportion of cross section variability explained by the model is 35.4%, implying that the idiosyncratic behaviour of individual firms still accounts for a large part of the total cross sectional variability. Long run idiosyncratic variance is decreasing in size and increasing in leverage. Volatility persistence does not have any significant explanatory variable, and in fact the amount of cross sectional variation of the coefficient is substantially small if not absent. This provides evidence that the degree of volatility memory is not firm specific but is common to all firms. The volatility smoothness coefficient increases with leverage. This coefficient is also related to the unconditional kurtosis of a firm's returns, implying that the returns of highly levered firms are heavier. I estimate the same specification from January 2005 to July 2007, a period of expansion and low volatility for the market and the financial industry. The relation between firm characteristics and the volatility dynamics differs, with the exception of leverage, which, roughly speaking, has the same impact. Overall results convey that a substantial amount of cross sectional variation in volatility dynamics is captured by firm characteristics. Leverage is the most influential determinant of volatility dynamics in the panel. On the other hand, illiquidity, small distance to default and, to a minor extent, large size, are harnessing features in the crisis period. The pre–crisis estimation results are also used to investigate which firms have higher systematic exposure before the beginning of the great fall. I consider the top 20% firms by size and rank them on the basis of their factor loading. The first four companies turn out to be, in order, Lehman, Goldman Sachs, Morgan Stanley and Merrill Lynch, that is, all the large investment banks. Pre-crisis results indicate that these firms had worrying level of exposure to systematic shocks and, indeed, these companies turned out to be, in different ways, the most severely harmed by the crisis.

Different strands of literature relate to this work. The intuition of treating coefficients as draws from some underlying population comes from Engle (2009) in the context of the Dynamic Conditional Correlation (DCC) model. This is motivated by the empirical observation that in this context, as well as in several other financial econometrics applications, parameter estimates of the same specification estimated over different assets often cluster. The idea is further developed in Engle et al. (2008) and Pakel et al. (2010) where individual coefficients are treated as incidental parameters and the focus is on estimating their averages. In this work, on the other hand, I explicitly model coefficients and estimate them. Independently, also the research of Bauwens and Rombouts (2007) is based on the same intuition. The use of factor modelling is widespread in the financial econometrics literature in many different shades. The list of recent contributions includes Calzolari et al. (2008), Hautsch (2008), Barigozzi et al. (2010). This paper also relates to the literature on panel data. Recent advances in the field are surveyed, among others, in Arellano and Honoré (2001). Hierarchical modelling is popular in several fields of the social sciences. The terminology is unfortunately highly heterogeneous and these models are also called Multilevel, Random Coefficient, Mixed, and Variance Components, to name a few. The use of random effects in nonlinear models is relatively young because of the numerical challenges in the computation of the likelihood. The list of contributions on the topic includes Davidian and Gallant (1993) and Green (2001). In the econometrics literature, simulation based methods are often used to approximate the intractable multidimensional integrals that arise in the likelihood of latent variable models. The list of proposed techniques includes Maximum Simulated Likelihood, Nonparametric Simulated Maximum Likelihood, Indirect Inference, Efficient Method of Moments, and Monte Carlo EM to name a few (cf. Gourieroux et al. (1993), Hajivassiliou and Ruud (1994), Gallant and Tauchen (1996), Nielsen (2000), Fermanian and Salanié (2004), Kristensen and Shin (2008)). In large dimensional problems simulation is the most effective approach, but for moderate dimensional problems, which are the ones relevant in this work, quadrature performs well. In fact, these methods are the common choice for latent variable modelling software (e.g. the GLAMM package for Stata or the procedure NLMIXED in SAS). The literature on quadrature rules includes, among others, Rabe-Hesketh et al. (2005).

The rest of the paper is structured as follows. Section 2 introduces the model and the inferential approach. Section 3 illustrates the methodology using simulated data. Section 4 presents the application to the 2007-2009 crisis. Concluding remarks follow in Section 5.

2 Hierarchical Nonlinear Dynamic Modelling

2.1 A Hierarchical Factor GARCH Model

The hierarchical modelling approach proposed in this work is illustrated by means of a Factor GARCH model with one observed common factor. Factor modelling is probably the main dimensional reduction technique in multivariate analysis and in the Multivariate GARCH literature this approach has been advocated since its early days, as in Engle et al. (1984), Diebold and Nerlove (1989), Engle et al. (1990) and King et al. (1994). It is is also a natural modelling framework in finance, where the Arbitrage Pricing Theory (APT) suggests that the unexpected return of a risky asset can be expressed as a function of a few common factors and an idiosyncratic component. The model described in this section is referred to as factor double ARCH in Engle (2009), but I will use the name Factor GARCH for simplicity.

Let r_{it} denote the return of the i^{th} firm on period t, with i in $1, ..., N$ and t in $1, ..., T$. Conditionally on the information set at time $t-1$ and the common observed factor r_{Ft}, the return on period t of the i^{th} firm is

$$r_{i\,t} = \beta_i \, r_{F\,t} \; + \; \epsilon_{i\,t}, \tag{1}$$

where the coefficient β_i is the *factor loading* and $\epsilon_{i\,t}$ is an unobserved idiosyncratic shock. For notational convenience I also use the short cut notation r_i to denote the series $\{r_{i\,t}\}_{t=1}^{T}$ for all the time series in the model. The r_F and ϵ_i processes are assumed to be heteroskedastic

$$r_{F\,t} \,|\, \mathcal{I}_{t-1} \sim \mathcal{N}(0, h_{F\,t}) \quad \text{and} \quad \epsilon_{i\,t} \,|\, \mathcal{I}_{t-1}, r_{F\,t} \sim \mathcal{N}(0, h_{i\,t}), \tag{2}$$

and for concreteness sake are assumed to be GARCH(1,1). The equation describing the evolution of the conditional factor variance $h_{F\,t}$ is

$$h_{F\,t} = \omega_F + \alpha_F \, r_{F\,t-1}^2 + \beta_F \, h_{F\,t-1}, \quad \omega_F > 0, \, \alpha_F > 0, \, \beta_F \geq 0, \tag{3}$$

and the conditional idiosyncratic variance $h_{i\,t}$ is

$$h_{i\,t} = \omega'_i + \alpha'_i \, \epsilon_{i\,t-1}^2 + \beta'_i \, h_{i\,t-1}, \quad \omega'_i > 0, \, \alpha'_i > 0, \, \beta'_i > 0,$$

where it is explicitly assumed that $\beta'_i > 0$ and that the idiosyncratic variance is stationary, that is $\alpha'_i + \beta'_i < 1$. This work resorts to an alternative parametrization of the conditional idiosyncratic variance equation that turns out to be more convenient in a hierarchical setting

$$h_{i\,t} = (1 - \pi_i) h_i + \pi_i \left[\lambda_i \epsilon_{i\,t-1}^2 + (1 - \lambda_i) h_{i\,t-1} \right], \quad h_i > 0, \, 0 < \pi_i < 1, \, 0 < \lambda_i < 1, \tag{4}$$

where one immediately sees that the mapping with the classical formulation is $h_i = \omega'_i/(1 - \alpha'_i - \beta'_i)$, $\pi_i = \alpha'_i + \beta'_i$ and $\lambda_i = \alpha'_i/(\alpha'_i + \beta'_i)$. Thus, h_i is the unconditional idiosyncratic variance, which I call *idiosyncratic variance* for short. The π_i coefficient is called *persistence* and it controls the amount of dependence of the idiosyncratic variance evolution. It is well known that the autocorrelation of the squared GARCH(1,1) process is

$$\mathsf{Corr}(\epsilon_{i\,t}^2, \epsilon_{i\,t-l}^2) = \rho_1 \pi_i^{l-1}$$

where ρ_1 is the lag one autocorrelation. It is evident from the formula that higher levels of π_i lead to a higher and a more slowly decaying pattern. Finally, λ_i determines the *smoothness* of the idiosyncratic variance path, the higher the value of the coefficient the rougher the volatility path. Also, π_i and λ_i relate to unconditional kurtosis of the idiosyncratic process

$$\mathsf{kurtosis}_i = 6 \, \frac{\pi_i^2 \lambda_i^2}{1 - \pi_i^2 (1 - 2\lambda_i^2)} \tag{5}$$

where it can be readily checked that the formula is an increasing function in both coefficients.[1] The alternative parameterization of Equation (4) has two advantages: the set of constraints on the idiosyncratic variance equation are independent of each other (which is an important feature to implement a hierarchical formulation) and coefficients have an easier interpretation. Using hierarchical terminology, Equations (1) to (4) will be referred to as the Level I model.

The properties of Factor GARCH models are well known and have been illustrated, among others, in Engle et al. (1990). The conditional variance of an asset is given by

$$\mathsf{Var}_{t-1}(r_{i\,t}) = \beta_i^2 h_{F\,t} + h_{i\,t},$$

[1] The kurtosis formula in Equation (5) applies to a GARCH(1,1) with Gaussian errors. Under more general distributional assumptions, Bai et al. (2003) show that kurtosis is still an increasing function of this quantity.

where the two terms of the expression are referred to as the systematic and idiosyncratic variance components; and conditional correlation between two assets is

$$\mathsf{Corr}_{t-1}(r_{i\,t}, r_{j\,t}) = \frac{\beta_i \beta_j h_{F\,t}}{\sqrt{(\beta_i^2 h_{F\,t} + h_{i\,t})(\beta_j^2 h_{F\,t} + h_{j\,t})}}.$$

Long run versions of these expressions are obtained by replacing the conditional variances with their unconditional analogs.

The departure from standard modelling consists of augmenting the specification with equations for the factor loading, idiosyncratic variance, persistence and smoothness coefficients $(\beta_i, h_i, \pi_i, \lambda_i)'$ that characterise the dynamics of each series. The model links the coefficients to a set of observed and unobserved firm characteristics. Let x_i denote a vector of p dimensional variables like industry group, size, leverage and so forth. The coefficients of the Level I model are given by the following set of equations

$$\begin{aligned}
\beta_i &= \delta_{\beta\,0} + \sum_{k=1}^{p} \delta_{\beta\,k}\, x_{i\,k} + u_{\beta\,i} & u_{\beta\,i} &\sim \mathcal{N}(0, \tau_\beta^2), & (6)\\
\log(h_i) &= \delta_{h\,0} + \sum_{k=1}^{p} \delta_{h\,k}\, x_{i\,k} + u_{h\,i} & u_{h\,i} &\sim \mathcal{N}(0, \tau_h^2), & (7)\\
\Phi^{-1}(\pi_i) &= \delta_{\pi\,0} + \sum_{k=1}^{p} \delta_{\pi\,k}\, x_{i\,k} + u_{\pi\,i} & u_{\pi\,i} &\sim \mathcal{N}(0, \tau_\pi^2), & (8)\\
\Phi^{-1}(\lambda_i) &= \delta_{\lambda\,0} + \sum_{k=1}^{p} \delta_{\lambda\,k}\, x_{i\,k} + u_{\lambda\,i} & u_{\lambda\,i} &\sim \mathcal{N}(0, \tau_\lambda^2), & (9)
\end{aligned}$$

where $u_i = (u_\beta, u_h, u_\pi, u_\lambda)'$ are independent Gaussian random effects and $\log(\cdot)$ and Φ^{-1} are link functions appropriately mapping the domain of the coefficients onto $\mathbb{R}$ ($\Phi^{-1}(p)$ is the inverse of the standard Gaussian cumulative density function, the probit link). The use of link functions to appropriately map the domain of a coefficient is a commonly used device in the financial econometrics literature (cf. Patton (2006), Distaso et al. (2009)) and is inspired by Generalized Liner Models. Let $z_i = (z_\beta, z_h, z_\pi, z_\lambda)' = (u_\beta/\tau_\beta, u_h/\tau_h, u_\pi/\tau_\pi, u_\lambda/\tau_\lambda)'$ be defined as the *specificity* of firm i, in the sense that z_i captures the deviation of the behaviour of company i from what the observed characteristics x_i would otherwise imply.[2] Equations (6) to (9) describe the so called Level II model using, again, hierarchical modelling terminology. They add the Hierarchical attribute to Factor GARCH in the sense that the coefficients that drive Level I model are the outcomes of the Level II specification.

The dynamics of the full multivariate system are governed by the *factor* parameter $\psi = (\omega_F, \alpha_F, \beta_F)'$ and the *hyper* parameter $\theta = (\delta_\beta', \delta_h', \delta_\pi', \delta_\lambda', \tau_\beta^2, \tau_h^2, \tau_\lambda^2, \tau_\pi^2)'$.

2.2 Discussion

The logic of the model is that idiosyncratic and systematic news drive the volatility of a firm, but that the way information is processed depends on its characteristics. Characteristics are time invariant. This assumption could obviously be relaxed at the expense of making the specification

[2]A more common term than specificity would probably be idiosyncrasy. However, I avoid using this terminology in order not to create confusion with the idiosyncratic process ϵ_i.

more complex. However, it is reasonable to believe that while volatility is inherently a high frequency phenomenon many firm features such as industry group classification, size, leverage, firm governance, etc. are slowly varying and often measured at low frequencies (quarterly or yearly). They can be considered constant to the extent that the time series span of the analysis has a reasonable length.

The proposed modelling approach differs substantially from adding lagged firm characteristics in the conditional variance equation, which is somehow more common (cf. Poon and Granger (2005), Brownlees and Gallo (2008)). One might be inclined to do so especially given that mixed data sampling (MIDAS) issues have recently been effectively tackled in the literature (see Ghysels et al. (2006)). If one is interested in determining the relation between volatility and asset characteristic, the challenge in this approach is that this source of information is not necessarily independent of the news process. For instance, if a firm decides to change its governance, it is likely that information concerning such a decision will gradually leek in the news process, corroding the relevance of the proxy capturing the characteristic.

On the other hand, the approach of this paper has connections with the credit scoring model literature, starting from Altman's seminal contribution (Altman (1968)). The aim of this literature is to identify characteristics that explain the default probability of a firm. One of the motivations of Altman's original contributions was to "*bridge the gap rather than sever the link between traditional ratio "analysis" and the more rigorous statistical techniques*" in order to better understand default risk. This paper follows this very same philosophy, with the only difference that the focus is on volatility dynamics.

The Hierarchical Factor GARCH also easily lends itself to a Bayesian interpretation. The Level II specification can be interpreted as a prior for the Level I model coefficients and the δ and τ^2 parameters are what a Bayesian calls hyper parameters. Hierarchical modelling is in fact at the core of the Bayesian way of thinking.

The model also has an interpretation in terms of shrinkage and pooled estimation. The estimator of the individual firm coefficients can be interpreted as a compromise between the information available for that company and the one available from all other companies in the panel. The random effect shrinks the coefficient of each asset to the expected value implied by its characteristics. The degree of shrinkage is determined by the variance of the random effect that is estimated from the data. Borrowing estimation strength from the ensemble is appealing for short time series and for coefficients with relatively large asymptotic variability. As one would hope, this is only a finite sample correction and as the time series dimension grows standard arguments deliver consistency of the coefficients. The advantage of the random effect formulation over shrinkage is to avoid many of the inferential problems associated with bandwidth selection optimality and post bandwidth selection inference (cf. Leeb and Pötscher (2005) and Leeb and Pötscher (2006)). In fact, in semiparametric modelling (cf. Ruppert et al. (2003)) a hierarchical estimation approach is often used for estimation and shrinkage selection.

The use of random coefficients also has an empirical motivation. In the analysis of panels of financial time series (like in Shephard and Sheppard (2010), Barigozzi et al. (2010)) it is often found that the estimation results of the same specification over the different series exhibit *coefficient clustering*. The typical example would be the one of the simple GARCH(1,1) on equity returns, which delivers "ARCH" and "GARCH" parameter estimates in the neighbourhood of, respectively, 0.05 and 0.95. The assumption that in an homogeneous panel of series coefficients are drawn by some underlying population with some mean and variance to be estimated from the data seems to be an empirically reasonable one.

Last, the hierarchical formulation can also be seen as a statistical hack to counter the curse of dimensionality, in that the number of parameters in the model is independent of the number of assets. The fixed effect version of the specification would have a number of parameters proportional to the number of assets, which is clearly numerically unfeasible for moderately large panels. At the same time, the random effect structure allows one to keep flexibility in the individual series dynamics. The price to pay to kill vicious parameter proliferation that poisons multivariate volatility modelling is a likelihood function involving a multidimensional integral that is typically not tractable analytically. However, as is detailed in Section 2.3, the integral is low dimensional, and straightforward numerical integration techniques can be applied.

2.3 Inference

There are several fixed and random quantities of interest that need to be estimated from the data. The fixed factor parameter $\psi = (\omega_F, \alpha_F, \beta_F)'$, the fixed hyper parameter $\theta = (\delta'_\beta, \delta'_h, \delta'_\pi, \delta'_\lambda, \tau^2_\beta, \tau^2_h, \tau^2_\pi, \tau^2_\lambda)'$, the random effects realizations u_i and the random coefficients realizations $(\beta_i, h_i, \pi_i, \lambda_i)'$.

The estimation of the fixed parameters ψ and θ is carried out by maximum likelihood (ML). The random effects u_i are predicted[3] using the customary Empirical Bayes (EB) predictor conditional on the ML estimate of θ. Finally, random coefficients $(\beta_i, h_i, \pi_i, \lambda_i)'$ are obtained by plugging into the respective equations the ML estimate of θ and EB predictions of u_i.

The estimators and predictors introduced have diverse sampling properties. The factor parameter is consistent for large T. On the other hand, the hyper parameter is consistent as N becomes large for any fixed T. Because of the random effect assumption, inference on the random effects and random coefficients realizations requires a consistent estimator of the hyper parameter, yet it can be carried out for any fixed T. However, it follows from standard arguments that with large T predictions consistently estimate the latent realizations.

The ML estimator of θ and EB predictors of u_i involve the computation of integrals that are not analytically tractable and quadrature rules based on Sparse Grid Integration (SGI) detailed in Section 2.4 are used to overcome this hurdle.

2.3.1 Fixed Parameters

The likelihood of the returns panel $\mathbf{r} = (r_1, ..., r_N, r_F)$ can be factorized in the joint likelihood of the N assets r_1 to r_N conditional on the common factor r_F and the likelihood of the factor r_F. Estimation of the two sets of parameters θ and ψ entering respectively the first and second likelihood components can be carried out independently. I shall not digress on the estimation of the factor parameter ψ, which is standard GARCH ML, and in this section I focus on hyper parameter θ only.

[3]It is standard practice to refer to the estimation of random effect realizations as prediction. See Robinson (1991) for an interesting discussion on the origin of the terminology.

The marginal log–likelihood of the N assets conditionally on the factor is

$$
\begin{aligned}
Q_N(\theta) = \log L_N(\theta; r_F, \mathbf{x}), \quad &= \log \int \prod_{i=1}^{N} f(r_i|x_i, u_i, r_F; \theta)\ p(u_i; \theta), \\
&= \sum_{i=1}^{N} \log \int f(r_i|x_i, u_i, r_F; \theta)\ p(u_i; \theta), \\
&= \sum_{i=1}^{N} \log \int f(r_i|x_i, z_i, r_F; \theta)\ \phi(z_i), \\
&= \sum_{i=1}^{N} q_i(\theta), \qquad (10)
\end{aligned}
$$

where $\phi(\cdot)$ is the pdf of a quadrivariate standard Gaussian random variable. The marginal log–likelihood contains a multi–dimensional integral that arises in integrating out the random effects u_i. However, the structure of the model allows one to factor the $4 \cdot N$ dimensional integral in the sum of N integrals of dimension 4, which albeit not being analytically tractable are feasible to compute numerically using standard quadrature techniques.

The intractable objective function in Equation (10) is replaced by

$$
\widetilde{Q}_N^{\,l}(\theta) = \sum_{i=1}^{N} \tilde{q}_i^{\,l}(\theta),
$$

where $\tilde{q}_i^{\,l}(\theta)$ is used to denote the approximation of $q_i(\theta)$. The Maximum (Approximated) Likelihood estimator is defined as

$$
\hat{\theta} = \arg\max_{\theta} \widetilde{Q}_N^{\,l}(\theta).
$$

Results on the consistency and asymptotic distribution of the estimator can be obtained from standard arguments similar to those used to establish the properties of Maximum Simulated Likelihood estimators (cf. Hajivassiliou and Ruud (1994)).

Some more notation on numerical integration has to be set.[4] To carry out inference it is necessary to evaluate the integral of functions $g : \mathbb{R}^4 \to \mathbb{R}$ with Gaussian weighting, that is

$$
G = \int g(z)\ \phi(z)\ \mathrm{d}\,z, \qquad (11)
$$

which are approximated using a quadrature rule, that is

$$
\widetilde{G}^{\,l} = \sum_{z \in \mathcal{Z}_l} g(z)\ w_l(z), \qquad (12)
$$

where $\mathcal{Z}_l$ is a set of abscissas in $\mathbb{R}^4$ called grid, $w_l(\cdot)$ is a function that associates every abscissa z in $\mathcal{Z}_l$ to a weight and l denotes the level of accuracy of the rule. The absolute error bound is denoted by $\mathcal{E}_l(g)$ or $\mathcal{E}_l(G)$ depending on convenience. Bounds on the error $\mathcal{E}_l$ depend on the quadrature rule used and smoothness conditions on the integrand g. Although the emphasis

[4]Admittedly, I avoid using detailed notation and I focus only on the subset needed in this context.

in this work is on quadrature, the approximation strategy of Equation (12) also encompasses simple Monte Carlo integration by letting the grid be a set of realizations from the standard Gaussian and setting the weight function to unity.

The following algorithm is used to approximate $q_i(\theta)$

$$\begin{aligned}\tilde{q}_i^l(\theta) &= \log \sum_{z \in \mathcal{Z}_l} \exp\left[\log f(r_i|x_i, z, r_F; \theta) - \log f(r_i|x_i, z^*, r_F; \theta)\right]\ w_l(z) \\ &\quad + \log f(r_i|x_i, z^*, r_F; \theta), \end{aligned} \tag{13}$$

where

$$\log f(r_i|x_i, z, r_F; \theta) = \sum_{t=1}^{T} \log f_t(r_{i\,t}|x_i, z, r_{F\,t}; \theta),$$

and with

$$z^* = \arg\max_{z \in \mathcal{Z}_l} \log f(r_i|x_i, z, r_F; \theta).$$

The procedure computes the log–likelihood of the i^{th} series using log–densities only. It is straightforward to verify that Equation (13) is *algebraically* equivalent to

$$\tilde{q}_i^l(\theta) = \log \sum_{z \in \mathcal{Z}_l} f(r_i|x_i, z, r_F; \theta)\ w_l(z)$$

however, from a *numerical* standpoint, the direct computation of the density on a computer using finite precision would lead to underflow for moderately large T while the proposed method does not. The numerical evaluation of the integral is relatively simple. Requires one to code the log–likelihood of the model conditionally on observing the random effects and then to compute a rescaled weighted average of these functions for different values of the random effects. Also note that z^* is the maximum over a set of finite dimension, so no numerical optimization is required to find it. The number of floating point operations required to compute the objective function increases (linearly) in the number of assets N, (linearly) in length of the time series T and desired level of accuracy l. The optimization of the objective function in large systems with high accuracy may require one to compute lengthy function evaluations but to the extent that θ is small dimensional the optimization problem is straightforward.

Consistency can be established starting from the assumption that the ML estimator based on the analytically intractable log–likelihood $Q_N(\cdot)$ is consistent (for instance, Theorem 2.1 of Newey and McFadden (1994)). Let θ_0 be the maximizer of $Q_0(\theta)$, the probability limit of $Q_N(\theta)$ for arbitrarily large N. It can be shown that $\hat{\theta}$ is consistent for θ_0 provided that the level of accuracy l is sufficiently high to make the approximation error negligible. An advantage of the algorithm described in Equation (13) is that the error bound is (Proposition 1 in the Appendix)

$$\mathcal{E}_l(q_i) = \mathcal{E}_l(f_i) \left[\frac{f(r_i|x_i, z^*, r_F; \theta)}{\int f(r_i|x_i, z, r_F; \theta)\ \phi(z)}\right],$$

instead of the one that one would get from the direct computation

$$\mathcal{E}_l(f_i) \left[\frac{1}{\int f(r_i|x_i, z, r_F; \theta)\ \phi(z)}\right],$$

which would suggest that for small values of the marginal density of f large levels of accuracy ought to be used.

This large bound is a consequence of not rescaling the integrand appropriately.

The asymptotic distribution of the estimator can be obtained via the customary mean value theorem expansion around θ_0

$$\sqrt{N}(\hat{\theta} - \theta_0) \quad = \quad \left[\frac{1}{N}\nabla^2_\theta \widetilde{Q}^{\,l}_N(\check{\theta})\right]^{-1} \cdot \frac{1}{\sqrt{N}}\nabla_\theta \widetilde{Q}^{\,l}_N(\theta_0), \tag{14}$$

where the elements of $\check{\theta}$ lie on the segment between θ_0 and $\hat{\theta}$. The task is then to show that appropriate LLN and CLT can be applied to the right hand side of Equation (14) and that the approximation error is sufficiently small so that the familiar result can be obtained, that is

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\theta_0)).$$

where $I(\theta_0)$ is $\mathsf{E}(\,\nabla^2_\theta Q_0(\theta_0)\,)$.

The derivative of the intractable log–likelihood is

$$\nabla_\theta q_i = \int \nabla_\theta \log f(r_i|x_i, z, r_F; \theta)\ \frac{f(r_i|x_i, z, r_F; \theta)}{f(r_i|x_i, r_F; \theta)}\ \phi(z),$$

and the derivative of the approximated log–likelihood is

$$\nabla_\theta \tilde{q}^{\,l}_i = \sum_{z \in \mathcal{Z}_l} \nabla_\theta \log f(r_i|x_i, z, r_F; \theta)\ \exp\left[\log f(r_i|x_i, z, r_F; \theta) - \tilde{q}^{\,l}_i(\theta)\right]\ w_l(z). \tag{15}$$

The assessment of the approximation error in Equation (15) is slightly complicated by the fact that the formula involves using an approximation of the log–likelihood which enters the equation nonlinearly. It turns out that (Proposition 2 in the Appendix) that

$$\nabla_\theta \tilde{q}^{\,l}_i = \nabla_\theta q_i + A_i + B_i$$

where A_i and B_i are bounded by

$$A_i \leq \exp(q_i - \tilde{q}^{\,l}_i)\ \mathcal{E}_l(\nabla q_i) \quad \text{and} \quad B_i \leq \nabla_\theta q_i\ \mathcal{E}_l(q_i).$$

The score of the objective function can then be expressed as

$$\frac{1}{\sqrt{N}}\nabla_\theta \widetilde{Q}^{\,l}_N(\theta_0) = \frac{1}{\sqrt{N}}\nabla_\theta Q_N(\theta_0) + \frac{1}{\sqrt{N}}\sum_{i=1}^N A_i + \frac{1}{\sqrt{N}}\sum_{i=1}^N B_i.$$

As long as the level of accuracy l in the least favourable approximations of A_i and B_i is such that $\sqrt{N}\max_i A_i$ and $\sqrt{N}\max_i B_i$ are negligible, the asymptotic distribution of the approximated and exact score is going to be the same. Consistency of $\check{\theta}$ for θ_0 and analog arguments can be used to established that

$$\frac{1}{N}\nabla^2_\theta \widetilde{Q}^{\,l}_N(\check{\theta}) \xrightarrow{p} \mathsf{E}\left(\,\nabla^2_\theta Q_0(\theta_0)\,\right) = I(\theta_0),$$

which leads to the familiar result.

2.3.2 Random Effects and Random Coefficients

Conditionally on the observed factor r_F, asset characteristic x_i and parameter θ, the random effect and return series pairs (r_i, u_i) of each asset are independent of each other. It follows from basic probability theory that inference about u_i is based on the conditional distribution

$$p(u_i|r_i, x_i, r_F; \theta) \propto \prod_{t=1}^{T} f(r_i|x_i, u_i, r_F; \theta)\, p(u_i; \theta),$$

where $\propto$ denotes that the distribution is equal to the quantity on the right hand side up to a multiplicative constant. An obvious predictor for the random effect u_i is its conditional mean of the random effect conditional on available information

$$\mu_i = \mathsf{E}(u_i|r_i, x_i, r_F; \theta), \tag{16}$$

and its variance can be measured as

$$\Sigma_i = \mathsf{Var}(u_i - \mu_i|r_i, x_i, r_F; \theta). \tag{17}$$

Alternatively, authors like Davidian and Gallant (1993) propose using the maximum of the conditional distribution, the mode

$$M_i = \arg\max_{u_i} \prod_{t=1}^{T} f(r_i|x_i, u_i, r_F; \theta)\, p(u_i; \theta),$$

straightforward manipulations allow us to see that this estimator can be interpreted as a penalized maximum likelihood estimator with ridge type penalty

$$\begin{aligned} M_i &= \arg\max_{u_i} \sum_{t=1}^{T} \log f(r_i|x_i, u_i, r_F; \theta) + \log p(u_i; \theta) \\ &= \arg\max_{u_i} \sum_{t=1}^{T} \log f(r_i|x_i, u_i, r_F; \delta) - \frac{1}{2} \sum \left| \frac{u_{ij}}{\tau_j} \right|^2 \end{aligned}$$

where the reciprocal of the random effects variance τ^2 is the equivalent of the so called "shrinkage" or "ridge" parameter. Note that for appropriately regular multivariate unimodal distributions the distance between the mean and the mode is bounded, where the bound is inversely proportional to the variance.[5] Furthermore, the law of total variance implies that the expectation of the variance in Equation (17) is a nonincreasing function in T, and *uniformly* so. Hence, the deviation between the two predictors is expected to get smaller in longer time series. Computationally speaking μ_i requires solving a numerical integration problem while M_i involves a numerical optimization. As it is more commonly done in the literature, I use μ_i to predict the random effect.

[5] It follows from basic probability that if X is a univariate unimodal distribution with finite second moment then

$$|\mathsf{mode}(X) - \mathsf{E}(X)| \leq \sqrt{3\,\mathsf{Var}(X)}.$$

Analogous inequalities hold for multivariate random variables (Basu and DasGupta (1996)) depending on the definiton of unimodality adopted in a multivariate setting.

The random coefficients $(\beta_i, h_i, \pi_i, \lambda_i)'$ predictions are obtained by substituting the latent random effect realizations with their predictions. In the case of the factor loading β_i is straightforward to see that one gets

$$\mathsf{E}(\beta_i|r_i, x_i, r_F; \theta) \;=\; \delta_{\beta\,0} + \sum_{k=1}^{p} \delta_{\beta\,k}\, x_{i\,k} + \mathsf{E}(u_{h\,i}|r_i, x_i, r_F; \theta). \tag{18}$$

This approach only delivers unbiased predictions for the coefficients with linear link (that is the factor loading β_i). In the case of log or probit link however, a Taylor expansion of the moments of a function of a random variable suggests that the bias ought to be small when the variance of the random effect predictor is sufficiently small, and, in fact

$$\begin{aligned}\mathsf{E}(h_i|r_i, x_i, r_F; \theta) \;\approx\; & \exp\left\{\delta_{h\,0} + \sum_{k=1}^{p} \delta_{h\,k}\, x_{i\,k} + \mathsf{E}(u_{h\,i}|r_i, x_i, r_F; \theta)\right\} \\ & + \frac{1}{2}\exp\left\{\delta_{h\,0} + \sum_{k=1}^{p} \delta_{h\,k}\, x_{i\,k} + \mathsf{E}(u_{h\,i}|r_i, x_i, r_F; \theta)\right\} \mathsf{Var}(u_{h\,i}|r_i, x_i, r_F; \theta).\end{aligned}$$

In practice random effects and random coefficients predictions are constructed by plugging in the ML estimator $\hat{\theta}$ into Equations (16) and (18) respectively, that is

$$\hat{u}_i = \hat{\mu}_i = \mathsf{E}(h_i|r_i, x_i, r_F; \hat{\theta})$$

and, in the case of the factor loading β_i, one gets

$$\hat{\beta}_i = \hat{\delta}_{\beta\,0} + \sum_{k=1}^{p} \hat{\delta}_{\beta\,k}\, x_{i\,k} + \hat{u}_{\beta\,i}.$$

Also, the naïve estimate of the standard error of $\hat{u}_i$ is defined as

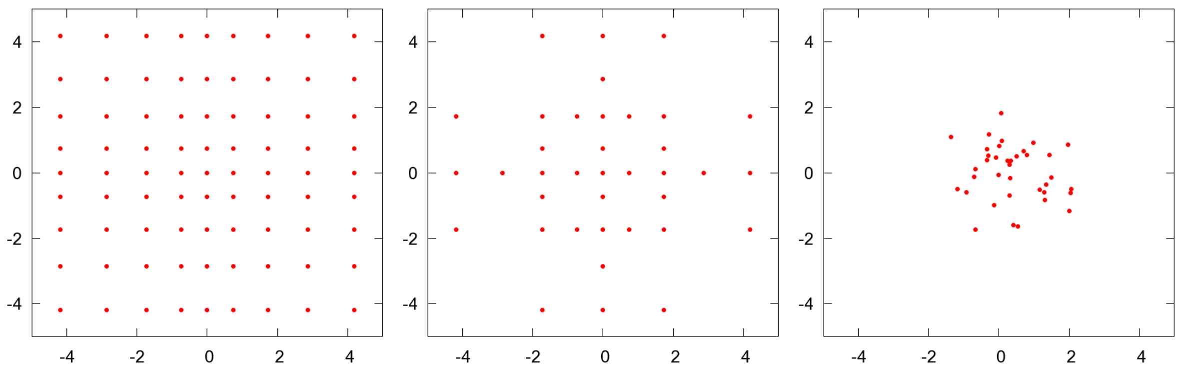
$$\widehat{se}_{\beta\,i} = \sqrt{\widehat{\Sigma}_{i\,\beta}}$$

where $\widehat{\Sigma}_{i\,\beta}$ denotes the variance on the diagonal of the covariance matrix of $\widehat{\Sigma}_i$ corresponding to the factor loading β. The estimator of the standard error predictions is called naïve as it does not take into account estimation variability. This is a topic of active research in the hierarchical modelling literature.

The predictors of the random effects do not require the time series dimension T to become arbitrarily large, however one might be interested in the large sample behaviour of these quantities. As the time-series length T increases, the two estimators are expected to be consistent for the latent realizations of u_i. The M_i predictor is consistent for the realization of u_i under standard conditions for maximum likelihood. Note that as the sample size increases the contribution to the likelihood of the random effect part becomes negligible (see also asymptotics for shrinkage type estimators, like Knight and Fu (2000)). Analogously, a Bernstein-Von Mises type argument can be used to show the consistency of μ_i.

2.4 Multidimensional Numerical Integration with Sparse Grids

Sparse Grid Integration (SGI) rules are used to approximate the intractable integrals in the likelihood of the model. SGI methods can be tracked back to Smolyak (1963) and were reproposed



(a) Product Rule Grid

(b) SGI Grid

(c) Monte Carlo Grid

Figure 1: Grid comparison. The figure plots the bivariate integration grids of the (a) Product Rule, (b) Sparse rule and (c) simple Monte Carlo integration.

in the 90's in the numerical analysis literature (cf. Bungartz and Griebel (2004)). Recently, these techniques have been introduced in econometrics by, among other, Heiss and Winschel (2008), which provide an excellent discussion on the topic. This section briefly sketches the main ideas along the lines of their contribution. A classic exposition on numerical integration is Davis and Rabinowitz (1984).

The objective is to approximate the type of integrals of Equation (11) using quadrature rules of the form given in Equation (12) that are able to deliver high approximation accuracy with the least number of function evaluations possible. In the one dimensional case there are many rules that precisely approximate integrals using relatively few function evaluations, provided the integrand is sufficiently regular. Integration problems get more challenging in multiple dimensions. One of the commonly used methods is the so called Product Rule (Tauchen and Hussey (1991)), which is a combination of one dimensional quadrature rules. This approach however is affected by the curse of dimensionality. The number of function evaluations grows exponentially with the dimension of the domain of integration, making the method feasible for small problems only. This limitation is typically the motivation to use simulation based approaches. SGI is an appealing solution for moderately large problems. SGI rules are still constructed combining one dimension rules but in a more parsimonious way, hence the name sparse. Asymptotically, the number of function evaluations grows polynomially rather than exponentially (cf. Theorem 2 in Heiss and Winschel (2008)) while still guaranteeing adequate approximation properties.

Figure 1 plots bivariate grids of the Product Rule and SGI, obtained by combining the same univariate quadrature rule, together with Monte Carlo Integration. One immediately sees how the Product Rule extends in multiple dimension in a more expensive way than SGI. The number of function evaluations of the former is 81 versus the 37 of the latter. Figuratively speaking, the figure shows how the SGI grid attempts to cover the domain of integration in a more clever way with respect to the way of the Product Rule. For comparison purposes, I also report the Monte Carlo integration grid that one could get by simulating 37 points from a standard Gaussian distribution. Among other things, simple simulation requires a large number of replications in order to cover appropriately the domain of integration and there is no guarantee that grid points are sufficiently far from each other, making some function evaluations redundant.

Several improvements are available to increase the accuracy and efficiency of quadrature rules, like a change of variables or adaptive methods (cf. Rabe-Hesketh et al. (2005)). Here I prefer to use to SGI in its basic form to stress how these rules work adequately in their simplest form.

In the computation of the likelihood and its derivatives in this work I use SGI based on univariate nested quadrature rules for integrals with Gaussian weighting. Grids and weights are the ones made publicly available as supplement material of Heiss and Winschel (2008).[6]

2.5 Allowing for Richer Dynamics

The specification introduced in Section 2.1 is sufficiently flexible for the purposes of this work, however, there is an array of straightforward extensions that allow one to handle specific modeling needs that often arise in volatility analysis.

In the standard GARCH(1,1) model, estimated persistence π_i is typically close to unity and this can be problematic in this framework in that the parameters of the π_i equation can be poorly identified. In fact, in this work I choose a factor model because, empirically, the issue of persistence close to unity is alleviated by the fact that the factor soaks parts of the variance dynamics (and this also turns out to be the case in the application). More generally, there are different ways to approach this issue that I am going to briefly sketch. In case all assets exhibit strong persistence, then π_i can be treated as a fixed coefficient common across assets which gets either estimated from the data or fixed to unity. If assets exhibit strong persistence, but there is a significant degree of variability, one could resort to a link function which converges to unit as slowly as possible, like log–log link function. This is also a device used in binary choice models to attenuate the parallel issue of over/under dispersion. Another solution consists of resorting to a specification that disentangles short and long run persistence such as the Component GARCH (cf. Engle and Lee (1999), Andersen et al. (2006)). In this model, the conditional idiosyncratic variance evolution is described by (using this work's parameterization)

$$h_{i\,t} = \zeta_{i\,t}(1 - \pi_i) + \pi_i \left[\lambda_i \epsilon^2_{i\,t-1} + (1 - \lambda_i) h_{i\,t-1}\right],$$

where $\zeta_{i\,t}$, the long term variance, is

$$\zeta_{i\,t} = \omega_i + \rho\zeta_{i\,t-1} + \phi(\epsilon^2_{i\,t-1} - h_{i\,t-1}),$$

with ρ and ϕ being fixed coefficients capturing the persistence of the long term variance assumed to be common across assets. Empirically, this model typically delivers estimates of ρ close to unity while the short term persistence captured by π_i is bounded away from one.

The asymmetric volatility effect, that is the tendency of volatility to react differently to positive or negative news, is often a relevant feature of volatility modelling and forecasting. In order to allow for asymmetric effects *à la* TARCH models (Rabemananjara and Zakoian (1993), Glosten et al. (1993)) it is convenient to resort to the parameterization of the APARCH model (Ding et al. (1993)), that is

$$h_{i\,t} = (1 - \pi_i)h_i + \pi_i \left[\lambda_i(|\epsilon_{i\,t-1}| - \gamma_i \epsilon_{i\,t-1})^2 + (1 - \lambda_i) h_{i\,t-1}\right],$$

with $\gamma_i \in (-1, 1)$. The equation for γ_i is

$$\mathsf{link}(\gamma_i) = \delta_{\gamma\,0} + \sum_{k=1}^{p} \delta_{\gamma\,k}\, x_{i\,k} + u_{\gamma\,i}, \qquad u_{\gamma\,i} \sim N(0, \tau^2_\gamma)$$

[6]Resources are posted on the website `http://www.sparse-grids.de/`.

where $\mathsf{link}(\cdot)$ is an appropriate link function from $(-1, 1)$ onto $\mathbb{R}$, for instance $\Phi^{-1}((x+1)/2)$.

The introduction of ultra–high frequency based measures of daily volatility, also known as *realized volatility*, has been one of the prominent developments of the financial econometrics literature (cf. Andersen et al. (2003), Aït-Sahalia et al. (2005), Bandi and Russell (2006), Barndorff-Nielsen et al. (2008)). After a first wave of papers focused on modelling realized volatility itself (*inter alia*, Deo et al. (2006), Engle and Gallo (2006), Ghysels et al. (2006), Corsi (2010)) a more recent strand of contributions has focused on using realized volatility to model the volatility of returns (cf. Giot and Laurent (2003) and Brownlees and Gallo (2010)). Several authors have now proposed alternative specifications for the variance of returns using realized volatility on the right hand side (cf. Engle and Gallo (2006), Shephard and Sheppard (2010), Hansen et al. (2010), see also Chen et al. (2010)). Without going into more sophisticated details and assuming the absence of a factor struture, realized volatility based modelling of the returns variance can be accomodated using

$$h_{i\,t} = (1 - \pi_i) h_i + \pi_i \left[\lambda_i rv_{i\,t-1} + (1 - \lambda_i) h_{i\,t-1} \right]$$

where $rv_{i\,t-1}$ is an appropriate estimator of the latent variance of asset i on period $t - 1$.

3 Monte Carlo Illustration

This section provides an illustration of the model and estimation methodology using simulated data. The exercise mimics some of the features of the empirical application in this work in order to give some insights on the performance of the estimator in that context.

I simulate a panel of 100 assets with a single characteristic x assumed to be a Gaussian random variable with mean zero and variance one. The Level II specification is

$$\begin{array}{rrlll} \beta_i = & 1.0 & +0.1 \cdot x_i & + u_{\beta\,i} & u_{\beta\,i} \sim \mathcal{N}(0, 0.10) \\ \log(h_i) = & -6.91 & +0.0 \cdot x_i & + u_{h\,i} & u_{\beta\,i} \sim \mathcal{N}(0, 0.10) \\ \Phi^{-1}(\pi_i) = & 1.28 & -0.3 \cdot x_i & + u_{\pi\,i} & u_{\beta\,i} \sim \mathcal{N}(0, 0.05) \\ \Phi^{-1}(\lambda_i) = & -0.84 & +0.0 \cdot x_i & + u_{\lambda\,i} & u_{\beta\,i} \sim \mathcal{N}(0, 0.05) \end{array}$$

The intercepts of each equation are set so that for an average asset (that is, when x is equal 0) the expectation of β is 1, annualized idiosyncratic volatility is 50%, persistence π_i is 0.90 and smoothness λ_i is 0.20. The Factor GARCH unconditional annualized volatility is 30% and α_F and β_F are respectively 0.02 and 0.97. For each of the 100 assets I simulate 500 days of data. The simulation is replicated 100 times keeping the values of the characteristics x fixed across simulations. The model is then estimated using the methodology described in Section 2 resorting to different levels of accuracy (2, 4, 6, 8, 10, 12).

I choose one of the one hundred replications for illustration purposes. The estimation results with accuracy level 12 deliver the following estimates (standard errors in parenthesis and significant characteristics in bold)

$$\begin{array}{rrlll} \beta_i = & 1.0 & +\underset{(0.0027)}{0.109} \cdot \boldsymbol{x}_i & + u_{\beta\,i} & u_{\beta\,i} \sim \mathcal{N}(0, \underset{(0.0021)}{0.099}), \\ \log(h_i) = & \underset{(0.0168)}{-6.911} & \underset{(0.0084)}{-0.008} \cdot x_i & + u_{h\,i} & u_{\beta\,i} \sim \mathcal{N}(0, \underset{(0.0046)}{0.094}), \\ \Phi^{-1}(\pi_i) = & \underset{(0.0449)}{1.321} & \underset{(0.0204)}{-0.277} \cdot \boldsymbol{x}_i & + u_{\pi\,i} & u_{\beta\,i} \sim \mathcal{N}(0, \underset{(0.0053)}{0.045}), \\ \Phi^{-1}(\lambda_i) = & \underset{(0.0233)}{-0.839} & \underset{(0.0300)}{+0.030} \cdot x_i & + u_{\lambda\,i} & u_{\beta\,i} \sim \mathcal{N}(0, \underset{(0.0052)}{0.046}). \end{array}$$

An R^2 type statistics is used to measure the goodness of fit of the equations with significant characteristics. Let the expected linear predictors for β_i and $\Phi^{-1}(\pi_i)$ be defined respectively as $\hat{\eta}_{\beta\, i} = \hat{\delta}_{\beta\, 0} + \hat{\delta}_{\beta\, 1} x_i$ and $\hat{\eta}_{\pi\, i} = \hat{\delta}_{\pi\, 0} + \hat{\delta}_{\pi\, 1} x_i$. Goodness of fit measures can be defined as

$$R^2_\beta = \frac{s^2(\hat{\eta}_{\beta\, i})}{s^2(\hat{\eta}_{\beta\, i}) + \hat{\tau}^2_\beta} \quad \text{and} \quad R^2_\pi = \frac{s^2(\hat{\eta}_{\pi\, i})}{s^2(\hat{\eta}_{\pi\, i}) + \hat{\tau}^2_\pi},$$

which are essentially variance ratios of the explained variance over total variance. In the sample estimates, the R^2_β and R^2_π are respectively 8.5% and 62.8% (the population analogs are 8.7% and 61.6%).

Figure 2 reports the plots of the actual and predicted random coefficients β and π against the characteristic x, together with the actual and estimated regression line. The figure also reports the estimates of the coefficients that can be obtained by estimating the 100 series individually (hence, ignoring the hierarchical structure) using ML. Visual inspection of the graphs suggests that the estimation of β is reasonably precise and that the random coefficient predictions (Hierarchical Factor GARCH) and the fixed coefficient estimates (Factor GARCH) are close. In fact, the rank correlation with the latent coefficients is 0.98 in both cases and the MSEs are, respectively, 2.56×10^{-3} and 2.58×10^{-3}. This does not make a strong case for using the hierarchical specification if one is only interested in the fitting the data, but it is nevertheless reassuring to see that the random and fixed methodology substantially coincide when the data contain sufficient information about the coefficient. The estimation of π delivers a different story. The pooling effect around the regression line of the random coefficient predictions is evident, and it appears to improve precision over the fixed coefficient estimates. Rank correlations with the latent coefficients are, respectively, 0.84 and 0.74 and the MSEs are 1.05×10^{-3} and 4.84×10^{-3}, almost five times more precise.

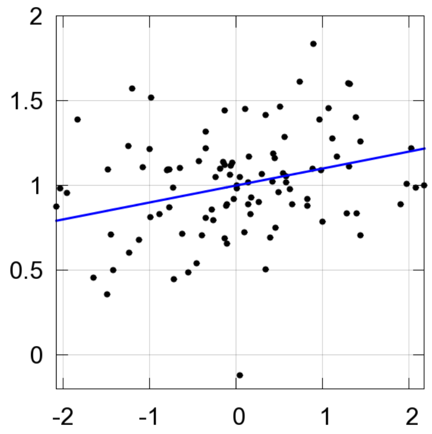
A useful feature of hierarchical modelling is that it allows one to predict/impute the random coefficients of a new asset with given characteristics x. Analogously, one can compute the expected coefficients for representative firms A and B with asset characteristics x_A and x_B. An insightful way to examine the differences implied by the model is to plot functions of the coefficients, such as moments, autocorrelation function, impulse response functions and so on. I illustrate the construction of a hierarchical impulse response function as an example. Let A be an average asset with $x_A = 0$ and B an asset with moderate levels of x ($x_B = 1.5\sigma_x$). In period $t = 0$, the variance of the factor and idiosyncratic innovation of both companies are at their steady states. In period $t = 1$, both assets receive a relative idiosyncratic shock ν of the same magnitude, equal to the 0.99 quantile of a Gaussian random variables with zero mean and unit variance. The response of the idiosyncratic variance at time $t + 1$ is

$$h_{i\, t+1} = h_i(1 - \pi_i) + \pi_i \left[\lambda h_i\, \nu^2 + (1 - \lambda) h_i\right] \quad i \in \{A, B\}$$

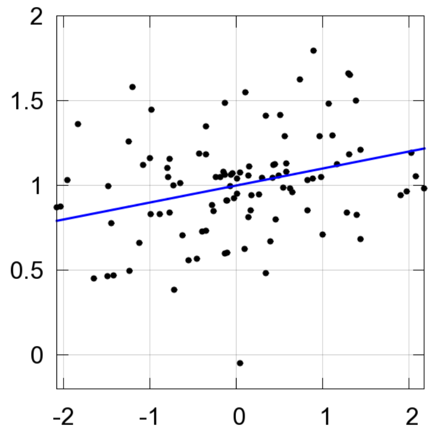
and the total asset variance from period $t + 1$ to the future is equal to

$$\mathsf{Var}_{t+k|t}(r_t) = \beta_i^2 h_m + h_i(1 - \pi_i) + \pi_i^{k-1}(h_{i\, t+1} - h_i).$$

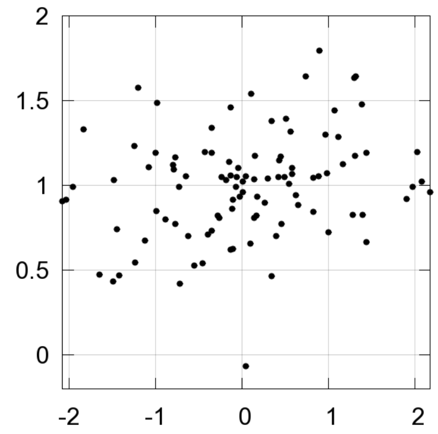
The graph of the hierarchical impulse response functions for companies A and B is reported in Figure 3. Since β_i increases in the characteristic x, company B has a higher long term variance because it has a higher systematic variance component. However, as π_i decrease in x, company A has higher short term because it has more variance persistence and it takes longer for A to absorb the shock.



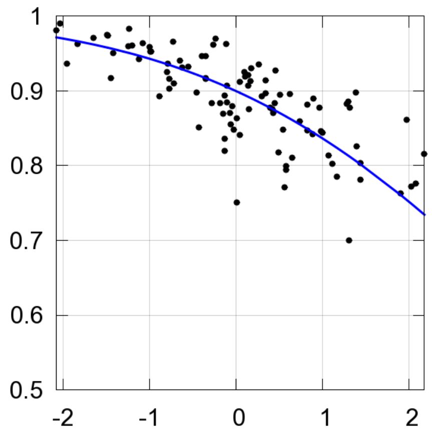
(a) β_i vs. x_i



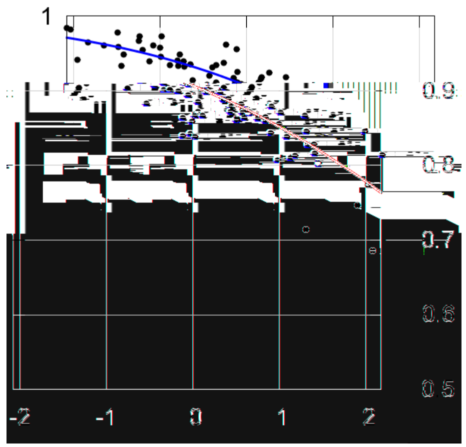
(b) $\hat{\beta}_i$ vs. x_i



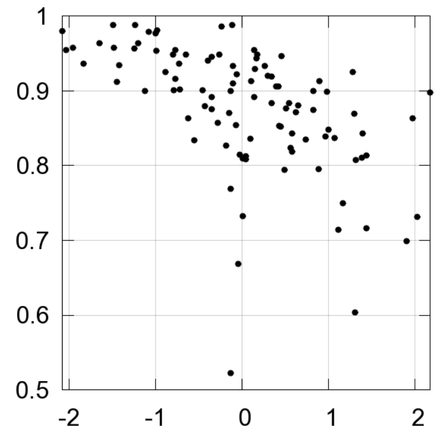
(c) $\tilde{\beta}_i$ vs. x_i



(d) π_i vs. x_i



(e) $\hat{\pi}_i$ vs. x_i



(f) $\tilde{\pi}_i$ vs. x_i

Figure 2: Coefficients versus asset characteristic. The figure plots the actual factor loading β_i and persistence π_i realizations versus the asset characteristic x togther with the actual regression curve ((a) and (d)); the Hierarchical Factor GARCH coefficients predictions with the estimation regression curve ((b) and (e)) and the Factor GARCH coefficient estimates ((c) and (f)).

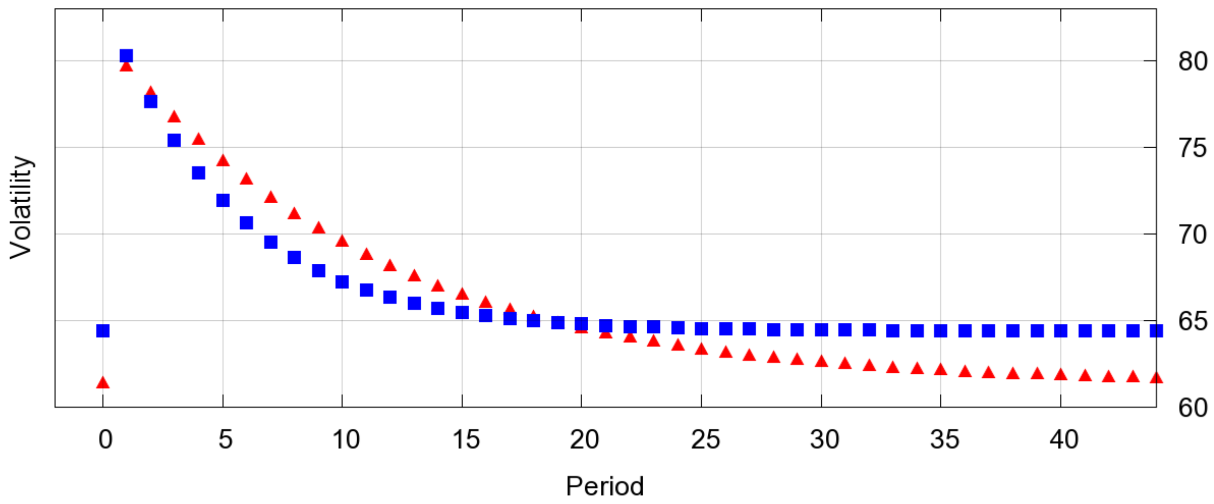


Figure 3: Hierarchical impulse response function. Model implied forecast function for representative company A (triangle) and company B (square).

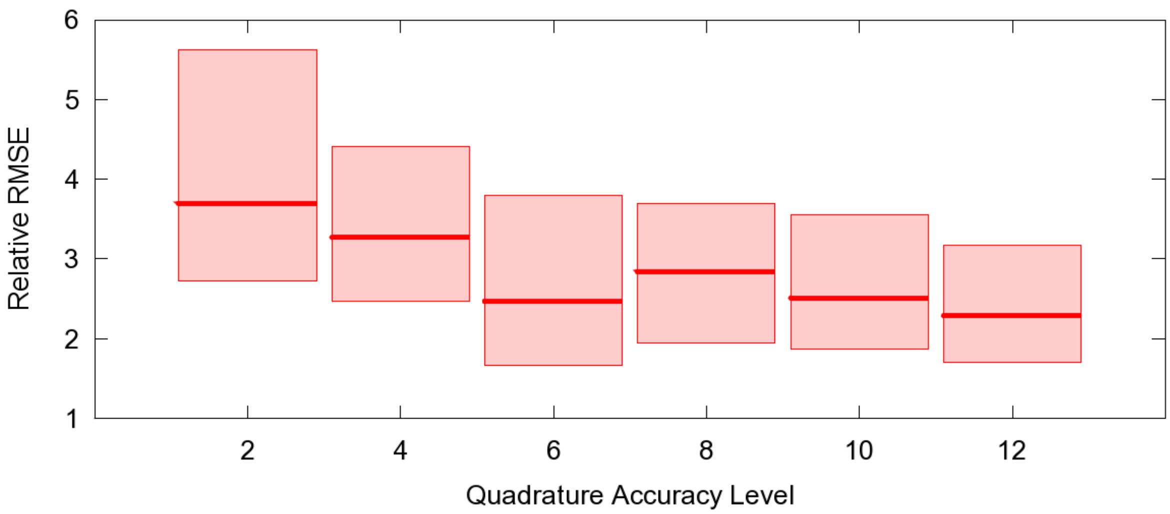


Figure 4: Relative Monte Carlo RMSE. Monte Carlo precision of the hyper parameter estimator.

The 100 replications give insights on the finite sample precision of the estimation procedure of the hyper parameter. Figure 4 reports the mean and 90% quantiles of the root mean square estimation error of the hyper parameter as a function of the accuracy level of the quadrature rule. The root mean square error is reported in relative terms, that is the mean and quantiles of the plot are scaled by the root mean square error of the unfeasible Maximum Likelihood estimator that also uses the random effect realizations. The RMSE is, not surprisingly, well above one. When the level of accuracy is low it is up to almost six times larger than the unfeasible benchmark. However, as the degree of accuracy increases, the RMSE steeply improves, being in bewteen two to three when the level of accuracy is high. Detailed inspection of the result suggests that the random effect variances τ^2 have higher RMSE, and also appear to be slighlty downward biased when the dimension of the cross sections is moderate.

Results show that the estimation approach works sufficiently well. I would like to stress that there is a feature of the data that is replicated in the Monte Carlo setting that makes the estimation work. That is *coefficient clustering*, the fact that there is cross sectional clustering of the coefficients. When the degree of cross sectional variability is high or absent, implying that assuming coefficients belong to some population becomes less evident, the efficiency of the RMSE deteriorates.

4 Application to the 2007/2009 Financial Crisis

I analyse top U.S. financial firms during the crisis in the attempt to identify which pre-crisis characteristics explain the cross sectional variation of volatility dynamics during the crisis. Asset dynamics are assumed to be described by the Hierarchical Factor GARCH model introduced in Section 2 and the set of firm characteristics considered are size, leverage, distance to default and liquidity. Of particular interest in this context is the assessment of how variables relate to the factor loading β_i. In this framework, a systemically important firm with factor exposure is systemically risky in that common factor drop will generate large losses to the firm and negative externalities to the rest of the economy.

The empirical application builds up on the ideas of Acharya, Pedersen, Philippe and Richardson (2010), which propose to measure systemic risk using the Marginal Expected Shortfall (MES), the expected loss of an asset when the whole sector has a downturn. In Acharya, Pedersen, Philippe and Richardson (2010) and Brownlees and Engle (2010) the aim of the analysis is *real time* exposure measurement and elaborate models of dependence are introduced. Brownlees and Engle (2010) propose a method for modelling MES based on time varying volatility, correlation and a tail correction to capture potential nonlinear dependence in the innovation process. On the other hand, here the focus is on *average* exposure measurement during the crisis and the linkage with cross sectional characteristics, and the simple factor loading β_i suffices.

On a general note, I would like to stress that a simple factor model with heteroskedastic components is able to describe the evolution of risk in the crisis. This is somehow in contrast with the popular view that has been stressing the inadequacies of the models proposed in the literature because of their inability to capture complex forms of dependence (cf. Taleb (2007) comments on ARCH models). Particularly, it has been emphasized that the clustering of the extreme losses of several financial institutions observed in the Fall of 2008 is not captured by standard models. While the research for more sophisticated methods to capture nonlinear dependence is of great importance (cf. Brownlees and Engle (2010)), it is inaccurate to claim that current financial econometrics machinery is inept for this task. In fact, in the simple Factor GARCH even if the *conditional* distribution of returns does not allow for extreme event clustering, the *unconditional* one allows for high levels of dependence in the extremes of the process. Conditional volatility and average correlation are increasing functions of the volatility of the factor, hence, extreme factor volatility determines extreme joint realizations of the process.

4.1 Data Description

The analysis focuses on all U.S. Financial firms with a market capitalization greater than 5 bln USD as of the end of June 2007; the same panel of institutions studied in Acharya, Pedersen, Philippe and Richardson (2010). Table 1 contains the full list of tickers and company names. The dataset is constructed using the CRSP and COMPUSTAT databases.

Ticker	Name	Ticker	Name
ABK	AMBAC Financial Group	HUM	Humana
ACAS	American CapitaL	ICE	Intercontinental Exchange
AET	Aetna	JNS	Janus Capital
AFL	Aflac	JPM	J.P. Morgan Chase
AGE	A. G. Edwards	KEY	KeyCorp
AIG	American International Group	LEH	Lehman Brothers
AIZ	Assurant	LM	LEGG Mason
ALL	Allstate	LNC	Lincoln National
AMP	Ameriprise Financial	MBI	MBIA
AMTD	TD Ameritrade	MER	Merrill Lynch
AON	Aon Corporation	MET	MetLife
AXP	American Express	MI	Marshall & Ilsley
BAC	Bank of America	MMC	Marsh & McLennan
BBT	BB&T	MS	Morgan Stanley
BEN	Franklin Resources	MTB	M&T Bank
BK	Bank of New York Mellon	NCC	National City Corporation
BLK	Blackrock	NMX	NYMEX
BOT	CBOT Holdings	NTRS	Northern Trust
BRK	Berkshire Hathaway	NYB	New York Community Bank
BSC	Bear Stearns	NYX	NYSE Euronext
C	Citigroup	PBCT	Peoples United
CB	Chubb Corporation	PFG	Pripal Financial Group
CBG	C. B. Richard Ellis	PGR	Progressive
CBH	Commerce Bank	PNC	PNC Financial
CBSS	Compass Bancshares	PRU	Prudential Financial
CFC	Countrywide Financial	RF	Regions Financial
CI	Cigna Corporation	SAF	Safeco
CINF	Cinnati Financial Corporation	SCHW	Schwab
CIT	CIT Group	SEIC	SEI Investments
CMA	Comerica	SLM	SLM
CME	CME Group	SNV	Synovus
CNA	CNA Financial Corporation	SOV	Sovereign Bank
COF	Capital One Financial	STI	SunTrust
CVH	Coventry Health Care	STT	State Street
ETFC	E-Trade	TMK	Torchmark
EV	Eaton Vance	TROW	T. Rowe Price
FITB	Fifth Third	TRV	St. Paul Companies
FNF	Fidelity National Financial	UB	UnionBanCal
FNM	Fannie	UNH	UnitedHealth Group
FRE	Freddie	UNM	Unum
GNW	Genworth Financial	USB	U. S. Bancorp
GS	Goldman Sachs	WB	Wachovia
HBAN	Huntington Bancshares	WFC	Wells Fargo
HCBK	Hudson City Bank	WM	Washington Mutual
HIG	Hartford Financial	WRB	W. R. Berkley Corporation
HNT	Health Net	WU	Western Union
HRB	H & R Block	ZION	Zions Bank

Table 1: Tickers and company names.

	$Q_{0.10}$	Median	$Q_{0.90}$	XLF
Ann. Mean	-153.28	-37.82	-6.22	-51.76
Ann. Vol.	48	80.1	146.3	68.1
Kurt.	5.48	8.41	58.82	5.68
$\rho_1(r^2)$	0.04	0.16	0.27	0.19

Table 2: Daily returns descriptive statistics in the financial crisis. The table reports the 10%, 50% and 90% quantiles of the annualized mean, annualized volatility, kurtosis and first order autocorrelation of the squared return together with the value of the same statistics for the XLF.

Figure 5: Financial (XLF) and Volatility (VIX) indexes.

I study daily return volatility dynamics in the two year period July 2, 2007 to June 30, 2009, which delivers a maximum of 504 trading days for each asset. The sample also roughly matches the NBER late 2000's recession defined as the period December 2007 to July 2009. As customary, daily returns are defined as log differences of the adjusted closing price. I choose to analyse this two year period only to specifically focus on dynamics during the crisis. Parameter estimates of GARCH type models typically exhibit slow moving variation (see for instance the evidence in Shephard and Sheppard (2010) or Brownlees et al. (2010)) and a wider sample would dampen crisis specific features. Out of the 94 companies in the panel, 15 do not make it to the end of the sample and have traded for as little as 8 days.[7] Even if the contribution of such companies is negligible, they have not been excluded from the sample to illustrate how the hierarchical methodology works in such conditions. A preliminary factor analysis suggests that a one factor structure is a reasonable assumption for the panel. The first principal component explains 49.0% of the total variability of the dataset followed by the second and third components capturing 9.9% and 4.8%. I take as a common factor the daily returns on the Financial Select Sector SPDR ETF (XLF). The series is strongly correlated with the equally weighted portfolio return constructed from the panel (96.0), the first factor from a principal components analysis (96.0) and also with the S&P 500 index (87.1). Figure 5 plots the XLF and the Volatility Index (VIX) from June 2004 through December 2009, the crisis period is the shaded gray area. Note that in the crisis the XLF index has dropped more than 80% from the peaks it reached in July 2007. Summary descriptive statistics on the financial returns panel and the common factor are reported in Table 2.

Asset characteristics are computed using quarterly accounting and daily market data as of the end of June 2007. The first characteristic is the size of the firm, denoted as siz and measured as market capitalization. I consider leverage (lvg), the proportion of debt to finance the firm. The abuse of leverage by financial institutions before the crisis is often blamed as one of the

[7] I adhere to CRSP for the determination of the end dates of each stock.

major sources of weakness that led to the crisis. I measure this characteristic as quasi market leverage

$$\mathsf{lvg}_i = \frac{\mathsf{BA}_i - \mathsf{BE}_i + \mathsf{ME}_i}{\mathsf{ME}_i},$$

where BA, BE and ME are respectively the book value of asset, book value of equity and market value of equity. I also employ distance to default to capture a firm's (in)ability to generate profits and distance from insolvency. There are different approaches in the literature to measure this quantity and I use the same indicator employed in, among others, Laeven and Levin (2009) and Beltratti and Stultz (2010), that is

$$\mathsf{dtd}_i = \frac{\mathsf{ROA}_i + \mathsf{CAR}_i}{sd(\mathsf{ROA}_i)},$$

where ROA stands for return on assets, CAR is the capital asset ratio and $sd(\cdot)$ denotes the (historical) standard deviation of ROA. The rationale of the measure is that the event $\mathsf{CAR}_i < \mathsf{ROA}_i$ can be considered as insolvency, hence $\mathsf{ROA}_i + \mathsf{CAR}_i$ measures the distance from insolvency. The adjustment by $sd(\mathsf{ROA}_i)$ is a customary normalization done in the literature. Lastly, I consider liquidity, the capacity of a firm to turn assets into cash to pay short term debt. Measuring liquidity for financial institutions is inherently challenging given the intermediary nature of the business itself. I proxy it using the proportion of assets that can be easily transformed in cash, that is

$$\mathsf{liq}_i = \frac{\mathsf{CS}_i}{\mathsf{BA}_i},$$

where CS is the sum of cash and short term investments and BA is the book value of assets. The choice of these accounting ratios is related to the financial statement analysis practice. Typically, the ratios used to infer the fundamental value of a firm are divided into three categories, which are stability, profitability and liquidity; and the characteristics I use attempt to broadly cover these three different sources of information coming from financial statements.

Table 3 reports descriptive statistics on the set of variables. It is interesting to notice that even if asset characteristics change in time, the cross section appears to be rather stable between 2004 and 2007. The correlations between the variables as of June 2007 and December 2004 are between 84% and 98%. The degree of dependence among regressors is modest, with maximum cross section correlation in absolute value being 0.26.

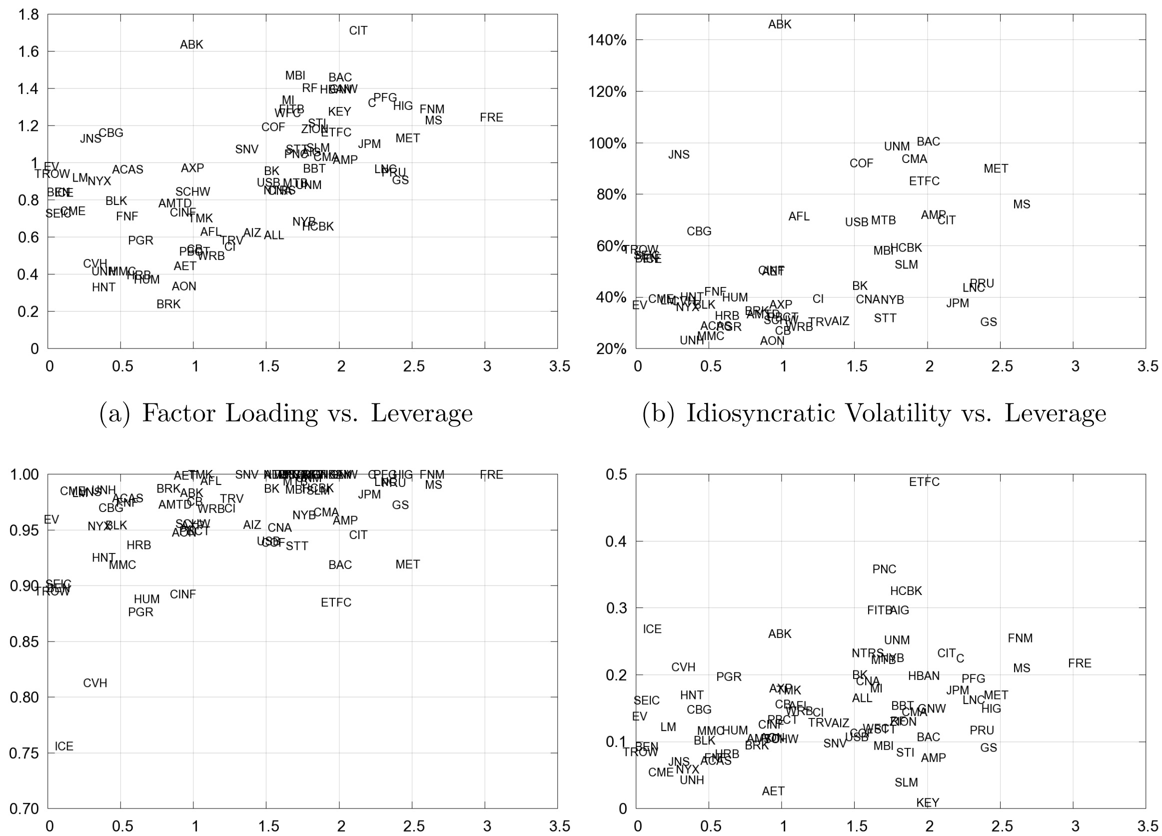
4.2 Preliminary Analysis

The time series analysis of volatility dynamics during the crisis turns out to be quite challenging because of the presence of strong heteroskedasticity, extreme returns and the limited time span of the period of interest.

For each asset in the panel, I estimate a Factor GARCH with fixed coefficients by Maximum Likelihood (ML). ML estimation under the Gaussian assumption turns out to be highly sensitive to the presence of outliers. In order to avoid extreme returns having too much of an impact on the estimates I switch to the Student t-distribution, which improves results.[8] Nevertheless, the nonlinear estimation procedure does not converge for 16 out of 94 series. As documented, among others, in Brownlees et al. (2010), GARCH models can be quite data hungry. The

[8] Trimming the outliers produces estimates which are substantially close to the ones that one gets using Student t-distribution.

parameter estimates exhibit a certain degree of cross sectional dispersion but, with the exception of smoothness parameter λ_i, the average values are close to the typical estimates. The λ_i parameter has an average value of 0.17, which can be interpreted as evidence of higher levels of roughness in the volatility path and unconditional kurtosis during the crisis period. This is also consistent with the evidence of Shephard and Sheppard (2010) on the performance of the GARCH model on the S&P500 during the crisis.[9]



(a) Factor Loading vs. Leverage

(b) Idiosyncratic Volatility vs. Leverage

(c) Persistence vs. Leverage

(d) Smoothness vs. Leverage

Figure 6: Preliminary analysis. Fixed coefficient estimates versus leverage. The fixed coefficients estimates are the ML estimates of the Factor GARCH model.

I construct scatter plots of (log) asset characteristics against parameter estimates to investigate the presence of cross sectional patterns in the coefficients. For sake of space, I only report the ones obtained using leverage in Figure 6. The scatters suggest that leverage is positively related to all the Factor GARCH coefficients. The most convincing case is seen in the factor loading β_i while the less clear cut case is probably found in the persistence π_i. Generally speaking, the visual inspection of these scatters gives interesting clues as to which characteristics are more promising explanatory variables.

[9] They predict conditional variance using GARCH estimates obtained from a longer sample period which delivers (in my parameterization) a low λ_i. However, they notice (cf. Section 4.6 of Shephard and Sheppard (2010)) that GARCH conditional volatility adjust too slowly to the turmoil of the crisis, implying that a higher λ_i would have performed better.

It is also interesting to comment on the behaviour of the simple least squares (LS) estimator of the factor loading β_i during the crisis.[10] As is well known, in the presence of heterskedasticity the estimator is consistent yet not efficient. During the financial crisis however, the degree of heteroskedascticity, together with the magnitude and number of extreme observations, is such that the LS estimates overestimate the factor loading β_i by an economically significant distance. Comparing the β_i estimate obtained by ML and LS, one can see that out of the 78 successfully converged ML estimates, in 10 cases the LS ones overshoot the parameter from 20% to 52%. In the remaining cases, LS provides 5 estimates above 2 and up to 4 (Lehman Brothers).

4.3 Estimation Results

I estimate the Hierarchical Factor GARCH presented in Section 2 by Maximum Likelihood using SGI with accuracy level 10. The set of explanatory variables siz, lvg, dtd and liq are transformed in logs and are demeaned so that the intercept of each equation can be interpreted as the average of a representative firm. The preliminary analysis suggests that the Gaussian assumption is too restrictive and I resort to Student t-distribution innovations using a common number of degrees of freedom for each series. This creates a minor glitch in the specification. Conditional moments of the process are still defined in the usual manner but the joint conditional distribution of the returns is not well known. As it is well known, multivariate generalisations of the Student t-distribution are available (cf. Kotz and Nadarajah (2004)) but I do not pursue the exercise here.

The model fitting process of hierarchical nonlinear dynamic models via numerical integration is not too common in the financial econometrics literature and I devote some length describing it. This application is also somehow demanding given that the full specification is richly parameterised (the number of parameters is twenty-four). Firstly, the intercept (that is, the average) of the β_i equation is set to unity so there is one less parameter to be estimated. I begin by estimating the Hierarchical Factor GARCH allowing for asset characteristics one equation at a time and using a lower integration accuracy (level 6). The initial values of the intercepts and random effect variances are set according to the cross sectional means and variances of the preliminary analysis while the characteristics parameters are set to zero. Initial values are also jittered in order to avoid potential local minima. Typically, all parameters that turn out to be significant seem to capture the right sign even when the accuracy level is low. The random effect variance parameter $\tau_{\cdot}^2$ on the other hand appears to become more reliable when the accuracy level is higher. The estimates I get from this procedure are used as initial values for the full maximum likelihood estimation of all equations simultaneously and with high accuracy (level 10). These initial values appear to be reasonably close to the final estimates. The main difference between the simultaneous and equation by equation estimation regards the significance of the parameters that is weakened in the full estimation case, which is probably due to the high number of parameters.

The maximum likelihood estimation results delivers the following estimates (standard errors

[10] I am thankful to Robert Engle for initially making me notice the issue.

in parenthesis and significant characteristics in bold)

$$\begin{array}{rrrrrrr}
\beta_i = & 1.0 & +\underset{(0.0119)}{0.021} \cdot \textbf{siz}_i & +\underset{(0.0197)}{0.164} \cdot \textbf{lvg}_i & -\underset{(0.0142)}{0.092} \cdot \textbf{dtd}_i & -\underset{(0.0131)}{0.022} \cdot \textbf{liq}_i & + u_{\beta\, i}, \\
\log(h_i) = & \underset{(0.1121)}{-6.759} & -\underset{(0.1068)}{0.176} \cdot \textbf{siz}_i & +\underset{(0.0853)}{0.145} \cdot \textbf{lvg}_i & -\underset{(0.0825)}{0.027} \cdot \mathsf{dtd}_i & +\underset{(0.0903)}{0.055} \cdot \mathsf{liq}_i & + u_{h\, i}, \\
\Phi^{-1}(\pi_i) = & \underset{(0.3228)}{1.990} & +\underset{(0.3663)}{0.049} \cdot \mathsf{siz}_i & +\underset{(0.3204)}{0.112} \cdot \mathsf{lvg}_i & +\underset{(0.2356)}{0.059} \cdot \mathsf{dtd}_i & -\underset{(0.2054)}{0.043} \cdot \mathsf{liq}_i & + u_{\pi\, i}, \\
\Phi^{-1}(\lambda_i) = & \underset{(0.0453)}{-0.936} & +\underset{(0.0519)}{0.009} \cdot \mathsf{siz}_i & +\underset{(0.0614)}{0.119} \cdot \textbf{lvg}_i & +\underset{(0.0495)}{0.034} \cdot \mathsf{dtd}_i & +\underset{(0.0435)}{0.058} \cdot \mathsf{liq}_i & + u_{\lambda\, i},
\end{array}$$

and the estimates of the random effects variances are

$$u_{\beta\, i} \sim N(0, \underset{(0.0027)}{0.055}) \quad u_{h\, i} \sim N(0, \underset{(0.0674)}{0.190}) \quad u_{\pi\, i} \sim N(0, \underset{(0.0719)}{0.030}) \quad u_{\lambda\, i} \sim N(0, \underset{(0.0211)}{0.048}).$$

The null hypothesis of no joint effect of the asset characteristics is strongly rejected by a Likelihood Ratio Test. In order to make the estimation easier to read, I report the Median and the 90% coverage probability of the coefficients of an average firm, together with the R^2 of each equation in Table 4.

The factor loading β_i equation is the one that has more significant relations with asset characteristics. In the crisis, the factor loading is increasing in size and leverage and is decreasing in distance to default and liquidity, with leverage being the most relevant effect. The link of the factor loading with size is somehow surprising as typically the larger the size of a company, the higher the degree of diversification, the lower the factor loading with respect to the common factor. Results show that, even after controlling for other characteristics, large firms have more systematic exposure than the small ones during the crisis. This can also be related to an interconnectedness effect. It is fair to say that, everything being equal, a large financial institution is bound to be more interconnected than a small one and more interconnected institutions are believed to have a higher exposure in the crisis. In the idiosyncratic variance h_i equation only size and leverage are significant with signs that are in line with expectations: idiosyncratic volatility is higher for smaller institutions and more levered institutions. Persistence π_i is not significantly explained by any of the asset characteristics and furthermore the variance of its random effect appears to be only mildly significant. This provides evidence that the degree of volatility memory is not firm specific but is common to all firms. Finally, the smoothness coefficient λ_i is positively related to leverage, meaning that levered firms have a rough volatility path and are more prone to abrupt volatility shocks. Overall estimation results show that the factor loading is the parameter with more cross sectional links and that leverage is the most relevant asset characteristic. The R^2 of the various coefficients reported in Table 4 show that the model is capturing a relevant portion of the overall variation, with the factor loading R^2_β being the highest.

Conditional volatility and correlation time series plots provide a synthesis of the dynamics in the crisis. I report the volatility of the common factor as well as the one of the global minimum variance portfolio (GMV) in the top panel of Figure 7. The two series are on a relative scale to facilitate comparisons. They are standardised by the level of volatility they have on the first day of the sample. The GMV volatility can actually be computed in close form in $O(N)$ operations, which is appealing for large cross sectional applications (see Proposition 3 in the appendix). The correlation of an average firm is reported on the bottom panel of Figure 7. The correlation is computed using the following formula

$$\rho_{\mathsf{x}\, t} = \frac{\beta_{\mathsf{x}}\, h_{F\, t}}{\sqrt{h_{F\, t}\, (\beta^2_{\mathsf{x}}\, h_{F\, t} + h_{\mathsf{x}})}},$$

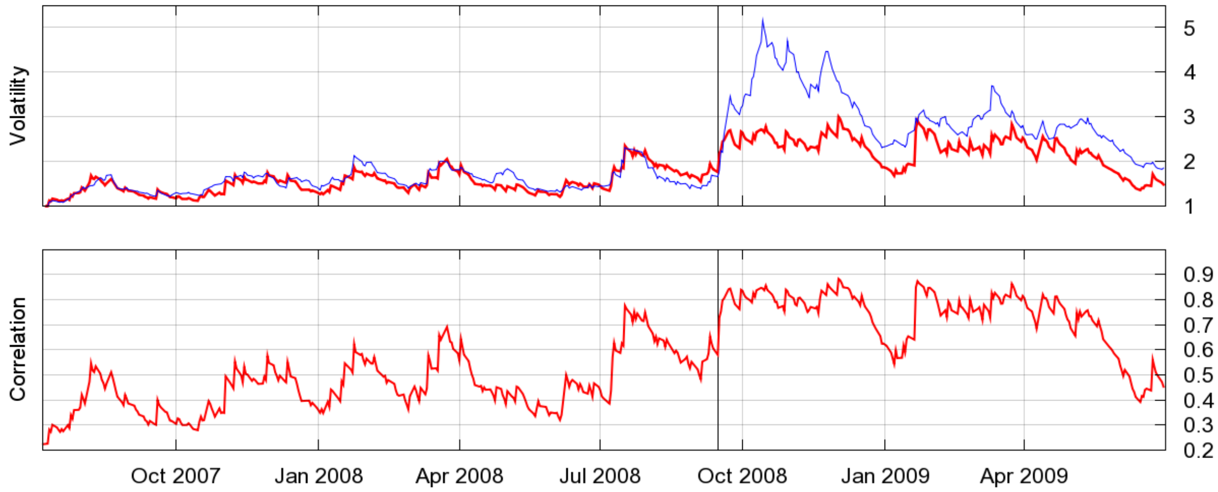


Figure 7: Volatility and correlation in the crisis. The top panel shows the volatility of the financial industry factor (thick line) and the volatility of the global minimum variance portfolio (thin line). The bottom panel shows the correlation of an average firm with the financial factor. The vertical line marks September 15, 2008, the day Lehman filed for Chapter 11, the largest bankruptcy in U.S. history.

where β_{x}, β^2_{x} and h_{x} are the expected factor loading, squared factor loading and idiosyncratic variance of a representative firm with characteristics x, that is

$$\beta_{\mathsf{x}} = \delta_{\beta\,0} + \sum_{k=1}^{p} \delta_{\beta\,k}\,\mathsf{x}_k, \;\; \beta^2_{\mathsf{x}} = \left(\delta_{\beta\,0} + \sum_{k=1}^{p} \delta_{\beta\,k}\,\mathsf{x}_k \right)^2 + \tau^2_\beta, \;\; h_{\mathsf{x}} = \exp\left\{ \delta_{h\,0} + \sum_{k=1}^{p} \delta_{h\,k}\,\mathsf{x}_k + \tau^2_h/2 \right\}.$$

The plot is then constructed by setting the characteristics x to their averages. The plot documents a steady increase in volatility and correlation in crisis. Also, factor volatility dynamics in this period are rougher than typical GARCH estimates (α_F and β_F are, respectively, 0.14 and 0.86). From the beginning of the sample to the mid of September 2008 volatilities and correlation increase by, respectively, roughly 100% and 200%. After the liquidation of Lehman on September 15th, the turbulence reaches its climax. To factor and GMV volatility have a further steep increase and correlation gets beyond 0.80. The failure of Lehman also marks a steep relative increase of GMV over the factor volatility. At the peak of the crisis, many companies suffered the most severe falls (Lehman, Washington Mutual, Freddie and Fannie, to name a few), which in turns leads to an increase in the level of idiosyncratic volatility and, consequently, GMV volatility.

Figure 8 shows the scatter of the crisis annualized mean return and the factor loading β_i. The inspection of the graph reveals that, not surprisingly, many of the most troubled companies were indeed the ones with higher factor exposure, to name a few, Lehman, Washington Mutual, Bear Sterns, Countrywide Financial, Freddie and Fannie. The rank correlation between the two series is 74.3%.

Inspection of the predicted specificities $\hat{z}_i = \hat{u}_i/\sqrt{\hat{\tau}^2}$ is used to check the adequacy of the normality assumption of the random effects. For each coefficient I construct a specificity histogram and test the Null hypothesis of Gaussianity using a Jarque-Bera test. Results are reported in Figure 9. The Gaussian assumption is adequate for the factor loading and idiosyncratic volatil-

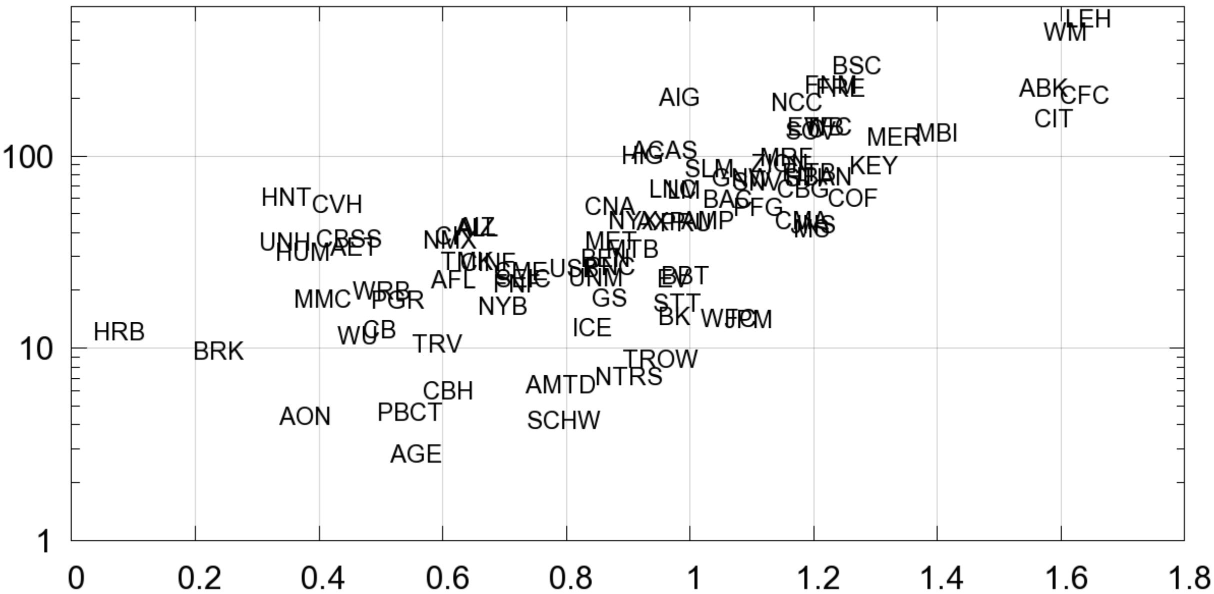


Figure 8: Financial crisis loss versus factor loading.

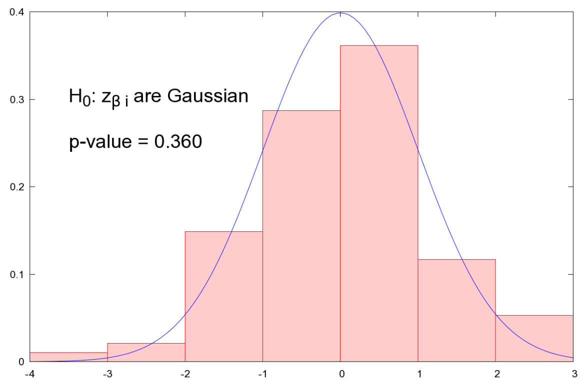
ity, but it is rejected for the persistence and smoothness. The histograms of these last two effects suggest that there are a couple of outlying firms which lead the null to be rejected.

The predictions of the individual specificities permit one to identify which firms behave in a substantially different manner from what their characteristics would otherwise imply. In the case of the factor loading, a large negative specificity prediction can be intepreted as the market evaluating the firm more market neutral than other similar companies.[11] In this context this can be judged as the firm ability of lowering its market exposure. Figure 10 plots ranked factor loading specificity predictions of selected companies with 95% confidence interval. The solid lines in the graph denote the 95% confidence interval of a standard Gaussian random variable. Interestingly, the plot shows that Lehman and Washington Mutual have high factor loading specificity, signaling that even after controlling for size, leverage, distance to default and liquidity, the exposure of these companies in the crisis is large.

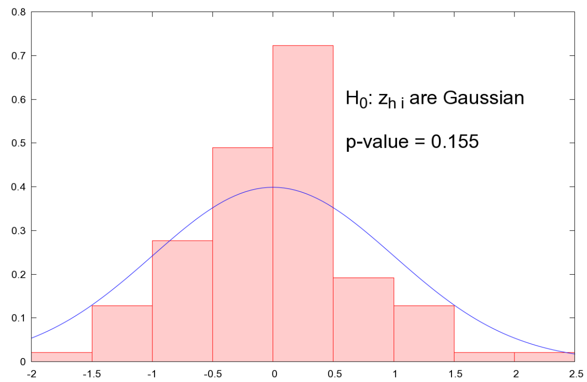
Figure 11 shows the scatter plot of the estimated coefficients versus leverage together with the estimated regression line. The pooling effect of the random effect is evident when comparing with fixed effect estimates in Figure 6. The shrinkage effect appears to adjust some of the outlying estimates obtained by fixed effect approach. For instance, the idiosyncratic variance of ABK which the fixed effect estimate set to above 140% is now slightly below 60%. The amount of shrinkage is not uniform across coefficients. The factor loading does not exhibit a strong shrinkage effect while the idiosyncratic variance gets notably shrunk. This is also in line with the evidence of the simulation exercise.

The estimation results allow one to construct impulse response functions whose shape is a function of firm characteristics to further explore model implied differences in the volatility dynamics. Figure 12 displays the impulse response function of three companies labelled as Average, Large and Levered. The Average firm has all characteristics set to their sample

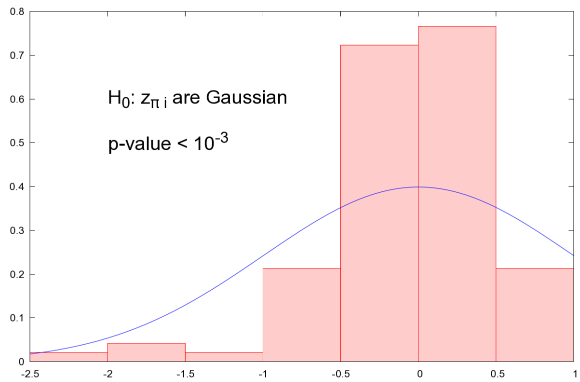
[11] This, of course, relies on controlling for enough observed characteristics which might not be the case in this application. In any case, I engage in this discussion for illustration purposes.



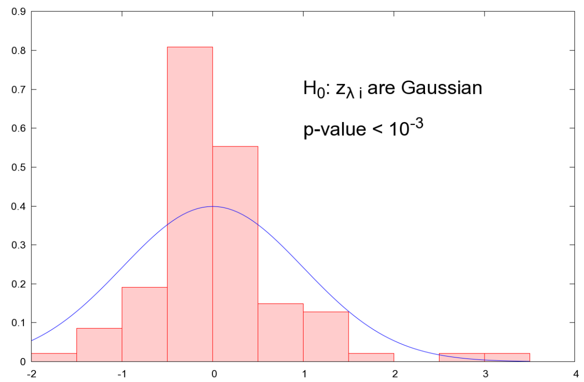
(a) Factor Loading Specificity Histogram



(b) Idiosyncratic Volatility Specificity Histogram



(c) Persistence Specificity Histogram



(d) Smoothness Specificity Histogram

Figure 9: Specificity histograms.

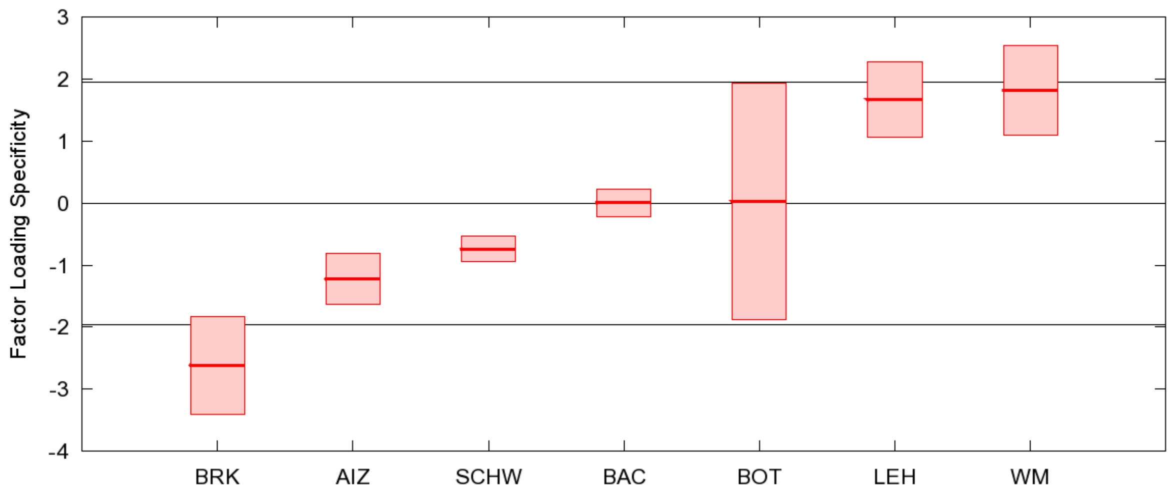
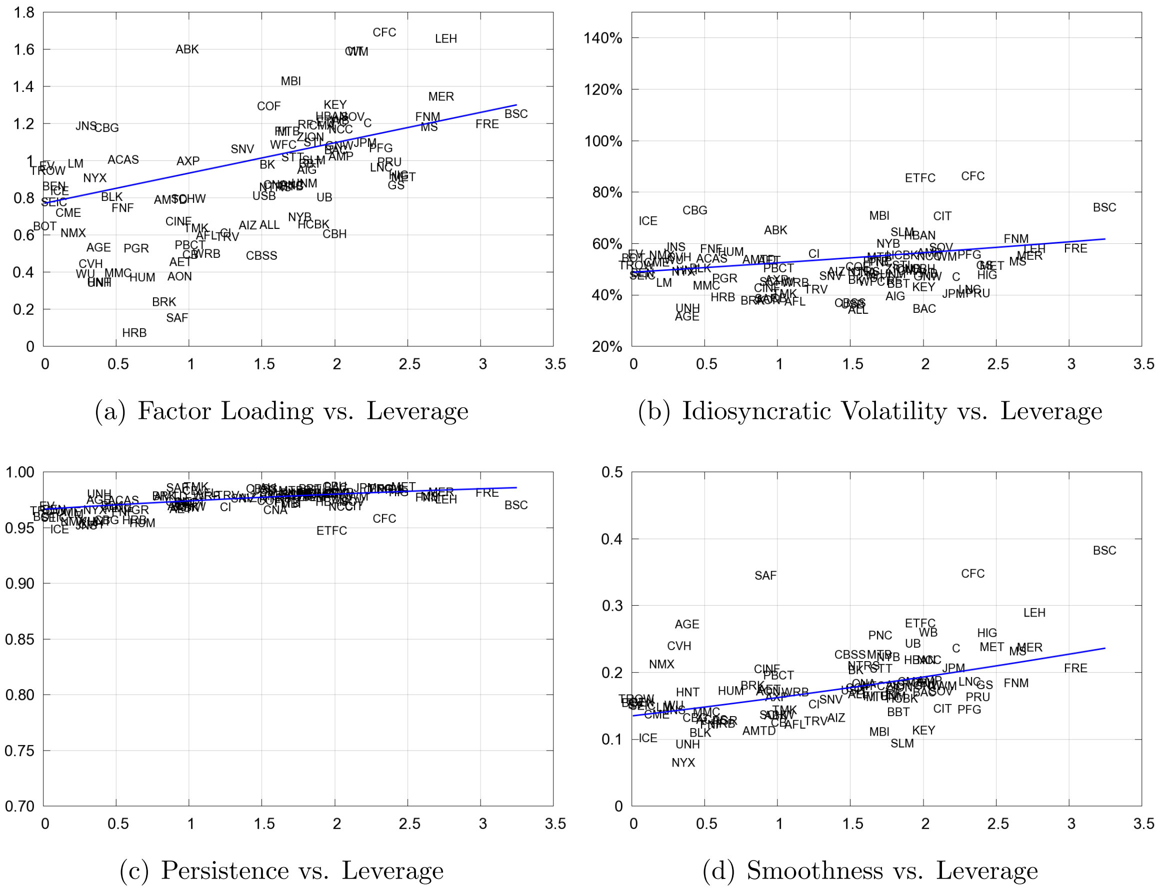


Figure 10: Factor loading specificity rankings for selected firms.

averages, while the Large and Levered firms are average firms with, respectively, siz and lvg set to 0.75% quantile of their respective empirical distributions. The impulse response functions



(a) Factor Loading vs. Leverage (b) Idiosyncratic Volatility vs. Leverage

(c) Persistence vs. Leverage (d) Smoothness vs. Leverage

Figure 11: Random effect predictions versus leverage.

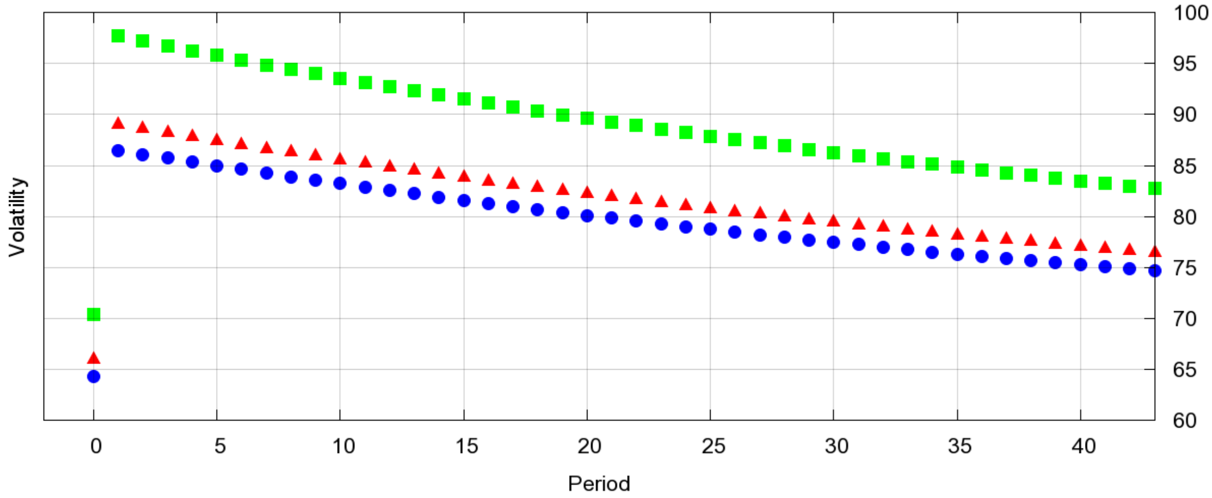


Figure 12: Hierarchical impulse response function for an average firm (triangle), levered firm (square) and large firm (circle).

are constructed using the same steps outlined in Section 3. The Average firm response to a volatility shock exhibits the typical pattern of GARCH forecasts, with a markedly slow decay in the absorption of volatility shocks. The Large firm is less volatile than the average in the steady state. During the crisis, β_i increases with size while h_i decreases, but ultimately the reduction of idiosyncratic volatility in h_i makes the total variance of the firm smaller. Volatility shocks to the Large firm behave exactly like those of the Average one. The Levered company has the highest level of volatility in the steady state, as leverage in the crisis buys both systematic and idiosyncratic volatility. On top of this, leverage also gets more volatility roughness, that is higher λ_i, and as a result the company has a more pronounced sensitivity to volatility shocks. Persistence of shocks is roughly equal for all firms, hence forecasts of all companies mean revert at the same speed. The plot also suggests some further remarks on the dynamics of volatility. It would be natural to consider tha persistence π_i and smoothness λ_i coefficients somehow less economically interesting in that they only capture temporary effects. However, the estimation results hint that temporary can be a pretty long time as the estimates of π_i are high and, roughly, the same for all firms. In such a context, the λ_i equation can be interpreted as the determinant of short term volatility in the presence of extreme shocks. From this perspective the estimation results show that in case of an extreme event levered firms are the ones affected by an extra increase in short term volatility with respect to all other firms.

It is of interest to carry out the same type of analysis in the period anticipating the turmoil to assess whether asset characteristics and volatility dynamics relate in the same way. Also, it is interesting to investigate if early symptoms can be detected. I estimate the same specification using data from January 2005 to June 2007 using asset characteristics measured using data as of December 2004. The maximum likelihood estimation results delivers the following estimates (standard errors in parenthesis and significant characteristics in bold)

$$\begin{aligned}
\beta_i &= \underset{}{1.0} \ \underset{(0.0196)}{-0.034}\cdot\mathbf{siz}_i \ \underset{(0.0345)}{+0.117}\cdot\mathbf{lvg}_i \ \underset{(0.0169)}{+0.010}\cdot\mathsf{dtd}_i \ \underset{(0.0221)}{-0.011}\cdot\mathsf{liq}_i + u_{\beta\, i},\\
\log(h_i) &= \underset{(0.1145)}{-6.682} \ \underset{(0.0543)}{-0.152}\cdot\mathbf{siz}_i \ \underset{(0.0960)}{+0.004}\cdot\mathsf{lvg}_i \ \underset{(0.0579)}{+0.004}\cdot\mathsf{dtd}_i \ \underset{(0.0564)}{+0.005}\cdot\mathsf{liq}_i + u_{h\, i},\\
\Phi^{-1}(\pi_i) &= \underset{(0.1214)}{2.016} \ \underset{(0.0837)}{+0.011}\cdot\mathsf{siz}_i \ \underset{(0.1367)}{+0.012}\cdot\mathsf{lvg}_i \ \underset{(0.0790)}{-0.014}\cdot\mathsf{dtd}_i \ \underset{(0.1058)}{-0.002}\cdot\mathsf{liq}_i + u_{\pi\, i},\\
\Phi^{-1}(\lambda_i) &= \underset{(0.0662)}{-1.285} \ \underset{(0.0338)}{+0.005}\cdot\mathsf{siz}_i \ \underset{(0.0511)}{+0.108}\cdot\mathbf{lvg}_i \ \underset{(0.0402)}{+0.008}\cdot\mathsf{dtd}_i \ \underset{(0.0380)}{+0.010}\cdot\mathsf{liq}_i + u_{\lambda\, i}.
\end{aligned}$$

and the estimates of the random effects variances are

$$u_{\beta\, i} \sim N(0, \underset{(0.0130)}{0.049}) \quad u_{h\, i} \sim N(0, \underset{(0.0168)}{0.091}) \quad u_{\pi\, i} \sim N(0, \underset{(0.0332)}{0.117}) \quad u_{\lambda\, i} \sim N(0, \underset{(0.0118)}{0.069}).$$

Again, a Likelihood Ratio Test overwhelmingly rejects the null hypothesis of no joint effect of the asset characteristics and Table 5 reports the Median and the 90% coverage probability of the coefficients of an average firm, together with the R^2 of each equation.

The factor loading equation has two significant explanatory variables only, size and leverage. Size restores its expected sign, that is large firms have small exposure while leverage has still a positive relation to systematic exposure. Idiosyncratic volatility is decreasing in size and leverage does not enter the equation significantly any more. The persistence equation and smoothness equation have the same significant links as in the crisis: persistence is not significantly related to any variables and smoothness is significantly explained by leverage, implying that levered firms have a rougher volatility dynamics.

It is also interesting to use the pre-crisis results to rank firms according to their exposure to the factor. I divide firms according to size quintiles (measured as of June 2007) and display

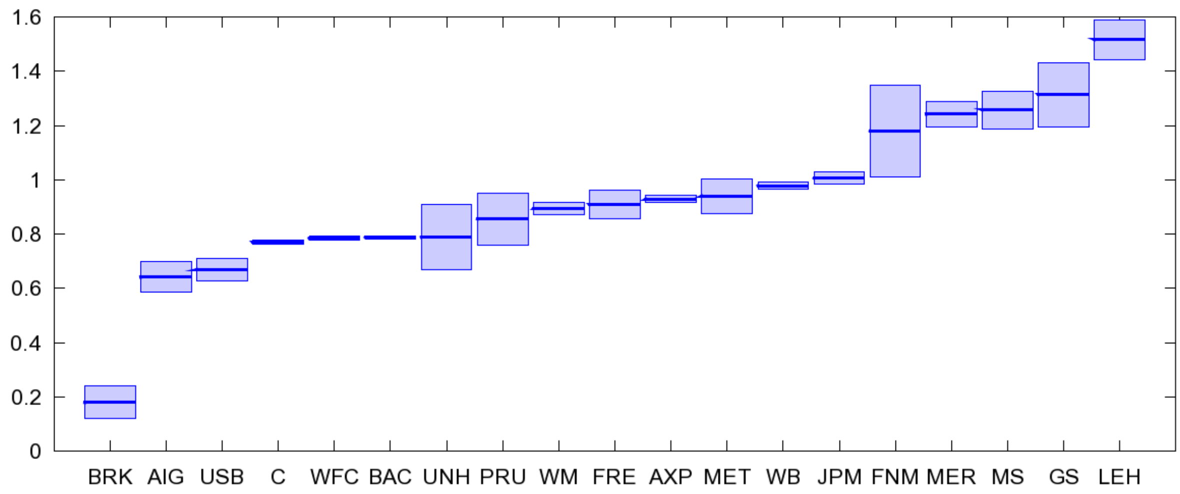


Figure 13: Pre-crisis factor loading rankings for the largest U.S. financials.

factor loading ranks of the top quintile in Figure 13.[12] This is in the spirit of Acharya, Pedersen, Philippe and Richardson (2010) who stress the importance of the construction of systemic risk ranking based on market data. Interestingly, the ranks clearly show Lehman Brothers (LEH) and top U.S. Investment banks are the most systematically exposed institutions before the crisis.

5 Conclusions

This work proposes the use of hierarchical modelling to analyse the relationship between firm characteristics and volatility dynamics. I introduce a Hierarchical Factor GARCH model in which firm variables determine factor exposure, idiosyncratic volatility, volatility persistence and smoothness of the volatility path. The proposed methodology is used to analyse the dynamics of top U.S. financial institutions during the 2007-2009 crisis using a financial index as common factor. Dynamics are a function of firm size, leverage, distance to default and liquidity before the beginning of the credit crunch. Of particular interest in this context is the factor loading coefficient capturing the sensitivity to systematic shocks. Results show that leverage is the most influential variable in the crisis, and firms with high leverage have high factor exposure, high idiosyncratic volatility as well as high sensitivity to temporary idiosyncratic volatility shocks. Factor exposure in the crisis is also high for firms that have a small distance to default, are illiquid and, to a minor extent, are large. The comparison with pre–crisis estimation results shows that role leverage is roughly stable but the impact on factor exposure of the other effects are harnessing features of the crisis period only. Overall, the model captures a substantial amount of cross sectional variation in volatility dynamics and gives economic insights on the nature of volatility in the crisis.

[12] One could also think of multiplying the factor loading by size. The problem with this approach is that the degree of concentration in size is such that the rankings one would get end up reflecting essentially rankings in this variable.

References

Acharya, V., Brownlees, C. T., Engle, R., Farazmand, F. and Richardson, M. (2010), Measuring systemic risk, *in* 'Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance', Wiley.

Acharya, V., Pedersen, L., Philippe, T. and Richardson, M. (2010), Measuring systemic risk, Technical report, Department of Finance, NYU.

Adrian, T. and Brunnermeier, M. K. (2009), CoVaR, Technical report, FRB of New York. Staff Report No. 348.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005), 'How often to sample a continuous–time process in the presence of market microstructure noise', *The Review of Financial Studies* **28**, 351–416.

Altman, E. I. (1968), 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *The Journal of Finance* **23**, 589–609.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006), Volatility and correlation forecasting, *in* G. Elliott, C. W. J. Granger and A. Timmermann, eds, 'Handbook of Economic Forecasting', North Holland.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Arellano, M. and Honoré, B. (2001), Panel data models: Some recent developments, *in* J. J. Heckman and E. Leamer, eds, 'Handbook of Econometrics', Elsevier.

Bai, X., Russell, J. R. and Tiao, G. C. (2003), 'Kurtosis of garch and stochastic volatility models with non-normal innovations', *Journal of Econometrics* **114**, 349 – 360.

Bandi, F. M. and Russell, J. R. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics* **79**, 655–692.

Barigozzi, M., Brownlees, C. T., Gallo, G. M. and Veredas, D. (2010), Disentangling systematic and idiosyncratic risk for large panels of assets, Technical report.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008), 'Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise', *Econometrica* **76**, 1481–1536.

Basu, R. and DasGupta, A. (1996), 'The mean, median, and mode of unimodal distributions : A characterization', *Theory of Probability and Its Applications* **2**, 336–352.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006), 'Multivariate GARCH models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Bauwens, L. and Rombouts, J. V. K. (2007), 'Bayesian clustering of many garch models', *Econometric Reviews* **26**, 365–386.

Beltratti, A. and Stultz, R. (2010), The Credit Crisis Around the Globe: Why Did Some Banks Perform Better During the Credit Crisis?, Technical Report WP 2010-05, Fisher College of Business Working Paper Series.

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994), ARCH models, *in* R. F. Engle and D. McFadden, eds, 'Handbook of Econometrics', Elsevier, pp. 2959–3038.

Brownlees, C. T. and Engle, R. (2010), Volatility, correlation and tails for systemic risk measurement, Technical report, Department of Finance, NYU.

Brownlees, C. T., Engle, R. and Kelly, B. T. (2010), A practical guide to volatility forecasting through calm and storm, Technical report, Department of Finance, NYU.

Brownlees, C. T. and Gallo, G. M. (2008), 'On variable selection for volatility forecasting: The role of focused selection criteria', *Journal of Financial Econometrics* **6**, 513 – 539.

Brownlees, C. T. and Gallo, G. M. (2010), 'Comparison of volatility measures: a risk management perspective', *Journal of Financial Econometrics* **8**, 2956.

Bungartz, H.-J. and Griebel, M. (2004), 'Sparse grids', *Acta Numerica* **13**, 147–269.

Calzolari, G., Fiorentini, G. and Sentana, E. (2008), 'Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks', *Journal of Econometrics* **146**, 10–25.

Chen, Ghysels and Wang (2010), The hybrid garch class of models, Technical report, University of North Carolina, Chapel Hill.

Corsi, F. (2010), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 174–196.

Davidian, M. and Gallant, A. R. (1993), 'The nonlinear mixed effects model with a smooth random effect density', *Biometrika* **80**, 475–488.

Davis, P. J. and Rabinowitz, P. (1984), *Methods of Numerical Integration*, Academic Press, New York.

Deo, R., Hurvich, C. and Lu, Y. (2006), 'Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment', *Journal of Econometrics* **131**, 29–58.

Diebold, F. X. and Nerlove, M. (1989), 'The dynamics of exchange rate volatility: A multivariate latent factor arch model', *Journal of Applied Econometrics* **4**, 1–21.

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance* **1**, 83–106.

Distaso, W., Fernandes, M. and Zikes, F. (2009), Tailing tail risk in the hedge fund industry, Technical report.

Engle, R. (2009), High dimensional dynamic correlations, *in* J. L. Castle and N. Sheppard, eds, 'The methodology and Practice of Econometrics: Papers in Honour of David F. Hendry', Oxford University Press.

Engle, R. F. and Gallo, G. M. (2006), 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics* **131**, 3–27.

Engle, R., Granger, C. W. J. and Kraft, D. (1984), 'Combining competing forecasts of inflation using a bivariate ARCH model', *Journal of Economic Dynamics and Control* **8**, 151 – 165.

Engle, R. and Lee, G. G. (1999), A Permanent and Transitory Component Model of Stock Return Volatility, *in* R. Engle and H. White, eds, 'Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger', Oxford.

Engle, R., Ng, V. K. and Rothschild, M. (1990), 'Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills', *Journal of Econometrics* **45**, 213–237.

Engle, R., Shephard, N. and Sheppard, K. (2008), Fitting vast dimensional time-varying covariance models, Technical report.

Fermanian, J.-D. and Salanié, B. (2004), 'A nonparametric simulated maximum likelihood estimation method', *Econometric Theory* **20**, 701–734.

Gallant, A. R. and Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657–681.

Ghysels, E., Harvey, A. C. and Renault, E. (1996), Stochastic volatility, *in* Maddala, G.S. and Rao, C.R., ed., 'Statistical Methods in Finance', Handbook of Statistics, Elsevier, pp. 119–191.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006), 'Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies', *Journal of Econometrics* **131**, 59–95.

Giot, P. and Laurent, S. (2003), 'Modelling daily value–at–risk using realized volatility and ARCH type models', *Journal of Empirical Finance* **11**, 379–398.

Glosten, L. R., Jagannanthan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48**, 1779–1801.

Gourieroux, C., Monfort, A. and Renault, E. (1993), 'Indirect inference', *Journal of Applied Econometrics* **8**, 85–118.

Green, W. (2001), Fixed and random effects in nonlinear models, Technical report, Department of Economics, Stern School of Business.

Hajivassiliou, V. A. and Ruud, P. A. (1994), Classical estimation methods for LDV models using simulation, *in* R. F. Engle and D. McFadden, eds, 'Handbook of Econometrics', Elsevier, pp. 2384–2441.

Hansen, P. R., Huang, Z. and Shek, H. H. (2010), Realized garch: A complete model of returns and realized measures of volatility, Technical report.

Hautsch, N. (2008), 'Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model', *Journal of Economic Dynamics and Control* **32**, 3978 – 4015.

Heiss, F. and Winschel, V. (2008), 'Likelihood approximation by numerical integration on sparse grids', *Journal of Econometrics* **144**, 62–80.

King, M., Sentana, E. and Wadhwani, S. (1994), 'Volatility and links between national stock markets', *Econometrica* **62**, 901–33.

Knight, K. and Fu, W. (2000), 'Asymptotics for lasso–type estimators', *The Annals of Statistics* **28**, 1356–1378.

Kotz, S. and Nadarajah, S. (2004), *Multivariate T-distributions & their applications*, Cambridge University Press.

Kristensen, D. and Shin, Y. (2008), Estimation of dynamic models with nonparametric simulated maximum likelihood, CREATES Research Papers 2008-58, School of Economics and Management, University of Aarhus.

Laeven, L. and Levin, R. (2009), 'Bank Governance, regulation and risk taking', *Journal of Financial Economics* **93**, 259–275.

Leeb, H. and Pötscher, B. M. (2005), 'Model selection and inference: Facts and fiction', *Econometric Theory* **21**, 21–59.

Leeb, H. and Pötscher, B. M. (2006), 'Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results', *Econometric Theory* **22**, 69–97.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. McFadden, eds, 'Handbook of Econometrics', Elsevier, pp. 2111–2245.

Nielsen, S. F. (2000), 'On simulated EM algorithms', *Journal of Econometrics* **96**, 267–292.

Pakel, C., Shephard, N. and Sheppard, K. (2010), 'Nuisance parameters, composite likelihoods and a panel of GARCH models', *Statistica Sinica* . forthcoming.

Patton, A. (2006), 'Modelling asymmetric exchange rate dependence', *International Economic Review* **47**, 527–556.

Poon, S. and Granger, C. W. J. (2003), 'Forecasting volatility in financial markets: A review', *Journal of Economic Literature* **91**, 478–539.

Poon, S. and Granger, C. W. J. (2005), 'Practical issues in forecasting volatility', *Financial Analysts Journal* **61**, 45–56.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2005), 'Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects', *Journal of Econometrics* **128**, 301–323.

Rabemananjara, R. and Zakoian, J. M. (1993), 'Threshold arch models and asymmetries in volatility', *Journal of Applied Econometrics* **8**, 31–49.

Robinson, G. K. (1991), 'That blup is a good thing: The estimation of random effects', *Statistical Science* **6**, 15–32.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

Shephard, N. and Sheppard, K. (2010), 'Realising the future: forecasting with high frequency based volatility (heavy) models', *Journal of Applied Econometrics* **25**, 197–231.

Silvennoinen, A. and Teräsvirta, T. (2009), Multivariate GARCH Models, *in* T. Mikosch, J.-P. Kreiß, R. A. Davis and T. G. Andersen, eds, 'Handbook of Financial Time Series', Springer Berlin Heidelberg, pp. 201–229.

Smolyak, S. A. (1963), 'Quadrature and interpolation formulas for tensor products of certain classes of functions', *Soviet Mathematics* **4**, 240 – 243.

Taleb, N. N. (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House, New York.

Tauchen, G. and Hussey, R. (1991), 'Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models', *Econometrica* **59**, 371–96.

Verde, M., Mancuso, P. and Altman, E. (2005), High yield bonds: Default and loss rate comparison - mid–cap versus large–cap issuers, Technical report, FitchRatings.

Winschel, V. and Krätzig, M. (2010), 'Solving, estimating, and selecting nonlinear dynamic models without the curse of dimensionality', *Econometrica* **78**, 803–821.

Proposition 1

$$
\begin{aligned}
\tilde{q}_i(\theta) - q_i(\theta) &= \log \sum_{z \in \mathcal{Z}_l} \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; w_l(z) + \log f(r_i|x_i,z^*,r_F;\theta) - \log \int f(r_i|x_i,z,r_F;\theta)\; \phi(z) \\
&= \log \sum_{z \in \mathcal{Z}_l} \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; w_l(z) + \log f(r_i|x_i,z^*,r_F;\theta) \\
&\quad - \log \int \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; \phi(z) - \log f(r_i|x_i,z^*,r_F;\theta) \\
&= \log \sum_{z \in \mathcal{Z}_l} \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; w_l(z) - \log \int \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; \phi(z) \\
&= \left(A - \int \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; \phi(z) \right) \left(\int \frac{f(r_i|x_i,z,r_F;\theta)}{f(r_i|x_i,z^*,r_F;\theta)}\; \phi(z) \right)^{-1} \\
&\leq \mathcal{E}_l(f) \left[\frac{f(r_i|x_i,z^*,r_F;\theta)}{\int f(r_i|x_i,z,r_F;\theta)\; \phi(z)} \right]
\end{aligned}
$$

Proposition 2 *Let the function h_i be defined as*

$$
h_i = \int \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\; \exp\left[\log f(r_i|x_i,z,r_F;\theta) - \tilde{q}_i(\theta)\right]\; \phi(z)
$$

and consider the factorization of the derivative log–likelihood approximation error as

$$
\begin{aligned}
\nabla_\theta \tilde{q}_i(\theta) - \nabla_\theta q_i(\theta) &= (\nabla_\theta \tilde{q}_i - h_i) + (h_i - \nabla_\theta q_i) \\
&= A_i + B_i
\end{aligned}
$$

where A_i is the approximation error in the computation of the derivative log–likelihood assuming that keeping the value of the likelihood fixed at $f^{\,l}$, and B_i is the contribution of the total approximation error due approximation of $f^{\,l}$, which can be conviently be expressed as

$$
\begin{aligned}
A_i &= \sum_{z \in Z_l} \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\;\; f(r_i|x_i,z,r_F;\theta) \frac{f(r_i|x_i,z,r_F;\theta)}{\exp \tilde{q}_i(\theta)} w_l(z) \\
&\quad - \int \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\; f(r_i|x_i,z,r_F;\theta) \frac{f(r_i|x_i,z,r_F;\theta)}{\exp \tilde{q}_i(\theta)} \phi(z) \\
&= \frac{f_i}{\exp \tilde{q}_i(\theta)} \Bigg(\sum_{z \in Z_l} \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\;\; f(r_i|x_i,z,r_F;\theta) \frac{f(r_i|x_i,z,r_F;\theta)}{f_i(\theta)} w_l(z) \\
&\quad - \int \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\; f(r_i|x_i,z,r_F;\theta) \frac{f(r_i|x_i,z,r_F;\theta)}{f_i(\theta)} \phi(z) \Bigg) \\
&\leq \exp(q_i - \tilde{q}_i)\; \mathcal{E}_l(\nabla q_i)
\end{aligned}
$$

$$
\begin{aligned}
B_i &= \int \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\; f(r_i|x_i,z,r_F;\theta) \left(\, \exp(-\tilde{q}_i) - \exp(-q_i) \,\right)\; \phi(z) \\
&= \left(\int \nabla_\theta \log f(r_i|x_i,z,r_F;\theta)\; \frac{f(r_i|x_i,z,r_F;\theta)}{f_i} \phi(z) \right) (q_i - \check{q}_i) \\
&\leq \nabla_\theta q_i\; \mathcal{E}_l(q_i)
\end{aligned}
$$

where the second expression is obtained using the mean value theorem and $\check{f}_i$ lies in segment joining f_i and $\tilde{f}_i^{\,l}$.

Proposition 3

	$Q_{0.10}$	Mean	Std. Dev.	$Q_{0.90}$	ρ_{2004}	Correlation siz	lvg	dtd	liq
siz	6406	31643	44450	72628	0.99				
lvg	1.34	5.53	4.43	10.64	0.87	0.26			
dtd	10.42	49.43	60.53	80.41	0.96	-0.06	-0.09		
liq	0.02	0.14	0.17	0.3	0.84	-0.02	-0.25	-0.06	

Table 3: Asset characteristics descriptive statistics.

Coefficient	$Q_{0.05}$	Median	$Q_{0.95}$	R^2
β	0.61	1.00	1.39	35.4%
$h_{\mathsf{ann}}^{1/2}$	37.78	54.07	77.39	10.9%
π	0.96	0.98	0.99	–
λ	0.10	0.17	0.28	16.4%

Table 4: Coefficient quantiles for an average firm and goodness of fit.

Coefficient	$Q_{0.05}$	Median	$Q_{0.95}$	R^2
β	0.63	1.00	1.36	13.7%
$h_{\mathsf{ann}}^{1/2}$	43.86	56.19	71.98	24.9%
π	0.93	0.98	1.00	–
λ	0.04	0.10	0.20	10.2%

Table 5: Coefficient quantiles for an average firm and goodness of fit.