Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2010/06

Disentangling Systematic and Idiosyncratic Risk for Large Panels of Assets

M. Barigozzi, C. T. Brownlees,
G. M. Gallo, D. Veredas

Università degli Studi
di Firenze

DISENTANGLING SYSTEMATIC AND IDIOSYNCRATIC RISK FOR LARGE PANELS OF ASSETS

Matteo BARIGOZZI[1], Christian T. BROWNLEES[2], Giampiero M. GALLO[3] and David VEREDAS[4]

Abstract

When observed over a large panel, measures of risk (such as realized volatilities) usually exhibit a secular trend around which individual risks cluster. In this article we propose a vector Multiplicative Error Model achieving a decomposition of each risk measure into a common systematic and an idiosyncratic component, while allowing for contemporaneous dependence in the innovation process. As a consequence, we can assess how much of the current asset risk is due to a system wide component, and measure the persistence of the deviation of an asset specific risk from that common level. We develop an estimation technique, based on a combination of seminonparametric methods and copula theory, that is suitable for large dimensional panels. The model is applied to two panels of daily realized volatilities between 2001 and 2008: the SPDR Sectoral Indices of the S&P500 and the constituents of the S&P100. Similar results are obtained on the two sets in terms of reverting behavior of the common nonstationary component and the idiosyncratic dynamics to with a variable speed that appears to be sector dependent.

Keywords: Systematic risk, idiosyncratic risk, Multiplicative Error Model, seminonparametric, copula.

JEL classification: C32, C51, G01.

[1]ECARES, Solvay Brussels School of Economics and Management, Université libre de Bruxelles; email: matteo.barigozzi@ulb.ac.be.

[2]New York University - Department of Finance; email: ctb@stern.nyu.edu.

[3]University of Florence - Dipartimento di Statistica; email: gallog@ds.unifi.it.

[4]ECARES, Solvay Brussels School of Economics and Management, Université libre de Bruxelles; email: david.veredas@ulb.ac.be.

Corresponding address: David Veredas, ECARES, Université libre de Bruxelles, 50 Av F.D. Roosevelt CP114, B1050 Brussels, Belgium. Phone: +3226504218. Fax: +3226504475.

We are grateful to Yacine Aït-Sahalia, Robert F. Engle, Jianqing Fan, Marc Hallin, Eric Jondeau, Oliver Linton, Alexei Onatsky, Davy Paindaveine, Eric Renault, Valeri Voev and Liuren Wu for insightful remarks. We also acknowledge the comments of the participants of the 4th Brussels-Waseda Seminar in Time Series and Financial Statistics (June 2009), the Stanford SITE summer 2009 workshop (June 2009), the 3rd International Conference on Computational and Financial Econometrics (October, 2009), the 2009 EC^2 conference in Aarhus (December 2009), the Oberwolfach Workshop of Semiparametric Modeling of Multivariate Time Series with Changing Dynamics (January, 2010), the NYU-Stern conference on Volatility and Systemic Risk (April 2010), the EUI Econometrics Workshop on Recent Advances in Time Series Econometrics (June 2010), CEQURA Conference on Advances in Financial and Insurance Risk Management (September 2010), and the seminar participants at Carlos III, CORE, ECARES, LSE, Princeton, Queen Mary College and Rutgers. Matteo Barigozzi gratefully acknowledges financial support from Wallonie-Bruxelles International (WBI). David Veredas acknowledges financial support from the Belgian National Bank and the IAP P6/07 contract, from the IAP program (Belgian Scientific Policy), 'Economic policy and finance in the global economy'. David Veredas is member of ECORE, the recently created association between CORE and ECARES. Any error and inaccuracy are ours.

1 Introduction

We are interested in the analysis of risk measures for large panels of assets and we start by adopting a pragmatic approach. By superimposing realized volatilities (a specific measure of risk) for many assets in the same market we find the empirical regularity that they exhibit a striking clustering around a common nonstationary trend. In this article we propose a model that disentangles this common trend from asset specific dynamics that oscillate around it. We refer to the former as *systematic* and to the latter as *idiosyncratic*. Identifying these two components allows us to know how much of the risk of an asset at any given date is due to a system wide level, and, in a dynamic perspective, how persistent idiosyncratic deviations are. The methodology is quite general and can be applied to other risk measures (Value-at-Risk, expected shortfall, etc.), or, for that matter, liquidity measures as well (traded volumes, spreads, trading intensities, etc.). To be clear, we are not attempting the identification of two separate sources of shocks (common and idiosyncratic) which could be both short–lived (an example of the former would be the sudden market drop of Thu. May 8, 2010), but rather two different types of dynamics, reckoning that the average level of volatility in a market is common to all assets considered, time varying, and slow moving.

We propose a novel specification that builds up on the literature on Multiplicative Error Models (MEM) (Engle (2002), Engle and Gallo (2006)) and dynamic models with slowly moving components as in, *inter alia*, Engle and Rangel (2008). We introduce a component vector MEM that decomposes the conditional expectation of each risk measure in a panel as the product of a systematic trend and an idiosyncratic dynamic component, while allowing for contemporaneous dependence in risk innovations. The systematic risk is modeled as a nonparametric curve while the dynamics of the idiosyncratic deviations are parametric. A simple estimation approach makes the model appealing especially when the number of assets in the panel is large. Indeed, estimation of the systematic and idiosyncratic risks boils down to a series of univariate maximum likelihood problems, and the dependence among risk innovations is estimated with a sample correlation.

Our model combines seminonparametric methods and copulas. The systematic risk, being a nonparametric curve, is estimated with kernels, while the idiosyncratic risks and the dependencies among risk innovations are fully parametric. We rely on profile likelihood (Staniswalis (1987) and Staniswalis (1989), Severini and Wong (1992) and Veredas *et al.* (2007)). This technique is based on the fact that the parametric Fisher information matrix in a seminonparametric model is smaller or equal than that in a fully parametric model. Profile likelihood finds the nonparametric curve that leads to the largest parametric Fisher information matrix. Such a curve is called the least favorable and is obtained by projections of the parametric and nonparametric scores. To compute the dependencies among risk innovations, we rely on copulas and inference from the marginals of Joe (1997) and Joe (2005), which is based on a two–step approach. The joint density has two parts, one due to the marginal conditional densities and another to the copula function. We first estimate the secular systematic and parametric idiosyncratic risks maximizing the marginal part of the log-likelihood. In the second step we estimate the dependencies among risk innovations maximizing the (Gaussian) copula part of the log-likelihood. Such a procedure brings down estimation to a simple problem: the estimated systematic

component has a closed-form expression, and estimation of the parametric idiosyncratic components and the dependencies among risk innovations boils down to the above commented univariate maximum likelihood problems and a sample correlation matrix. We show the asymptotic properties of the estimators. In particular, since the systematic component is a least favorable curve, the parametric Fisher information matrix is the largest given the seminonparametric and two–steps nature of the estimation.

We apply the model to two panels of daily realized volatility measures (Andersen *et al.* (2003), Aït-Sahalia *et al.* (2005), Bandi and Russell (2006), Barndorff-Nielsen *et al.* (2008)) spanning from January 2, 2001 to December 31, 2008. The first panel consists of the nine sectoral indices of the SPDR S&P500 index while the second contains the ninety constituents of the S&P100 that have continuously been trading in the sample period. The datasets are related to each other in that the constituents of the S&P100 are also some of the main underlying assets of the SPDR sectoral indices; reassuringly, the empirical results of the two applications are consistent to one another in what concerns the estimated shape of the systematic risk, showing that it is essentially the same in the two panels, and its level can be associated with the global level of uncertainty in the economy: it reaches its peaks at the beginning and end of the 2000s in correspondence to the dot com bubble burst and the financial crisis, and its trough at the end of 2006. While all realized volatility series exhibit nonstationary, once a systematic risk is accounted for all idiosyncratic dynamics are reverting to the systematic component. However, the speed of reversion is rather heterogeneous across assets. Consumer, materials and healthcare assets have a steady reversion while it is much slower for the technology and energy related ones. Also, the S&P100 panel exhibits on average more idiosyncratic dynamics in the sense of slower mean reversion than the SPDR. Idiosyncratic dynamics show some interesting patterns, like technology being higher during the dot-com bubble burst, energy sectors being higher during the energy crisis in 2005-2006, and financial sector being most distressed during the credit-crunch crisis. We also find that there is a non negligible dependence in the volatility innovations with average correlations of 0.40 and 0.26 in the SDPR and S&P100 panels respectively. The S&P100 innovations also have some clear sectoral correlation patterns within the technology, financial, energy and utilities assets.

Different strands of literature relate to our work. Starting from the contribution of Engle and Rangel (2008), there has been interest in capturing secular trends in financial volatility. Among others, the list of contributions in a univariate setting includes Amado and Teräsvirta (2008), Engle *et al.* (2008) and Brownlees and Gallo (2010). Feng (2006), Rangel and Engle (2009), Hafner and Linton (2009), Long *et al.* (2009) and Colacito *et al.* (2010) extend these ideas in a multivariate setting. The paper relates also to the literature on multivariate extensions of the MEM model, like the works of Cipollini *et al.* (2007) and Hautsch (2008), and the analysis of large panels with factor models, either linear and homecesdatic (Chamberlain and Rothschild (1983), Forni *et al.* (2000), Forni *et al.* (2004), Forni *et al.* (2005), Bai and Ng (2002)) or heteroskedastic (**?**, **?**), and nonlinear (Gagliardini and Gourieroux (2009)). This work also fits with the larger segment of the literature that finds evidence of long range dependence in volatility and have proposed ways to capture it. Significant contributions include long memory models (Andersen *et al.* (2003), Deo *et al.* (2006)) and the pseudo long memory HAR (Corsi (2010), Andersen *et al.* (2007)). There are also connections with the growing literature on modeling daily

volatility using intra-daily information. Research in this area includes Andersen *et al.* (2007), Patton and Sheppard (2009), Shephard and Sheppard (2010), Hansen *et al.* (2010) and Chen *et al.* (2010). Chiriac and Voev (2010) explore models for realized covariance matrices. Our approach lies somehow in the middle between univariate and multivariate realized volatility models in that we analyze panels of volatilities but we do not model covariances.

The paper is structured as follows. Section 2 describes the panels of realized volatility measures that we use in the empirical application and reports some descriptive statistics that motivate the modeling approach. Section 3 describes the model and Section 4 details the estimation strategy and the asymptotic properties of the estimator. Section 5 presents the results for the SPDR sectoral indices and the constituents of the S&P100. We conclude in Section 6. The Appendix gathers the assumptions and the proofs of the propositions and theorems.

2 Stylized facts of large panels of volatility

We study two panels of realized volatility measures from January 2, 2001 to December 31, 2008. The first, referred to as SPDR, consists of the nine Select Sector SPDRs Exchange Traded Funds (ETF) that divide the S&P500 index into sector index funds. The sectors (with the abbreviation we use and the more cryptic original ticker) are Materials (Mat, XLB), Energy (Ener, XLE), Financial (Fin, XLF), Industrial (Ind, XLI), Technology (Tech, XLK), Consumer Staples (Stap, XLP), Utilities (Util, XLU), Health Care (Heal, XLV), and Consumer Discretionary (Disc, XLY). The second, named S&P100, consists of U.S. equity companies that are part of the S&P100 index. This panel contains all the constituents of the S&P100 index as of December 2008 that have been trading in the full sample period (90 in total). The complete list of S&P100 tickers, company names and industry sectors is reported in Table 1 of the Appendix.

Among the available estimator of the daily integrated volatility based on intraday returns, we adopt the realized kernels (Barndorff-Nielsen *et al.* (2008)).[1] They are a family of heteroskedastic and autocorrelation consistent (HAC) type estimators, robust to various forms of market microstructure noise present in high frequency data. We support our choice with theoretical results and good forecasting properties in empirical studies (e.g. for predicting Value at Risk, Brownlees and Gallo (2010)). Yet, we trust that our results do not depend on the specific measure of volatility chosen.

We compute optimal realized kernels following the procedure detailed in Barndorff-Nielsen *et al.* (2009). Our primary source of data are tick–by–tick intra-daily quotes from the TAQ database. Data are extracted and filtered using the methods described in Brownlees and Gallo (2006) and Barndorff-Nielsen *et al.* (2009). Let r_{itj} denote the 1-minute frequency returns (sampled in tick time) at minute j on day t for ticker i. The

[1] Other alternative estimators are the range (Parkinson (1980), Alizadeh *et al.* (2002)), the "vanilla" 5–minute realized volatility (Andersen *et al.* (2003)), or the two-scales estimator (Aït-Sahalia *et al.* (2005)).

realized (Parzen) kernel estimator is defined as

$$x_{i\,t} = \sum_{h=-H}^{H} \mathrm{K}_p\left(\frac{h}{H+1}\right)\gamma_h,$$

with

$$\gamma_h = \sum_{j=|h|+1}^{J} r_{i\,t\,j} r_{i\,t\,j-|h|},$$

and where H is both the bandwidth of the kernel and the maximum order of the autocovariance (the optimal selection for each day and ticker in the sample is slightly involved and closely follows the guidelines Barndorff-Nielsen *et al.* (2009)), J is the number of 1–minute frequency returns within the day, and $\mathrm{K}_p(\cdot)$ denotes the Parzen kernel

$$\mathrm{K}_p(y) = \begin{cases} 1 - 6y^2 + 6y^3 & 0 \leq y \leq 1/2 \\ 2(1-y)^3 & 1/2 \leq y \leq 1 \\ 0 & y > 1. \end{cases}$$

Under appropriate conditions, Barndorff-Nielsen *et al.* (2008) show that the realized kernel estimator converges to the integrated variance of returns.

[FIGURE 1 ABOUT HERE]

Figure 1 shows plots of the two panels of annualized volatilities $\sqrt{252 x_{i\,t}}$; top for SPDR and bottom for S&P100.[2] For both plots the volatilities co–move and they are barely distinguishable from each other. The overall pattern suggests that they all cluster around a common time-varying level, which qualifies as the *systematic volatility*. The secular movements of the systematic volatility can be easily attached to well known economic events or system wide innovations. The downturn in volatility in 2001 corresponds to the aftermath of the burst of the dot com bubble. The rise around 2002 and 2003 due to the accounting scandals (Enron and Worldcom among others). Volatility then drops from 2004 to July 2007 when it starts to rise with the beginning of the financial crisis. It then skyrockets to the highest level of volatility in the last 20 years in the fall of 2008, following the demise of Lehman Brothers.

Table 2 displays summary descriptive statistics. The table reports average annualized volatility, standard deviation of volatility (a.k.a. volatility of volatility), daily, weekly (5 days) and monthly (22 days) autocorrelations, average correlation with the other series and the percentage of variance explained by the first principal component. For SPDR we report statistics for each sectoral index while for S&P100 we report the values across the same sectors as SPDR and, for each sector, we report the 25, 50, and 75 quantile. Mean and variability levels of the series are higher for the S&P100 panel rather then the SPDR, due to the fact the sectoral aggregation decreases the average and dispersion of volatility.

[2]For mere scale reasons we exclude from the figure the energy stock Williams Companies, which had very large volatilities between May and July 2002. The series itself is considered in the estimation since, regardless of the 2002 turmoil, it was part of the S&P100 over the sample period.

Autocorrelations decay slowly, consistently with the evidence of long range dependence widely documented in volatility studies. The proportion of variance explained by the first principal components is always above 50%, which confirms the existence of strong co-movements in volatility. But there are substantial differences across sectors and stocks, which qualify as *idiosyncrasies*, that can be understood as firm (sector) specific movements that oscillate around the systematic risk. This is particularly clear for S&P100, for which the dissimilarity among volatilities is the largest.

[TABLE 2 ABOUT HERE]

In all, we understand volatility for large panels of assets as a combination of systematic (common secular trend) and idiosyncratic (firm or sector specific movements) components. In the next section we propose a model that disentangles them.

3 Disentangling: Theory

Let $\mathbf{x}_t = (x_{1\,t} \dots x_{N\,t})'$ be a vector of N non–negative real–valued risk measures at time $t \in \mathbb{N}$. The vector Multiplicative Error Model in the current context is defined as

$$\mathbf{x}_t = \phi(z_t)\ \mathbf{a} \odot \boldsymbol{\mu}(\mathcal{F}_{t-1}, \boldsymbol{\delta}) \odot \boldsymbol{\epsilon}_t \tag{1}$$

and describes the behavior of $\mathbf{x}_t$ as the product of the systematic slow moving risk component $\phi(z_t)$ depending on a time index z_t, the asset specific overall average risk levels $\mathbf{a}$, the idiosyncratic risks $\boldsymbol{\mu}(\mathcal{F}_{t-1}, \boldsymbol{\delta})$ capturing short-lived dynamics around $\phi(z_t)$, and a mean one innovation term $\boldsymbol{\epsilon}_t$.

The idiosyncratic risks $\boldsymbol{\mu}(\mathcal{F}_{t-1}, \boldsymbol{\delta}) \subset \mathbb{R}_+^N$ form a $N \times 1$ vector:

$$\boldsymbol{\mu}(\mathcal{F}_{t-1}, \boldsymbol{\delta}) = \begin{pmatrix} \mu_1(\mathcal{F}_{t-1}, \boldsymbol{\delta}_1) \\ \mu_2(\mathcal{F}_{t-1}, \boldsymbol{\delta}_2) \\ \vdots \\ \mu_N(\mathcal{F}_{t-1}, \boldsymbol{\delta}_N) \end{pmatrix},$$

where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_N) \in \mathcal{D} \subset \mathbb{R}^{p_\delta}$ with $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_N$. Let $\mu_{i\,t}$ denote $\mu_i(\mathcal{F}_{t-1}, \boldsymbol{\delta}_i)$. Similarly to Engle and Rangel (2008), the i-th component is

$$\mu_{i\,t} = \left(1 - \alpha_i - \beta_i - \frac{\gamma_i}{2}\right) + \alpha_i \frac{x_{i\,t-1}}{a_i \phi(z_{t-1})} + \beta_i \mu_{i\,t-1} + \gamma_i \frac{x_{i\,t-1}}{a_i \phi(z_{t-1})} \mathbf{1}_{r_{i\,t-1}<0}, \tag{2}$$

so that $\boldsymbol{\delta}_i = (\alpha_i, \beta_i, \gamma_i)$. This is a dynamic process with unit unconditional expected value under the assumption of equiprobability of positive or negative returns.[3]

[3] In the case of an asymmetric distribution of returns it is enough to replace the first term on the right-hand side of (2) by $(1 - \alpha_i - \beta_i - \gamma_i \mathrm{E}[r_{i\,t} < 0])$.

The N-dimensional vector $\mathbf{a} = (a_1, \ldots, a_N) \in \mathcal{A} \subset \mathbb{R}_+^N$ is the average level of the risk measures. The intercept term in (2) ensures identification of $\boldsymbol{\delta}_i$ and a_i.[4]

The systematic risk is captured by the scalar smooth function $\phi(z_t) : [0,1] \to \mathcal{P} \subset \mathbb{R}_+$ taking values on the set $\Gamma = \{p \in C^2[0,1] : p(z_t) \in int(\mathcal{P}) \text{ for all } z_t \in [0,1]\}$. The variable z_t is the driver of the systematic risk. In our application $z_t = t/T$, where T is the total number of observations, but z_t can be any other standardized variable that is believed to drive $\phi(z_t)$. Since $\mathrm{E}[\epsilon_{it}] = 1$, the unconditional expectation is $\mathrm{E}[x_{it}|\mathcal{F}_{t-1}] = a_i\phi(z_t)\mu_{it}$ and $\mathrm{E}[x_{it}] = a_i\phi(z_t)$, which shows that the model is non–stationary, recognizing the fact that risks may not be stationary.

The innovation $\boldsymbol{\epsilon}_t$ is a $N \times 1$ vector of $i.i.d.$ errors with positive support, mean one and distributed according to a probability density function (pdf) $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t; \boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{p_\theta}$, and corresponding cumulative density function (cdf) $F(\boldsymbol{\epsilon}_t; \boldsymbol{\theta})$. Denote by $F_{\mathbf{x}}(\mathbf{x}_t|\mathcal{F}_{t-1}; \boldsymbol{\eta}, \phi(z_t))$ the joint cdf of $\mathbf{x}_t$ entailed by (1) and $F_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t; \boldsymbol{\theta})$. Finally, let $\boldsymbol{\eta} = (\mathbf{a}, \boldsymbol{\delta}, \boldsymbol{\theta}) \in \boldsymbol{\Lambda} \subset \mathbb{R}^{N+p_\delta+p_\theta}$, $\boldsymbol{\Lambda} = \mathcal{A} \times \mathcal{D} \times \boldsymbol{\Theta}$ being a compact set, be the vector that gathers all the parameters.

We assume that we do not know $F_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t; \boldsymbol{\theta})$ but we have knowledge about the marginal cdf's $F_{\epsilon_i}(\epsilon_{it}; \boldsymbol{\theta}_i)$, $i = 1, \ldots, N$ and hence about the marginal conditional cdf's $F_{x_i}(x_{it}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi(z_t))$, with $x_{it} = \phi(z_t)a_i\mu_i(\mathcal{F}_{t-1}, \boldsymbol{\delta}_i)\epsilon_{it}$, and $\boldsymbol{\xi}_i = (a_i, \boldsymbol{\delta}_i, \boldsymbol{\theta}_i)$ a $p_{\xi_i} \times 1$ vector containing the parameters of the i-th conditional distribution. Lack of information on the joint distribution for $\mathbf{x}_t$ is realistic. There are barely multivariate distributions for positive real–valued random vectors (see Johnson *et al.* (2000)). The multivariate exponential and gamma are the two most prominent but they are cumbersome and their properties may not always dovetail with those of the risk measures (see Cipollini *et al.* (2007)). We gather the parameters of all the marginals in $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N) \in \boldsymbol{\Xi} \subset \mathbb{R}^{p_{\xi_1}+\ldots+p_{\xi_N}}$, where $\boldsymbol{\Xi} = \boldsymbol{\Xi}_1 \times \ldots \times \boldsymbol{\Xi}_N$.

The dependencies among risk innovations are explained by a copula function that depends on a set of parameters $\boldsymbol{\psi} \in \boldsymbol{\Psi} \subset \mathbb{R}^{p_\psi}$ specifying its shape. The parameter set $\boldsymbol{\eta}$ containing all the parameters can therefore be split as $\boldsymbol{\eta} = (\boldsymbol{\xi}, \boldsymbol{\psi}) \in \boldsymbol{\Lambda} \subset \mathbb{R}^{p_{\xi_1}+\ldots+p_{\xi_N}+p_\psi}$, where $\boldsymbol{\Lambda} \equiv \boldsymbol{\Xi} \times \boldsymbol{\Psi}$.[5] By Sklar's theorem the joint conditional cdf of $\mathbf{x}_t$ is

$$F_{\mathbf{x}}(\mathbf{x}_t|\mathcal{F}_{t-1}; \boldsymbol{\eta}, \phi(z_t)) = C(F_{x_1}(x_{1,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_1, \phi(z_t)), \ldots, F_{x_N}(x_{N,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_N, \phi(z_t)); \boldsymbol{\psi}),$$

and the joint pdf is

$$\begin{aligned} f_{\mathbf{x}}(\mathbf{x}_t|\mathcal{F}_{t-1}; \boldsymbol{\eta}, \phi(z_t)) &= \prod_{i=1}^{N} f_{x_i}(x_{it}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi(z_t)) \\ &\quad \cdot c(F_{x_1}(x_{1,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_1, \phi(z_t)), \ldots, F_{x_N}(x_{N,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_N, \phi(z_t)); \boldsymbol{\psi}), \end{aligned}$$

[4] An equivalent parametrization can be obtained by defining

$$\mu'_{it} = \omega'_i + \alpha'_i \frac{x_{it-1}}{\phi(z_{t-1})} + \beta'_i \mu'_{it-1} + \gamma'_i \frac{x_{it-1}}{\phi(z_{t-1})} \mathbf{1}_{r_{it-1}<0},$$

such that $\mu'_{it} = a_i\mu_{it}$, and $\boldsymbol{\delta}'_i = (\omega'_i, \alpha'_i, \beta'_i, \gamma'_i)$ with $\omega'_i = a_i\left(1 - \alpha_i - \beta_i - \frac{\gamma_i}{2}\right)$, $\alpha'_i = \alpha_i$, $\beta'_i = \beta_i$, and $\gamma'_i = \gamma_i$. This alternative parametrization may be computationally simpler and the original parameters $\boldsymbol{\delta}_i$ and a_i can be uniquely backed up.

[5] Note that $p_{\xi_1} + \ldots + p_{\xi_N} + p_\psi = N + p_\delta + p_\theta$ and that, since the variables are not independent, $\boldsymbol{\theta} \neq (\boldsymbol{\theta}_1 \ldots \boldsymbol{\theta}_N)$ and $\boldsymbol{\eta} \neq (\boldsymbol{\xi}_1 \ldots \boldsymbol{\xi}_N)$.

where c is the derivative of C with respect to $(F_{x_1}, \dots F_{x_N})$.

We define $\mathcal{L}(\boldsymbol{\eta}, \phi(z_t)) = \sum_{t=1}^{T} \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)) = \sum_{t=1}^{T} \log f_{\mathbf{x}}(\mathbf{x}_t | \mathcal{F}_{t-1}; \boldsymbol{\eta}, \phi(z_t))$ to be the log–likelihood, which can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\eta}, \phi(z_t)) &= \left[\sum_{t=1}^{T}\sum_{i=1}^{N} \log f_{x_i}(x_{it}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi(z_t))\right] \\ &\quad + \sum_{t=1}^{T} \log c(F_{x_1}(x_{1,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_1, \phi(z_t)), \dots, F_{x_N}(x_{N,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_N, \phi(z_t)); \boldsymbol{\psi}).\end{aligned}$$

To short the notation we denote

$$\begin{aligned}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t)) &= \log f_{x_i}(x_{it}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi(z_t)) \text{ and} \\ \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t)) &= \log c(F_{x_1}(x_{1,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_1, \phi(z_t)), \dots, F_{x_N}(x_{N,t}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_N, \phi(z_t)); \boldsymbol{\psi}),\end{aligned}$$

so that

$$\mathcal{L}(\boldsymbol{\eta}, \phi(z_t)) = \left[\sum_{t=1}^{T}\sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right] + \sum_{t=1}^{T} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t)). \tag{3}$$

In order to estimate the model, three decisions have to be taken.

The first one are the marginal distributions. We opt for gamma distributions. This choice is motivated from the facts that it is a rather flexible distribution, it belongs to the exponential family, and it nests the exponential and the chi–square distributions. If ϵ_{it} follows a gamma distribution with parameters $\boldsymbol{\theta}_i = (k_i, \nu_i)$, the conditional distribution of x_{it} is gamma with parameters $((a_i\phi(z_t)\mu_{it})^{-1}k_i, \nu_i)$. The conditional pdf for x_{it} is then

$$f_{x_i}(x_{it}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi(z_t)) = \frac{k_i}{\Gamma(\nu_i)a_i\phi(z_t)\mu_{it}} \left(\frac{x_{it}k_i}{a_i\phi(z_t)\mu_{it}}\right)^{\nu_i - 1} \exp\left(-\frac{x_{it}k_i}{a_i\phi(z_t)\mu_{it}}\right).$$

Let $k_i = \nu_i$, which ensures that ϵ_{it} has unit mean (and variance ν_i^{-1}). Note that if $\nu_i = 1$ all the marginals are exponentially distributed with parameter $(a_i\phi(z_t)\mu_{it})^{-1}$. With this choice, and given (2), the parameter vector of the i-th marginal is $\boldsymbol{\xi}_i = (a_i, \alpha_i, \beta_i, \gamma_i, \nu_i)$.

The second choice is the copula function. The main empirical advantage of copulas is their ability to capture correlations and tail dependences. As far as the latter is concerned, knowledge of the copula function is crucial as it determines the behavior at extreme quantiles. Since our interest lies in correlations between risk measures, we rely on Gaussian copulas. Define $u_{it} = F_{x_i}(x_{it}|\mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi(z_t))$ for $i = 1, \dots, N$. The standard approach is to bend $u_{1\,t}, \dots, u_{N\,t}$ with a Gaussian copula, yielding a meta copula (see Song, 2000 and Cipollini, Engle and Gallo, 2009, for its use on dispersion distributions generated from a Gaussian copula). We proceed differently by applying another monotone increasing transformation to u_{it}: $y_{it} = \Phi^{-1}(u_{it})$.[6] Then we know that $y_{it} \sim \mathcal{N}(0, 1)$. By Sklar's theorem the joint cdf of $\mathbf{y}_t$ can be written as a copula function C with standard Normal

[6]Notice that $y_{i\,t}$ is a function of $\boldsymbol{\xi}_i$ and $\phi(z_t)$.

marginals and depending on some generic parameter $\mathbf{R}$:

$$G_{\mathbf{y}}(y_{1\,t},\ldots,y_{N\,t}) = C(\Phi(y_{1\,t}),\ldots,\Phi(y_{N\,t});\mathbf{R}), \tag{4}$$

with copula density

$$g_{\mathbf{y}}(y_{i\,t},\ldots,y_{N\,t}) = c(\Phi(y_{1\,t}),\ldots,\Phi(y_{N\,t});\mathbf{R})\prod_{i=1}^{N}\varphi(y_{it}),$$

and where φ is the univariate standard Gaussian pdf. We seek a function c such that i) it is a copula density, ii) if $y_{1\,t},\ldots,y_{N\,t}$ are independent then $c(\Phi(y_{1\,t}),\ldots,\Phi(y_{N\,t});\mathbf{R}) = 1$ and $g_{\mathbf{y}}(y_{1\,t},\ldots,y_{N\,t}) = \prod_{i=1}^{N}\varphi(y_{i\,t})$, and iii) it is suitable for large dimensions. We opt for the Gaussian copula density

$$c_{\Phi}(\Phi(y_{1\,t}),\ldots,\Phi(y_{N\,t});\mathbf{R}) = |\mathbf{R}|^{-1/2}\text{exp}\left(-\frac{1}{2}\mathbf{y}_t'(\mathbf{R}-\mathbf{I})\mathbf{y}_t\right) \tag{5}$$

where $\mathbf{R}$ is the Pearson correlation matrix and $\mathbf{I}$ is the identity matrix. If $(y_{1\,t},\ldots,y_{N\,t})$ are independent then $\mathbf{R} = \mathbf{I}$ and the joint density is the product of the marginals. Under this choice $g_{\mathbf{y}}$ is the pdf of a multivariate standard Gaussian and $G_{\mathbf{y}}(y_{1\,t},\ldots,y_{N\,t}) = \mathbf{\Phi}(y_{1\,t},\ldots,y_{N\,t})$ is a multivariate standard Gaussian cdf.[7]

The last choice is the estimation strategy. With respect to the fully parametric problems, model (1) entails two additional difficulties. The first is the presence of the systematic risk $\phi(z_t)$ which, statistically wise, is a nonparametric curve. We rely on profile likelihood that allows to obtain efficient estimators for $\boldsymbol{\eta}$ by considering $\phi(z_t)$ as an infinite dimensional nuisance parameter, and do correct inference for $\phi(z_t)$. We show in the first part of the next Section the asymptotic properties of the estimators for $\boldsymbol{\eta}$ and $\phi(z_t)$. The second difficulty is the computational complexity. Although theoretically feasible, joint estimation of $\boldsymbol{\eta}$ from (3) and of $\phi(z_t)$ is cumbersome in practice. We rely on the Inference from the Marginals of Joe (1997) and Joe (2005), which is based on a two–step approach. We first estimate $\boldsymbol{\xi}$ and $\phi(z_t)$ making use of the marginal part of the log-likelihood and, in a second step, we estimate $\boldsymbol{\psi}$ maximizing $\sum_{t=1}^{T}\boldsymbol{\ell}^{c}(\hat{\boldsymbol{\xi}}_T,\boldsymbol{\psi},\hat{\phi}_{T,\boldsymbol{\xi}}(z_t))$, where $\hat{\boldsymbol{\xi}}_T$ and $\hat{\phi}_{T,\boldsymbol{\xi}}(z_t)$ are the estimates of the first step. In the second part of the next Section we show how to do correct inference for the estimators of $\boldsymbol{\xi}$, $\phi(z_t)$ and $\boldsymbol{\psi}$ combining profile likelihood and Inference from the Marginals.

4 Estimation and Asymptotic Properties

Standard ML techniques do not apply directly since the estimation of the parameters $\boldsymbol{\eta} = (\boldsymbol{\xi},\boldsymbol{\psi})$ does not necessarily provide consistent estimators in the presence of the

[7] By mapping $u_{i\,t}$ into a Gaussian frame we ignore the possible tail dependencies that may be present between risk measures. This is not harmful since our aim is not to capture these dependencies. Yet we acknowledge that in some instances measuring tail dependences may be of interest, such as in financial products based on volatilities. The estimation strategy and the asymptotic theory of next section considers a generic copula.

infinite dimensional parameter $\phi(z_t)$. In order to implement valid inference not only on $\boldsymbol{\xi}$ but on $\phi(z_t)$ as well, we use profile likelihood. It is based on the fact that the marginal Fisher information of a seminonparametric model for $\boldsymbol{\eta}$ is smaller or equal than the marginal Fisher information of a parametric model. So the aim is to find the curve such that the marginal Fisher information is the largest possible. This curve is called the least favorable curve – the systematic risk – and the maximal marginal Fisher information is obtained by projections of the nonparametric and parametric scores.

We denote by $\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$ the gradient with respect to $\boldsymbol{\eta}$ and by $\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$ the derivative with respect to ϕ (the Fréchet derivative of $\boldsymbol{\ell}$ with respect to ϕ).[8] Given an estimate of a least favorable curve, the marginal Fisher information matrix for $\boldsymbol{\eta}$ is the expected squared length of the residual of the parametric score $\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$ after projection onto the nonparametric score $\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$. To compute this projection we need the least favorable direction $\mathbf{v}$, which is the tangent to the least favorable curve. Hence, the marginal Fisher information matrix for $\boldsymbol{\eta}$ is given by

$$\mathbf{I}^*_{\boldsymbol{\eta}} = \mathrm{E}\left[\left(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)) + \mathbf{v}\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))\right)\left(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)) + \mathbf{v}\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))\right)'\right]. \quad (6)$$

4.1 Joint Estimation

An estimator of the least favorable systematic risk is obtained by maximizing a local (and hence smoothed) likelihood function (see Staniswalis (1987) and Staniswalis (1989)), and simultaneously estimate the parameter vector $\boldsymbol{\eta}$ by maximizing the un–smoothed likelihood function. For a given value $z_0 \in [0, 1]$, and fixed values of $\boldsymbol{\eta}$, we estimate $\phi(z_0)$ as the solution of the problem

$$\hat{\phi}_{T\,\boldsymbol{\eta}}(z_0) = \arg\sup_{\phi \in \mathcal{P}} \sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)), \quad (7)$$

where K is a suitable kernel function and h_T is the corresponding bandwidth (see assumptions in the Appendix for details). Note that the estimator depends on $\boldsymbol{\eta}$ and it is point wise, i.e. there are as many optimizations of (7) as observations. Given the estimates for the nonparametric function, an un–smoothed ML estimation for $\boldsymbol{\eta}$ is performed:

$$\hat{\boldsymbol{\eta}}_T = \arg\max_{\boldsymbol{\eta} \in \boldsymbol{\Lambda}} \sum_{t=1}^{T} \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)). \quad (8)$$

Staniswalis (Staniswalis (1987) and Staniswalis (1989)), Severini and Wong (1992) and Veredas *et al.* (2007) show the asymptotic properties of the joint estimator $(\hat{\boldsymbol{\eta}}_T, \hat{\phi}_{T\,\boldsymbol{\eta}}(z_0))$ obtained by satisfying jointly (7) and (8). The next two Theorems show their consistency and asymptotic distribution.

Theorem 1 - (Consistency and asymptotic normality of $\hat{\phi}_{T\,\boldsymbol{\eta}}$) *Under the assumptions in the Appendix, if for each $z_0 \in [0, 1]$ and $\boldsymbol{\eta} \in \boldsymbol{\Lambda}$, $\hat{\phi}_{T\,\boldsymbol{\eta}}(z_0)$ satisfies (7). Then*

[8] Notice that $\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$ is a vector and $\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$ is a scalar.

a) $\hat{\phi}_{T\boldsymbol{\eta}}(z_0)$ *is an estimator of the least favorable curve;*

b) $\hat{\phi}_{T\boldsymbol{\eta}}(z_0)$ *is consistent, so for any* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that*

$$\lim_{Th\to\infty} \text{Prob}\left\{ \sup_{z_0\in[0,1]} \sup_{\boldsymbol{\eta}\in\boldsymbol{\Lambda}} |\hat{\phi}_{T,\boldsymbol{\eta}}(z_0) - \phi(z_0)| > \delta \right\} < \varepsilon;$$

c) for any $z_0 \in [0,1]$

$$\sqrt{Th_T}\left(\hat{\phi}_{T\boldsymbol{\eta}}(z_0) - \phi(z_0)\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\boldsymbol{\eta}}(z_0)\right) \text{ as } Th_T \to \infty,$$

where

$$V_{\boldsymbol{\eta}}(z_0) = \left[\int_{-1}^{1} \mathrm{K}^2(u)\mathrm{d}u\right] [i_{\boldsymbol{\eta}}(z_0)]^{-1}, \text{ and } i_{\boldsymbol{\eta}}(z_0) = \mathrm{E}\left[\left(\nabla_\phi \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))\right)^2 \middle| z_t = z_0\right].$$

Theorem 2 - (Consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_T$) *Under the assumptions in the Appendix, let* $\hat{\boldsymbol{\eta}}_T$ *satisfy (8). Then*

a) $\hat{\boldsymbol{\eta}}_T$ *is consistent, given* $\hat{\phi}_{T,\boldsymbol{\eta}}(z_0)$*, so for any* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that*

$$\lim_{T\to\infty} \text{Prob}\left\{||\hat{\boldsymbol{\eta}}_T - \boldsymbol{\eta}|| > \delta\right\} < \varepsilon;$$

b) $\hat{\boldsymbol{\eta}}_T$ *is asymptotically normal:*

$$\sqrt{T}\left(\hat{\boldsymbol{\eta}}_T - \boldsymbol{\eta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{\boldsymbol{\eta}}^{*-1}\right) \text{ as } T \to \infty,$$

where

$$\begin{aligned}
\mathbf{I}_{\boldsymbol{\eta}}^{*} &= \mathrm{E}\left[(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))'\right] + \\
&- \mathrm{E}\left[(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))(\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))\right] \mathrm{E}\left[\nabla_{\phi\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0))\right]^{-1} \\
&\quad \mathrm{E}\left[(\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))'\right].
\end{aligned}$$

From Lemma 1 in Severini and Wong (1992) the explicit form of the least favorable direction is

$$\mathbf{v} = -\mathrm{E}\left[(\nabla_{\boldsymbol{\eta}}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))(\nabla_{\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0)))\right] \mathrm{E}\left[\nabla_{\phi\phi}\boldsymbol{\ell}(\boldsymbol{\eta}, \hat{\phi}_{T\boldsymbol{\eta}}(z_0))\right]^{-1}. \quad (9)$$

By substituting (9) into (6) we get the explicit form of the seminonparametric Fisher information matrix $\mathbf{I}_{\boldsymbol{\eta}}^{*}$. The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\eta}}_T$ is the inverse of the marginal Fisher information matrix of a seminonparametric model, so that it is the natural lower bound for seminonparametric models. Therefore, according to the definition of the least favorable curve and, given the ML estimators obtained from (7) and (8), we attain asymptotic efficiency.

4.2 Two–step Estimation

In practice optimization of the full log-likelihood in (3) is cumbersome. The procedure to estimate $\boldsymbol{\eta}$ and $\phi(z_t)$ is based on iterating between the smoothed and the un–smoothed optimizations. At each iteration, the smooth optimization has to be done T times and the un–smoothed optimization is with respect to $5N$ parameters in $\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\phi(z_t))$ plus the p_ψ parameters in $\sum_{t=1}^{T}\boldsymbol{\ell}^c(\boldsymbol{\xi},\boldsymbol{\psi},\phi(z_t))$.

This procedure is computationally intensive and it is significantly alleviated if we consider Inference from the Marginals (Joe, 1997 and Joe, 2005), a two–step estimation procedure. The first step consists of estimating $\phi(z_0)$ and $\boldsymbol{\xi}$ by making use only of the first term of the log-likelihood (3). For any $z_0 \in [0,1]$, $\hat{\phi}_{T,\boldsymbol{\xi}}(z_0)$ must now fulfill

$$\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0) = \arg\sup_{\phi\in\mathcal{P}} \sum_{t=1}^{T}\sum_{i=1}^{N} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right)\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\phi(z_t)). \tag{10}$$

This optimization has a closed form solution

$$\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0) = \frac{\sum_{t=1}^{T}\mathrm{K}\left(\frac{z_0-z_t}{h_T}\right)\sum_{i=1}^{N}\frac{x_{i\,t}}{a_i\mu_{i\,t}(\mathcal{F}_{j-1},\boldsymbol{\delta}_i)}\frac{\nu_i}{\sum_{i=1}^{N}\nu_i}}{\sum_{t=1}^{T}\mathrm{K}\left(\frac{z_0-z_t}{h_T}\right)}, \tag{11}$$

which has an intuitive meaning: the systematic risk on z_0 is estimated as a nonparametric regression of a weighted sum (across N) of risks adjusted by the idiosyncratic component and a_i where the proportional weights are a function of all the ν_i's. Since ν_i is the reciprocal of the variance of the i-th innovation, the weights also have an intuitive interpretation: the less erratic volatilities (denoted by larger ν's) have more weight in the estimation of the systematic risk.

Consistency and asymptotic normality of $\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0)$ are proved in the following Theorem.

Theorem 3 - (Consistency and asymptotic normality of $\hat{\phi}_{T\,\boldsymbol{\xi}}$) *Under the assumptions in the Appendix:*

a) $\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0)$ *given by* (11) *is an estimator of the least favorable curve;*

b) $\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0)$ *is consistent, so for any* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that*

$$\lim_{Th\to\infty} \text{Prob}\left\{\sup_{z_0\in[0,1]}\sup_{\boldsymbol{\eta}\in\boldsymbol{\Lambda}}|\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0) - \phi(z_0)| > \delta\right\} < \varepsilon;$$

b) for any $z_0 \in [0,1]$,

$$\sqrt{Th_T}\left(\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0) - \phi(z_0)\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\boldsymbol{\xi}}(z_0)\right) \text{ as } Th_T \to \infty,$$

where

$$V_{\boldsymbol{\xi}}(z_0) = \left[\int_{-1}^{1}\mathrm{K}^2(u)\mathrm{d}u\right] i_{\boldsymbol{\xi}}(z_0)\; h_{\boldsymbol{\xi}}^{-2}(z_0),$$

and

$$
\begin{aligned}
i_{\boldsymbol{\xi}}(z_0) &= \mathrm{E}\left[\left(\nabla_{\phi}\sum_{i=1}^{N}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\phi(z_t))\right)^2\middle| z_t=z_0\right],\\
h_{\boldsymbol{\xi}}(z_0) &= -\mathrm{E}\left[\nabla_{\phi\phi}\sum_{i=1}^{N}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\phi(z_t))\middle| z_t=z_0\right].
\end{aligned}
$$

To do inference (i.e. standard errors, testing and confidence intervals) the expressions in $V_{\boldsymbol{\xi}}(z_0)$ have to be replaced by sample equivalents. The integral of the squared kernel is a constant specific to the chosen kernel. Since we use the quartic kernel, it equals $5/7$. The Fisher information matrix $i_{\boldsymbol{\xi}}(z_0)$ and the second derivative $h_{\boldsymbol{\xi}}(z_0)$ are replaced by

$$
\begin{aligned}
\hat{i}_{\boldsymbol{\xi}}(z_0) &= \frac{1}{T}\sum_{t=1}^{T}\mathrm{K}\left(\frac{z_0-z_t}{h_T}\right)\left[\sum_{i=1}^{N}\frac{\hat{\nu}_i}{\hat{\phi}_{T\boldsymbol{\xi}}(z_0)}\left(\frac{x_{it}}{\hat{a}_i\mu_{it}(\mathcal{F}_{j-1},\hat{\boldsymbol{\delta}}_i)\hat{\phi}_{T\boldsymbol{\xi}}(z_0)}-1\right)\right]^2 \text{ and}\\
\hat{h}_{\boldsymbol{\xi}}(z_0) &= \frac{1}{T}\sum_{t=1}^{T}\mathrm{K}\left(\frac{z_0-z_t}{h_T}\right)\sum_{i=1}^{N}\frac{\hat{\nu}_i}{\hat{\phi}_{T\boldsymbol{\xi}}(z_0)^2}\left(\frac{2x_{it}}{\hat{a}_i\mu_{it}(\mathcal{F}_{j-1},\hat{\boldsymbol{\delta}}_i)\hat{\phi}_{T\boldsymbol{\xi}}(z_0)}-1\right),
\end{aligned}
$$

which are smoothed expectations of weighted standardized residuals (squared for $\hat{i}_{\boldsymbol{\xi}}(z_0)$). The efficiency loss due to neglecting the presence of $\phi(z_0)$ in the copula function $\sum_{t=1}^{T}\boldsymbol{\ell}^c(\boldsymbol{\xi},\boldsymbol{\psi},\phi(z_t))$ is shown in the following Proposition.

Proposition 1 - (Asymptotic efficiency of $\hat{\phi}_{T\boldsymbol{\xi}}$) *Under the assumptions in the Appendix, for any $z_0\in[0,1]$, the estimator $\hat{\phi}_{T\boldsymbol{\xi}}(z_0)$ is less efficient than $\hat{\phi}_{T\boldsymbol{\eta}}(z_0)$, i.e. $V_{\boldsymbol{\xi}}(z_0)\geq V_{\boldsymbol{\eta}}(z_0)$.*

Given $\hat{\phi}_{T\boldsymbol{\xi}}(z_0)$, and since $\boldsymbol{\Xi}=\boldsymbol{\Xi}_1\times\ldots\boldsymbol{\Xi}_N$, the optimization of $\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\phi(z_t))$ with respect to $(\boldsymbol{\xi}_1\ldots\boldsymbol{\xi}_N)$ boils down to N independent optimizations of the marginals:

$$
\hat{\boldsymbol{\xi}}_{Ti}=\arg\max_{\boldsymbol{\xi}_i\in\boldsymbol{\Xi}_i}\sum_{t=1}^{T}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\hat{\phi}_{T\boldsymbol{\xi}}(z_t))\;\; i=1,\ldots,N. \tag{12}
$$

We now clearly see the gain of the Inference from the Marginals. The complicated iterative process between the smoothed optimization with respect to the systematic risk and the un–smoothed univariate optimizations with respect to $5N+p_\psi$ parameters, reduces to a simple iterative process between a closed form estimator, $\hat{\phi}_{T\boldsymbol{\xi}}(z_t)$, and N un–smoothed optimizations with respect to 5 parameters each ($\boldsymbol{\xi}_i=(a_i,\alpha_i,\beta_i,\gamma_i,\nu_i)$). Table 3 provides detailed explanations of the iterative process.

[TABLE 3 ABOUT HERE]

The initialization of the systematic risk – $\phi^0(z_0)$ – needs some further clarification. Theorem 3 states that $\hat{\phi}_{T\boldsymbol{\xi}}(z_0)$ is a consistent estimator for any $\boldsymbol{\xi}$ and z_0. Therefore the initial estimator $\phi^0(z_0)$ needs to be evaluated in an educated initialization of $a_i\mu_{it}(\mathcal{F}_{j-1},\boldsymbol{\delta}_i)$

and ν_i, $\forall i$. As for $a_i\mu_{i\,t}(\mathcal{F}_{j-1}, \boldsymbol{\delta}_i)$, since it is the conditional mean of the i-th risk adjusted for the systematic component, we consider the sample means $\bar{x}_i$. As for ν_i, since it is the inverse of the variance of a gamma distribution, we consider the inverse of the sample variance s_i^2. This is analogous to the device used in Hafner and Linton (2009) where an educated initialization of the smooth component of the variance-covariance matrix is a Naradaya Watson estimator of the standardized time on the outer product of vector of returns.[9]

In the second step we estimate $\boldsymbol{\psi}$ by maximum likelihood:

$$\hat{\boldsymbol{\psi}}_T = \arg\max_{\boldsymbol{\psi}\in\,\boldsymbol{\Psi}} \sum_{t=1}^{T} \boldsymbol{\ell}^c(\hat{\boldsymbol{\xi}}_T, \boldsymbol{\psi}, \hat{\phi}_{T\,\boldsymbol{\xi}}(z_t)). \tag{13}$$

Since $\sum_{t=1}^{T} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))$, and $\sum_{t=1}^{T} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t))$ are proper log–likelihoods, consistency is ensured (Joe, 1997, p. 301). Asymptotic normality is also achieved but, due to the two–step nature of the optimization, the asymptotic covariance matrix is modified, following Joe (2005), with respect to that of Theorem 2. This modification is in order to take into account the parameter uncertainty of the first step into the second step estimators.

Theorem 4 - (Consistency and asymptotic normality of $\hat{\boldsymbol{\xi}}_T$ and $\hat{\boldsymbol{\psi}}_T$) *Let $(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T)$ be the vector of parametric estimates obtained by solving first* (12) *and second* (13)*. Consider the consistent estimator of a least favorable curve $\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0)$. Under the assumptions in the Appendix:*

a) $(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T)$ is consistent, so for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\lim_{T\to\infty} \text{Prob}\left\{\left\|\begin{pmatrix} \hat{\boldsymbol{\xi}}_T \\ \hat{\boldsymbol{\psi}}_T \end{pmatrix} - \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\psi} \end{pmatrix}\right\| > \delta\right\} < \varepsilon;$$

b) $(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T)$ is asymptotically normal

$$\sqrt{T}\left(\begin{pmatrix} \hat{\boldsymbol{\xi}}_T \\ \hat{\boldsymbol{\psi}}_T \end{pmatrix} - \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\psi} \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{H}^{-1}\,\mathbf{I}^*_{\boldsymbol{\xi},\boldsymbol{\psi}}\,\mathbf{H}^{-1\prime}\right) \qquad \text{as } T \to \infty.$$

[9] An alternative avenue to avoid the use of profile likelihood for the estimation of the systematic risk is to consider the nonparametric regression $\tilde{x}_{t\,\boldsymbol{\xi}} = m(z_t) + \tilde{\epsilon}_t$ where

$$\tilde{x}_{t\,\boldsymbol{\xi}} = \sum_{i=1}^{N} \frac{x_{i\,t}}{a_i\mu_{i\,t}(\mathcal{F}_{j-1}, \boldsymbol{\delta}_i)} \frac{\nu_i}{\sum_{i=1}^{N}\nu_i}.$$

The Nadaraya-Watson estimator is also (11). While the issue of the initialization $\phi^0(z_0)$ is still present, the advantage of proceeding in this way is that the parameters do not need to be maximized with a gamma distribution but more general distributions can be considered. However, it is not clear that the resulting estimator is a least favorable curve and the estimated parameters reach the seminonparametric efficiency bound.

The matrix $\mathbf{H}$ *is defined as*

$$\mathbf{H} = \left(\begin{array}{ccc|c} \mathcal{H}_{\boldsymbol{\xi}_1 \boldsymbol{\xi}_1} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathcal{H}_{\boldsymbol{\xi}_N \boldsymbol{\xi}_N} & \mathbf{0} \\ \hline \mathcal{H}_{\boldsymbol{\psi} \boldsymbol{\xi}_1} & \dots & \mathcal{H}_{\boldsymbol{\psi} \boldsymbol{\xi}_N} & \mathcal{H}_{\boldsymbol{\psi} \boldsymbol{\psi}} \end{array} \right),$$

where, for any $i = 1, \dots, N$*,*

$$\mathcal{H}_{\boldsymbol{\xi}_i \boldsymbol{\xi}_i} = -\mathrm{E}\left[\nabla_{\boldsymbol{\xi}_i \boldsymbol{\xi}_i} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right], \;\; \mathcal{H}_{\boldsymbol{\psi} \boldsymbol{\xi}_i} = -\mathrm{E}\left[\nabla_{\boldsymbol{\psi} \boldsymbol{\xi}_i} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t))\right],$$

and

$$\mathcal{H}_{\boldsymbol{\psi} \boldsymbol{\psi}} = -\mathrm{E}\left[\nabla_{\boldsymbol{\psi} \boldsymbol{\psi}} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t))\right].$$

The information matrix is defined as $\mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}} = \mathbf{I}_{\boldsymbol{\xi}\,\boldsymbol{\psi}} - \mathbf{W}_\phi$*, where* $\mathbf{I}_{\boldsymbol{\xi}\,\boldsymbol{\psi}}$ *is the block–diagonal information matrix of the fully parametric case and* $\mathbf{W}_\phi$ *is a correction term due to the presence of the curve. Namely,*

$$\begin{aligned} \mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}} \;=\; & \underbrace{\left(\begin{array}{ccc|c} \mathcal{I}_{\boldsymbol{\xi}_1 \boldsymbol{\xi}_1} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathcal{I}_{\boldsymbol{\xi}_N \boldsymbol{\xi}_N} & \mathbf{0} \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathcal{I}_{\boldsymbol{\psi} \boldsymbol{\psi}} \end{array} \right)}_{\mathbf{I}_{\boldsymbol{\xi}, \boldsymbol{\psi}}} + \\ & - \underbrace{\left(\begin{array}{ccc|c} \mathcal{I}_{\boldsymbol{\xi}_1 \phi} \mathcal{H}^{-1}_{\phi\,\phi} \mathcal{I}_{\phi\,\boldsymbol{\xi}_1} & \dots & \mathcal{I}_{\boldsymbol{\xi}_1 \phi} \mathcal{H}^{-1}_{\phi\,\phi} \mathcal{I}_{\phi\,\boldsymbol{\xi}_N} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{I}_{\boldsymbol{\xi}_N \phi} \mathcal{H}^{-1}_{\phi\,\phi} \mathcal{I}_{\phi\,\boldsymbol{\xi}_1} & \dots & \mathcal{I}_{\boldsymbol{\xi}_N \phi} \mathcal{H}^{-1}_{\phi\,\phi} \mathcal{I}_{\phi\,\boldsymbol{\xi}_N} & \mathbf{0} \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathcal{I}_{\boldsymbol{\psi}\,\phi} \mathcal{H}^{-1}_{\phi,\phi} \mathcal{I}_{\phi\,\boldsymbol{\psi}} \end{array} \right)}_{\mathbf{W}_\phi} \end{aligned}$$

where, for any $i = 1, \dots, N$*, the generic elements of* $\mathbf{I}_{\boldsymbol{\xi}\,\boldsymbol{\psi}}$ *are*

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\xi}_i \boldsymbol{\xi}_i} &= \mathrm{E}\left[\left(\nabla_{\boldsymbol{\xi}_i} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right)\left(\nabla_{\boldsymbol{\xi}_i} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right)'\right], \\ \mathcal{I}_{\boldsymbol{\psi} \boldsymbol{\psi}} &= \mathrm{E}\left[\left(\nabla_{\boldsymbol{\psi}} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t))\right)\left(\nabla_{\boldsymbol{\psi}} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t))\right)'\right], \end{aligned}$$

and the generic elements of $\mathbf{W}_\phi$ *are*

$$\begin{aligned} \mathcal{H}_{\phi\,\phi} &= \mathrm{E}\left[\nabla_{\phi\phi} \sum_{i=1}^N \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right], \\ \mathcal{I}_{\boldsymbol{\xi}_i \phi} &= \mathrm{E}\left[\left(\nabla_{\boldsymbol{\xi}_i} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right)\left(\nabla_\phi \sum_{i=1}^N \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right)\right], \;\; \mathcal{I}_{\phi\,\boldsymbol{\xi}_i} = \mathcal{I}'_{\boldsymbol{\xi}_i \phi}, \\ \mathcal{I}_{\boldsymbol{\psi}\,\phi} &= \mathrm{E}\left[\left(\nabla_{\boldsymbol{\psi}} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \psi, \phi(z_t))\right)\left(\nabla_\phi \boldsymbol{\ell}^c(\boldsymbol{\xi}, \psi, \phi(z_t))\right)\right], \;\; \mathcal{I}_{\phi\,\boldsymbol{\psi}} = \mathcal{I}'_{\boldsymbol{\psi}\,\phi}. \end{aligned}$$

Given the estimated parameters $\hat{\boldsymbol{\xi}}_{T\,i}$ and the systematic risk $\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t)$, let $\hat{u}_{i\,t}$ be the cdf $F_{x_i}(x_{i\,t}|\mathcal{F}_{t-1}; \hat{\boldsymbol{\xi}}_{T\,i}, \hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))$ that is used to construct $\boldsymbol{\ell}^c(\hat{\boldsymbol{\xi}}_T, \boldsymbol{\psi}, \hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))$. As explained

in Section 3, since we map $\hat{u}_{i\,t}$ into a Gaussian frame the dependence parameter $\boldsymbol{\psi}$ becomes $\mathbf{R}$ that is estimated as the sample correlation matrix of $\hat{\mathbf{y}}_t = (\hat{y}_{1,t} \dots \hat{y}_{N,t}) = (\Phi^{-1}(\hat{u}_{1,t}) \dots \Phi^{-1}(\hat{u}_{N,t}))$.

In a fully parametric case, i.e. the case where $\mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}}$ does not contain the term $\mathbf{W}_\phi$, the variance–covariance matrix of $(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T)$ boils down to $\mathbf{I}_{\boldsymbol{\xi}\,\boldsymbol{\psi}}$. This latter term is not block diagonal as the systematic risk is contained in all the marginal distributions. Notice also that the matrix $\mathbf{H}$ remains even in the fully parametric case as it is due to the two–step procedure.

We define the upper left block of $\mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}}$ relative to $\boldsymbol{\xi}$ as $\mathbf{I}^*_{\boldsymbol{\xi}}$ and the lower right block relative to $\boldsymbol{\psi}$ as $\mathcal{I}^*_{\boldsymbol{\psi}\,\boldsymbol{\psi}}$. We also define the marginal Fisher information matrix relative to $\boldsymbol{\xi}$ when maximizing the whole likelihood as $\mathbf{I}^*_{\boldsymbol{\eta}|\boldsymbol{\xi}}$, and analogously we define $\mathbf{I}^*_{\boldsymbol{\eta}|\boldsymbol{\psi}}$. The following Proposition states the efficiency loss due to the Inference from the Marginal.[10]

Proposition 2 - (Asymptotic efficiency of $\hat{\boldsymbol{\xi}}_T$ and $\hat{\boldsymbol{\psi}}_T$) *Under the assumptions in the Appendix, the estimator $(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T)$ is less efficient than $\hat{\boldsymbol{\eta}}_T$, i.e.*

$$\mathbf{H}^{-1} \begin{pmatrix} \mathbf{I}^*_{\boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}^*_{\boldsymbol{\psi}\,\boldsymbol{\psi}} \end{pmatrix} \mathbf{H}^{-1'} \succeq \begin{pmatrix} \mathbf{I}^{*-1}_{\boldsymbol{\eta}|\boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}^{*-1}_{\boldsymbol{\psi}\,\boldsymbol{\psi}} \end{pmatrix}.$$

5 Disentangling: Practice

Consistently with Section 2, we provide detailed results for SPDR and summary results for S&P100 and, though estimations are done on variances, all the results are reported in terms of annualized volatilities. For comparison purposes, we also fit univariate asymmetric MEMs (simply referred to as MEM) on each series to provide baseline results. These models can be formulated as a constrained version of our specification when the trend $\phi(z_t)$ is set to be equal to one for all z_t and there is no copula.

[TABLE 4 ABOUT HERE]

Table 4 reports the estimated parameter for the idiosyncratic risks (columns α_i, β_i and γ_i), the estimated persistence $\pi_i = \alpha_i + \beta_i + \frac{1}{2}\gamma_i$ that captures durability of departures of the idiosyncratic risks around the systematic risk, and the Quasi Log Likelihood Loss (QLL) that measures the in–sample fit of the model (Patton (2010)). The left part of the panel shows the results for our model while the right part shows the results for the MEM.

We first focus on the estimation results of our model. In comparison to typical GARCH estimates, values of α and γ are higher and β's are lower (see Brownlees and Gallo (2010) and Shephard and Sheppard (2010) for similar evidence). This is a consequence of the fact that realized kernels provide more accurate measurement of volatility in comparison to squared returns, hence past realizations of the process turn out to be more informative

[10] Given two matrices $\mathbf{A}$ and $\mathbf{B}$, of the same dimensions, we use the notation $\mathbf{A} \succeq \mathbf{B}$ to indicate that $(\mathbf{A} - \mathbf{B})$ is a positive semidefinite matrix.

and are associated with higher estimates of the α and γ parameters. As expected, the asymmetry is positive suggesting that negative news increase the level of volatility more than positive ones. The persistence reveals important differences across assets. For SPDR the sectors with higher persistence are energy and technology, meaning that these are the sectors with longer lasting idiosyncratic departures from the systematic risk. For S&P100 the differences between the least and the most persistent idiosyncratic risks are wide and, in general, they appear to be higher than the ones in SPDR. We interpret this as the result that individual assets have a higher level of idiosyncrasy in comparison to SPDR. The systematic loadings are fairly close to the sample average volatilities (see Table 3). Last, the estimates of the gamma distribution also differ substantially across sectors and assets meaning differences in the marginals probability distributions. By contrast, in the vast majority of cases, the persistences entailed by the MEMs are essentially equal to one, hinting at the presence of integrated dynamics. The unconditional volatilities (denoted by a_i for comparison purposes with our model), which are relevant for long-run forecasting, are often far –especially for SPDR– from the sample average volatilities, suggesting that the stationarity condition may be violated. The estimated variances of the innovations terms in the univariate MEM are also higher than the ones of our model in the great majority of cases. Lastly, the QLL loss function is always smaller for our model, meaning that our model always improves the fit over the simple MEM.

[FIGURE 2 ABOUT HERE]

Figure 2 displays, for SPDR, the scatter plot of the log of the systematic loadings (for our model) and the log of the unconditional expected volatilities (for the MEMs) versus the persistence. Results from our model suggest that sectors cluster in groups. Technology is the sector with the largest persistence and volatility and it is closely followed by Energy and Utilities. The Financial, Industrial, and Health Care sectors have median levels of volatility and persistence. Material and Consumer Discretionary have low volatility and the lowest persistence. Lastly, Consumer Staples has the lowest volatility and average persistence. The same results using MEMs turn out to be harder to interpret. The non-stationarity in volatility makes the estimated level of persistence collapsing to one, and differences in parameter estimates across assets ultimately reflect estimation variability.

[FIGURE 3 ABOUT HERE]

Figure 3 shows (from top to bottom) the estimated fit, the systematic volatility, and idiosyncratic volatilities for SPDR (left) and S&P100 (right). The fit $-\sqrt{252\ \hat{a}_{T\,i}\ \hat{\phi}_{T\,\boldsymbol{\xi}}(z_t)\ \hat{\mu}_{i\,t}}-$ fairly mimics the movements of realized volatilities, even during the 2008 financial crisis. The systematic volatility $-\sqrt{\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t)}-$ for both panels are essentially the same with minor differences in scale and pattern that reflect the difference composition of assets. The line at one is used to emphasize periods of amplification and contraction. The systematic volatility captures low frequency movements in the volatility trend which reflects the general economic conditions commented in Section 2: In mid 2002 the systematic volatility increased by almost half, during the volatility moderation period systematic volatility was

around 3/4 of its unconditional value and, finally, in the recession of the late 2000s the index increased to more than double in the last quarter of 2008.

The idiosyncratic volatilities $-\sqrt{252\ \hat{a}_{T\,i}\ \hat{\mu}_{i\,t}}-$ are stationary and vary around the unconditional means. To get deeper insights, Figure 4 displays the idiosyncratic volatilities of the Energy, Financial and Technology sector only, which allows to visually identify periods of idiosyncratic distress of these sectors. In the beginning of the 2001, Technology was the most volatile sector due the aftermath of the burst of the dot com bubble. In between 2005 and 2007 concerns on oil prices generated an increased level of uncertainty in the Energy sector. Finally, the Financial sector had a surge in volatility starting from July 2007 with the beginning of the credit crunch.

[FIGURE 4 ABOUT HERE]

Table 5 shows the estimates for the dependence among volatility innovations for SPDR while Figure 5 shows them for S&P100 in a heat map. In the plot sectors are delimited by horizontal and vertical lines and, for comparison purposes, we also show the heat map of the realized volatilities. The correlatons range from 0.32 to 0.50 for SPDR and from 0.07 to 0.60 for S&P100, and the average are 0.40 and 0.26 respectively. This difference reflects the already mentioned effect that the sectoral aggregation may entail an increase in the correlations. The row $\bar{\psi}_i$ in the Table shows the cross-sectional average dependence, which are all around 0.40. Though important, they are substantially smaller than the average correlations $\bar{\rho}$ in Table 2. In fact, last row of Table 5 shows the percentage difference of $\bar{\psi}_i$ with respect to $\bar{\rho}$. After taking into account the systematic and idiosyncratic components of volatilities, around 50% of the dependencies are still present. More interestingly these dependencies are not homogeneous. The heat map unveils clustering among stocks that belong to the same sector that are not visible from the raw volatilities. This is the case for Technology, Financials, Energy and Utilities. This finding dovetails with the conclusions extracted from the idiosyncratic volatilities in Figure 4 and puts forward evidence for sectoral systematic components in volatilities.[11]

[TABLE 5 ABOUT HERE]

[FIGURE 5 ABOUT HERE]

A natural question that rises at this point is what the proportion of volatility would be that is explained by the idiosyncratic components. To do so we introduce a goodness of fit measure that is in spirit of the pseudo R^2 used in the discrete choice models literature. For the i-th asset we define

$$GoF_{i\,\mu} = 1 - \frac{\boldsymbol{\ell}_i^m(x_{i\,t}, \widehat{\boldsymbol{\xi}}_{T\,i}, \widehat{\phi}_{T\,\boldsymbol{\xi}}(z_t))}{\boldsymbol{\ell}_i^m(x_{i\,t}, \widehat{\nu}_{T\,i}, \widetilde{\phi}_{T\,\boldsymbol{\xi}}(z_t))}.$$

[11] Since the independence among risk innovations is estimated with a Pearson correlation matrix on $y_{1\,t}, \ldots, y_{N\,t}$, we also estimated the rank-based Spearman's correlation matrix, as a check of the potential lost of dependencies beyond linearity (available under request). The two matrices are very similar.

The numerator is the log–likelihood of the i-th marginal model evaluated at estimates of the model, while the denominator is the same log–likelihood but imposing $\mu_{i\,t} = 1$ (so that $\widehat{\boldsymbol{\xi}}_{T\,i}$ contains only the estimated parameters of the marginal distribution) and a_i are fixed to the sample mean of the risk of the i-th asset. The systematic risk is estimated as

$$\widetilde{\phi}_{T\,\boldsymbol{\xi}}(z_t) = \frac{\sum_{t=1}^{T} K\left(\frac{z_j - z_t}{h_T}\right) \sum_{i=1}^{N} \frac{x_{i\,t}}{a_i} \frac{\nu_i}{\sum_i \nu_i}}{\sum_{t=1}^{T} K\left(\frac{z_j - z_t}{h_T}\right)}.$$

$GoF_{i\,\mu}$ measures the relative contribution of the idiosyncratic risk to the log–likelihood. Table 6 reports full results for SPDR and quantiles for the S&P100. The contributions of the idiosyncratic risks are substantial. The range of $GoF_{i\,\mu}$ is between 0.14 and 0.33 for SPDR and 0.12 and 0.55 for S&P100, with the latter panel exhibiting more idiosyncrasy (average of 0.46 for S&P100 vs 0.28 for SPDR).

[TABLE 6 ABOUT HERE]

Last, we find the residuals of the model to be adequate. The top panel of Table 7 reports the residual autocorrelations and the p-values of Ljung–Box tests at 1, 5 and 22 lags (daily, weekly, monthly), together with the average lagged correlations with the other residuals at 1, 5 and 22 lags. The autocorrelations are negligible and there is virtually no evidence against Ljung–Box null of no correlation. Also, the average cross autocorrelations are exiguous. This is verified in the bottom part of the table that shows the lagged one residuals autocorrelations.

[TABLE 7 ABOUT HERE]

Likewise, Figure 6 reports, for S&P100, the residual lagged autocorrelation matrices for orders 1 and 5 in the form of heat maps. The cell $i\,j$ corresponds to the correlation between $\hat{\epsilon}_{i\,t}$ and $\hat{\epsilon}_{j\,t-l}$, where l equals one (top plot) or five (bottom plot). We conclude that, as for SPDR, the lagged cross autocorrelations are small and no dynamic pattern in the residuals is detected.

[FIGURE 6 ABOUT HERE]

6 Conclusion

Modeling large panels of volatilities may prove a formidable task if one allows for dynamic interdependence. We follow parsimony in parametric specification, exploiting the stylized fact that volatilities appear to be driven by an underlying factor that captures the secular systematic trend. This is consistent with other studies that have shown that the secular volatility deserves separate attention from the short run idiosyncratic dynamics.

The stylized facts that emerge in the analysis of such panels motivate us to propose a novel vector MEM specification that decomposes risk measures in a systematic and idiosyncratic components, and it further allows for cross-section dependence in the innovations. The systematic component is a slowly varying nonparametric function and the idiosyncratic components follow a parametric dynamics. We develop a theoretically sound estimation technique which rests on the theories of profile likelihood and copulas. Regardless of the dimension of the panel, estimation boils down to univariate maximum likelihood problems and one sample correlation matrix. The ease in estimation of our model makes it appealing in large dimensional applications.

We engage in the analysis of two panels of daily realized volatility measures between 2001 and 2008. The first panel consists of the nine SPDR Sectoral Indices of the S&P500 and the second panel contains the ninety constituents of the S&P100 that have been continuously trading in the sample period. We find that while there is strong evidence of nonstationary dynamics in all series, once a common component is accounted for all series exhibit mean reversion around it. The speed of reversion is rather heterogeneous across assets. For instance energy and technology companies have a slow reversion while consumer, materials and healthcare assets are faster. The model also unveils dependencies in volatility innovations across assets and sectoral clusters for technology, financial, energy and utilities companies.

Further refinements or uses of the model can be envisaged. As realized measures are estimators, we would need to investigate the benefits of taking the measurement error in the volatility estimation into explicit account in the modeling step (cf. Hansen and Lunde (2010)). As in other contributions, a relationship between macroeconomic variables and the common component can help highlight some determinants of the changes in risk levels or, reverting the perspective, the spillover effects of market volatility onto the real economy.

Appendix A - Assumptions

The following assumptions are drawn from Severini and Wong (1992) and Veredas *et al.* (2007). They are stated for $\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))$ but hold also for $\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))$ and $\boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t))$. Moreover all the assumptions at the beginning of Section 2 hold and we assume also the standard regularity conditions for maximum likelihood estimation.

Assumption I.1 For fixed but arbitrary $\tilde{\boldsymbol{\eta}} \in \boldsymbol{\Lambda}$ and $\tilde{\phi}(z_t) \in \mathcal{P}$, let

$$\rho(\boldsymbol{\eta}, \phi(z_t)) = \int \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)) \mathrm{d}F_{\boldsymbol{x}}(\tilde{\boldsymbol{\eta}}, \tilde{\phi}(z_t)).$$

If $\phi(z_t) \neq \tilde{\phi}(z_t)$, then $\rho(\boldsymbol{\eta}, \phi(z_t)) < \rho(\tilde{\boldsymbol{\eta}}, \tilde{\phi}(z_t))$.

Assumption I.2 The seminonparametric marginal Fisher information matrix $\mathbf{I}^*_{\boldsymbol{\eta}}$ as defined in Theorem 2 is positive definite for all $\boldsymbol{\eta} \in \boldsymbol{\Lambda}$.

Assumption S Assume that for all $r, s = 0, \ldots, 4$, $r + s \leq 4$, the derivative

$$\frac{\partial^{r+s} \boldsymbol{\ell}}{\partial \boldsymbol{\eta}^r \partial \phi^s}(\boldsymbol{\eta}, \phi(z_t))$$

exist for almost all $\boldsymbol{x}_t$ and assume that

$$\mathrm{E}\left[\sup_{\boldsymbol{\eta} \in \Lambda} \sup_{\phi \in \mathcal{P}} \left|\left| \left(\frac{\partial^{r+s} \boldsymbol{\ell}}{\partial \boldsymbol{\eta}^r \partial \phi^s}(\boldsymbol{\eta}, \phi(z_t))\right) \left(\frac{\partial^{r+s} \boldsymbol{\ell}}{\partial \boldsymbol{\eta}^r \partial \phi^s}(\boldsymbol{\eta}, \phi(z_t))\right)' \right|\right|\right] < \infty.$$

Assumption K Assume that the kernel function $\mathrm{K}(\cdot)$ in (7) is of order $k > 3/2$ with support $[-1, 1]$ and it has bounded $k + 2$ derivatives. Assume also that

$$\int_{-1}^{1} \mathrm{K}(u)\mathrm{d}u = 1, \qquad \int_{-1}^{1} u\mathrm{K}(u)\mathrm{d}u = 0, \qquad \int_{-1}^{1} u^2\mathrm{K}(u)\mathrm{d}u < \infty,$$

and

$$\sup_u \left|\frac{\partial^r \mathrm{K}(u)}{\partial u^r}\right| < \infty, \qquad r = 0, \ldots, 4.$$

The bandwidth h_T is such that $h_T \to 0$, $Th_T \to \infty$, $Th_T^2 \to 0$ as $T \to \infty$. Moreover, it satisfies $h_T = O(T^{-\alpha})$ as $T \to \infty$ where

$$\frac{1}{8} < \alpha < \frac{1}{16}\frac{(q+3)(q-2)}{(q+6)(q+2)},$$

for some even integer $q \geq 10$.

Assumption A.1 The observations $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ are a sequence of ergodic random vectors.

Assumption A.2 $x_{i\,t}$ is a strong mixing process where the mixing coefficients $\{\alpha_j\}$ must satisfy

$$\sum_{j=1}^{\infty} j^{r-1}\alpha_j^{1-2/p} < \infty, \qquad \text{for } p > 2 \text{ and } r \in \mathbb{N}.$$

Furthermore, for some even integer $q \leq 2r$

$$\mathrm{E}[|x_{i\,t}|^q] < \nu,$$

where ν is a constant not depending on z_t.

Assumption B.1 Define

$$\mathcal{L}_0(\boldsymbol{\eta}, \phi(z_t)) \equiv \mathrm{E}_0\left[\boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t))\right]$$

where E_0 indicates expectation with respect to the true value of the parameters. Then, for each $\boldsymbol{\eta} \in \boldsymbol{\Lambda}$ and $z_t \in [0, 1]$, we assume that

$$\sup_{\boldsymbol{\eta}} \sup_{\phi} \sup_{z_t} \left| \frac{\partial^k}{\partial \boldsymbol{\eta}^k} \frac{\partial^h}{\partial z_t^h} \frac{\partial^j}{\partial \phi^j} \mathcal{L}_0(\boldsymbol{\eta}, \phi(z_t)) \right| < \infty,$$

for $j = 1, 2, 3$, $k = 0, 1, 2$ and $h = 0, 1$ such that $j + k + h \leq 4$.

Assumption B.2 Let $\phi_{\boldsymbol{\eta}}(z_t)$ denote a solution of

$$\frac{\partial \mathcal{L}_0}{\partial \phi}(\boldsymbol{\eta}, \phi(z_t)) = 0,$$

with respect to ϕ for each fixed $\boldsymbol{\eta}$ and z_t. Then we assume that $\phi_{\boldsymbol{\eta}}(z_t)$ is unique and that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_{\boldsymbol{\eta}} \sup_{z_t} \left| \frac{\partial \mathcal{L}_0}{\partial \phi}(\boldsymbol{\eta}, \bar{\phi}_{\boldsymbol{\eta}}(z_t)) \right| \leq \delta,$$

implies that

$$\sup_{\boldsymbol{\eta}} \sup_{z_t} \left| \bar{\phi}_{\boldsymbol{\eta}}(z_t) - \phi_{\boldsymbol{\eta}}(z_t) \right| \leq \varepsilon.$$

Assumption B.3 Let

$$T^{(j\,k)}(\boldsymbol{\eta}, \phi(z_t)) = \frac{\partial^j}{\partial \eta^j} \frac{\partial^k}{\partial \phi^k} \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)),$$

and let $f_{\boldsymbol{\eta}}^{(j\,k)}(\mathbf{x}_t | z_t)$ the conditional density of $T^{(j\,k)}(\boldsymbol{\eta}, \phi(z_t))$ given z_t. We assume that

a) $\mathrm{E}\left[\sup_{\boldsymbol{\eta}} \sup_{\phi} \left| T^{(j\,k)}(\boldsymbol{\eta}, \phi(z_t)) \right| \right] < \infty$, for $j = 0, \ldots, 3$ and $k = 0, \ldots, 5$.

b) For some even integer $q \geq 10$, $\sup_{\boldsymbol{\eta}} \sup_{\phi} \mathrm{E}\left[\left| T^{(j\,k)}(\boldsymbol{\eta}, \phi(z_t)) \right|^q \right] < \infty$, for $j = 0, \ldots, 3$ and $k = 0, \ldots, 4$.

c) $\sup_{\boldsymbol{\eta}} \sup_{\phi} \sup_{\mathbf{x}_t\, z_t} \left| f_{\boldsymbol{\eta}}^{(j\,k)}(\mathbf{x}_t | z_t) \right| < \infty$, for $j = 0, \ldots, 3$ and $k = 0, \ldots, 4$.

Appendix B - Proofs

Proof of Theorem 1

See Veredas *et al.* (2007). Details are also in Severini and Wong (1992) for part *a)* and Staniswalis (1987) for parts *b)* and *c)*.

Proof of Theorem 2

See Veredas *et al.* (2007).

Proof of Theorem 3

This Theorem is a special case of Theorem 1 in Staniswalis (1987) and we show only the main steps of the proof. Given an estimated curve $\hat{\phi}_{T\boldsymbol{\xi}}(z_0)$, the Taylor expansion of $\nabla_\phi \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \hat{\phi}_{T\boldsymbol{\xi}}(z_0))$ around the true curve $\phi(z_t)$ is

$$\nabla_\phi \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \hat{\phi}_{T\boldsymbol{\xi}}(z_0)) = \nabla_\phi \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t)) + \nabla_{\phi\phi} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \bar{\phi}(z_t))(\hat{\phi}_{T\boldsymbol{\xi}}(z_0) - \phi(z_t)), \quad \forall z_t \in [0,1],$$

where $\bar{\phi}$ lies between $\hat{\phi}_{T\boldsymbol{\xi}}$ and ϕ. Define the sum of the smoothed marginal log–likelihoods computed in a generic curve $\hat{\phi}_{\boldsymbol{\xi}}(z_0)$ as

$$\widetilde{\mathcal{L}}(\boldsymbol{\xi}, \hat{\phi}_{T\boldsymbol{\xi}}(z_0)) = \sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \hat{\phi}_{T\boldsymbol{\xi}}(z_0)). \tag{B-1}$$

From the first order conditions of (10) we have $\nabla_\phi \widetilde{\mathcal{L}}(\boldsymbol{\xi}, \hat{\phi}_{T\boldsymbol{\xi}}(z_0)) = 0$. Therefore, for any $z_0 \in [0,1]$,

$$
\begin{aligned}
0 \;=\; & \underbrace{\sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \nabla_\phi \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))}_{A_T} + \qquad \text{(B-2)} \\
& + \underbrace{\sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \bar{\phi}(z_t))\right] (\phi(z_0) - \phi(z_t))}_{D_T} + \\
& + \underbrace{\sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right]}_{B_T} (\hat{\phi}_{T\boldsymbol{\xi}}(z_0) - \phi(z_0)) + \\
& + \underbrace{\sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \bar{(}\phi)(z_t))\right]}_{C_{1T}} (\hat{\phi}_{T\boldsymbol{\xi}}(z_0) - \phi(z_0)) + \\
& - \underbrace{\sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right]}_{C_{2T}} (\hat{\phi}_{T\boldsymbol{\xi}}(z_0) - \phi(z_0)).
\end{aligned}
$$

Re-arranging we obtain

$$
\sqrt{Th_T}(\hat{\phi}_{T\boldsymbol{\xi}}(z_0) - \phi(z_0)) = -\frac{\frac{1}{\sqrt{Th_T}}(A_T + D_T)}{B_T + (C_{1T} - C_{2T})}. \qquad \text{(B-3)}
$$

Staniswalis (1987) proves that, as $Th_T \to \infty$ and $h_T \to 0$,

$$
A_T \xrightarrow{P} 0, \quad (C_{1T} - C_{2T}) \xrightarrow{P} 0, \quad D_T \xrightarrow{P} 0, \qquad \text{(B-4)}
$$

and

$$
\begin{aligned}
B_T \;\xrightarrow{P}\; & \sum_{t=1}^{T} \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \mathrm{E}\left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right] \\
= \; & \mathrm{E}\left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t)) \,\middle|\, z_t = z_0\right] = -h_{\boldsymbol{\xi}}(z_0).
\end{aligned}
$$

From (B-4) we attain consistency.

Moreover, as $Th_T \to \infty$ and $h_T \to 0$, (B-3) is asymptotically equivalent to $-\frac{1}{\sqrt{Th_T}} \frac{A_T}{B_T}$, and we have convergence in distribution

$$
A_T \xrightarrow{d} \mathcal{N}\left(0, V(A_T | z_t = z_0)\right).
$$

From (B-3)

$$V(A_T|z_t = z_0) = \left[\int_{-1}^{1} \mathrm{K}^2(u)\mathrm{d}u\right] \mathrm{E}\left[\left(\nabla_\phi \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t))\right)^2 \middle| z_t = z_0\right] = i_{\boldsymbol{\xi}}(z_0).$$

Therefore

$$\sqrt{Th_T}(\hat{\phi}_{T\,\boldsymbol{\xi}}(z_0) - \phi(z_0)) \xrightarrow{d} \mathcal{N}\left(0, \left[\int_{-1}^{1} \mathrm{K}^2(u)\mathrm{d}u\right] [i_{\boldsymbol{\xi}}(z_0) h_{\boldsymbol{\xi}}(z_0)^{-2}]\right)$$

which completes the proof of asymptotic normality. □

Proof of Proposition 1

For any $z_0 \in [0, 1]$, the statement $V_{\boldsymbol{\xi}}(z_0) \geq V_{\boldsymbol{\eta}}(z_0)$ is equivalent to

$$h_{\boldsymbol{\xi}}^2(z_0) \leq i_{\boldsymbol{\eta}}(z_0) i_{\boldsymbol{\xi}}(z_0). \tag{B-5}$$

We know that

$$h_{\boldsymbol{\xi}}^2(z_0) \leq i_{\boldsymbol{\xi}}^2(z_0), \tag{B-6}$$
$$h_{\boldsymbol{\eta}}^2(z_0) = i_{\boldsymbol{\eta}}^2(z_0). \tag{B-7}$$

The first relation is due to the fact that, when using only a limited amount of information, the squared second derivative of the log–likelihood and the Fisher information does not coincide as the latter is larger than the inverse of the asymptotic variance. The second relation states that in the full-information case the squared second derivative and the Fisher information are equal.

Second derivatives of the full log–likelihood, of the marginals, and of the copula density have the additivity property:

$$\begin{aligned} h_{\boldsymbol{\eta}}(z_0) &= -\mathrm{E}\left[\nabla_{\phi\phi} \boldsymbol{\ell}(\boldsymbol{\eta}, \phi(z_t)) \middle| z_t = z_0\right] = \\ &= -\mathrm{E}\left[\nabla_{\phi\phi} \sum_{i=1}^{N} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \phi(z_t)) + \nabla_{\phi\phi} \boldsymbol{\ell}^c(\boldsymbol{\xi}, \boldsymbol{\psi}, \phi(z_t)) \middle| z_t = z_0\right] = h_{\boldsymbol{\xi}}(z_0) + h_{\boldsymbol{\psi}}(z_0). \end{aligned} \tag{B-8}$$

We have the following inequalities:

$$h_{\boldsymbol{\xi}}^4(z_0) \leq h_{\boldsymbol{\eta}}^2(z_0) i_{\boldsymbol{\xi}}^2(z_0) \leq \left[h_{\boldsymbol{\xi}}^2(z_0) + h_{\boldsymbol{\psi}}^2(z_0) + 2h_{\boldsymbol{\xi}}(z_0) h_{\boldsymbol{\psi}}(z_0)\right] i_{\boldsymbol{\xi}}^2(z_0) = i_{\boldsymbol{\eta}}^2(z_0) i_{\boldsymbol{\xi}}^2(z_0). \tag{B-9}$$

The first inequality is given by (B-6); the second inequality is obtained since the existence of an internal maximum of the log–likelihoods implies that the second derivative computed in the maximum is negative, i.e. $h_{\boldsymbol{\xi}}(z_0) h_{\boldsymbol{\psi}}(z_0) \geq 0$, and clearly $h_{\boldsymbol{\psi}}^2(z_0) \geq 0$; the last equality is given by taking the square of (B-8) and by (B-7). By taking square roots in (B-9), we prove (B-5). □

We need the following Lemma to prove Theorem 4. For simplicity we indicate only the parameters $\boldsymbol{\xi}_i$ and $\boldsymbol{\psi}$ among the arguments of the log-likelihood functions.

Lemma 1

Under the assumptions in the Appendix A, for any $i, j = 1, \ldots, N$

$$\mathrm{E}\left[\left(\nabla_{\boldsymbol{\xi}_i}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i)\right)\left(\nabla_{\boldsymbol{\psi}}\boldsymbol{\ell}^c(\boldsymbol{\xi},\boldsymbol{\psi})\right)'\right] = \mathbf{0}, \tag{B-10}$$

and

$$\mathrm{E}\left[\left(\nabla_{\boldsymbol{\xi}_i}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i)\right)\left(\nabla_{\boldsymbol{\xi}_j}\boldsymbol{\ell}_j^m(\boldsymbol{\xi}_j)\right)'\right] = \mathbf{0}. \tag{B-11}$$

Proof of Lemma 1

(B-10) is in the Appendix in Joe (2005). The proof of (B-11) is similar. Let $\mathbf{x}_{-i}$ be the vector $\mathbf{x}$ omitting the i-th component. Then the expectation in (B-11) is equivalent to

$$\begin{aligned}
&\mathrm{E}\left[\left(\nabla_{\boldsymbol{\xi}_i}\log f_{x_i}(\boldsymbol{\xi}_i)\right)\left(\nabla_{\boldsymbol{\xi}_j}\log f_{x_j}(\boldsymbol{\xi}_j)\right)'\right] = \\
&= \int_{\mathbf{x}}\left(\nabla_{\boldsymbol{\xi}_i}\log f_{x_i}(\boldsymbol{\xi}_i)\right)\left(\nabla_{\boldsymbol{\xi}_j}\log f_{x_j}(\boldsymbol{\xi}_j)\right)'\left(\prod_{k=1}^N f_{x_k}(\boldsymbol{\xi}_k)\right)c(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_N,\boldsymbol{\psi})\mathrm{d}\mathbf{x} = \\
&= \int_{x_i}\left(\nabla_{\boldsymbol{\xi}_i}\log f_{x_i}(\boldsymbol{\xi}_i)\right)\left[\int_{\mathbf{x}_{-i}}\left(\nabla_{\boldsymbol{\xi}_j}\log f_{x_j}(\boldsymbol{\xi}_j)\right)'\left(\prod_{k=1}^N f_{x_k}(\boldsymbol{\xi}_k)\right)c(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_N,\boldsymbol{\psi})\mathrm{d}\mathbf{x}_{-i}\right]\mathrm{d}x_i = \\
&= \int_{x_i}\left(\nabla_{\boldsymbol{\xi}_i}\log f_{x_i}(\boldsymbol{\xi}_i)\right)\left[\int_{\mathbf{x}_{-i}}\left(\nabla_{\boldsymbol{\xi}_j} f_{x_j}(\boldsymbol{\xi}_j)\right)'\left(\prod_{k\neq j;k=1}^N f_{x_k}(\boldsymbol{\xi}_k)\right)c(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_N,\boldsymbol{\psi})\mathrm{d}\mathbf{x}_{-i}\right]\mathrm{d}x_i = \\
&= \int_{x_i}\left(\nabla_{\boldsymbol{\xi}_i}\log f_{x_i}(\boldsymbol{\xi}_i)\right)\left[\nabla_{\boldsymbol{\xi}_j}\int_{\mathbf{x}_{-i}}\left(\prod_{k=1}^N f_{x_k}(\boldsymbol{\xi}_k)\right)c(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_N,\boldsymbol{\psi})\mathrm{d}\mathbf{x}_{-i}\right]'\mathrm{d}x_i = \\
&= \int_{x_i}\left(\nabla_{\boldsymbol{\xi}_i}\log f_{x_i}(\boldsymbol{\xi}_i)\right)\left(\nabla_{\boldsymbol{\xi}_j} f_{x_i}(\boldsymbol{\xi}_i)\right)'\mathrm{d}x_i = \mathbf{0}.
\end{aligned}$$

An interchange of integration and differentiation is made in the before the last equality. □

Proof of Theorem 4

Given a consistent estimator of the curve $\hat{\phi}_{T\boldsymbol{\xi}}(z_0)$, the estimated parameters $(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T)$ have to satisfy (12) and (13). Hence, the first order conditions are

$$\left(\begin{array}{l} \nabla_{\boldsymbol{\xi}_1} \sum_{t=1}^{T} \boldsymbol{\ell}_1^m(\hat{\boldsymbol{\xi}}_{T1}, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \\ \cdots\cdots\cdots \\ \nabla_{\boldsymbol{\xi}_N} \sum_{t=1}^{T} \boldsymbol{\ell}_t^m(\hat{\boldsymbol{\xi}}_{TN}, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \\ \nabla_{\boldsymbol{\psi}} \sum_{t=1}^{T} \boldsymbol{\ell}^c(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \end{array} \right) = \mathbf{0}. \tag{B-12}$$

We define the score function $\boldsymbol{g} = (\boldsymbol{g}_1' \ldots \boldsymbol{g}_N' \;\; \boldsymbol{g}_{N+1}')'$ as

$$\begin{aligned} \boldsymbol{g}_i(\boldsymbol{\xi}_i, \hat{\phi}_{T\boldsymbol{\xi}}(z_0)) &= \nabla_{\boldsymbol{\xi}_i} \boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \text{ for any } i = 1, \ldots, N, \\ \boldsymbol{g}_{N+1}(\boldsymbol{\xi}, \boldsymbol{\psi}, \hat{\phi}_{T\boldsymbol{\xi}}(z_0)) &= \nabla_{\boldsymbol{\psi}} \boldsymbol{\ell}_t^c(\boldsymbol{\xi}_T, \boldsymbol{\psi}_T, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)). \end{aligned}$$

So (B-12) can be written as

$$\sum_{t=1}^{T} \boldsymbol{g}(\hat{\boldsymbol{\xi}}_T, \hat{\boldsymbol{\psi}}_T, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) = \mathbf{0}. \tag{B-13}$$

The first N components of g have dimension 5 each since $\boldsymbol{\xi}_i = (a_i, \alpha_i, \beta_i, \gamma_i, \nu_i)'$, while the $N+1$ component of g has dimension p_ψ. Therefore, g is a $(5N + p_\psi)$–dimensional column vector. Define also

$$\begin{aligned} \mathcal{A}_T(\boldsymbol{\xi}, \boldsymbol{\psi}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathrm{g}(\boldsymbol{\xi}, \boldsymbol{\psi}, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \text{ and} \\ \mathcal{B}_T &= -\frac{1}{T} \nabla_{\boldsymbol{\xi}\boldsymbol{\psi}} \left(\sum_{t=1}^{T} \mathrm{g}(\bar{\boldsymbol{\xi}}, \bar{\boldsymbol{\psi}}, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \right)' \end{aligned}$$

where $\bar{\boldsymbol{\xi}}$ lies between $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}_T$ and $\bar{\boldsymbol{\psi}}$ lies between $\boldsymbol{\psi}$ and $\hat{\boldsymbol{\psi}}_T$. By $\nabla_{\boldsymbol{\xi}\boldsymbol{\psi}}$ we denote the $(5N+p_\psi)$-dimensional column vector of first-order differential operators $(\nabla_{\boldsymbol{\xi}_1} \cdots \nabla_{\boldsymbol{\xi}_N} \; \nabla_{\boldsymbol{\psi}})'$. The first-order Taylor expansion of (B-13) around the true value of the parameters $(\boldsymbol{\xi}, \boldsymbol{\psi})$ is

$$\sqrt{T} \left(\left(\begin{array}{c} \hat{\boldsymbol{\xi}}_T \\ \hat{\boldsymbol{\psi}}_T \end{array} \right) - \left(\begin{array}{c} \boldsymbol{\xi} \\ \boldsymbol{\psi} \end{array} \right) \right) = \mathcal{B}_T^{-1} \mathcal{A}_T(\boldsymbol{\xi}, \boldsymbol{\psi}). \tag{B-14}$$

By the Law of Large Numbers, as $T \to \infty$, $\mathcal{A}_T(\boldsymbol{\xi}, \boldsymbol{\psi}) \xrightarrow{P} \mathbf{0}$ (a $(5N + p_\psi)$-dimensional column vector of zeros) and

$$\mathcal{B}_T \xrightarrow{P} -\mathrm{E}\left[\nabla_{\boldsymbol{\xi}\boldsymbol{\psi}} \left(\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\psi}, \hat{\phi}_{T\boldsymbol{\xi}}(z_t)) \right)' \right]. \tag{B-15}$$

This is enough to prove consistency.

From (B-15) the probability limit of $\mathcal{B}_T$ is a $(5N + p_\psi) \times (5N + p_\psi)$ matrix with generic

elements

$$\begin{aligned}
\mathcal{H}_{\boldsymbol{\xi}_i\,\boldsymbol{\xi}_j} &= -\mathrm{E}\left[\nabla_{\boldsymbol{\xi}_i\boldsymbol{\xi}_j}\boldsymbol{\ell}_j^m(\boldsymbol{\xi}_j,\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))\right], \\
\mathcal{H}_{\boldsymbol{\xi}_i\,\boldsymbol{\psi}} &= -\mathrm{E}\left[\nabla_{\boldsymbol{\xi}_i\boldsymbol{\psi}}\boldsymbol{\ell}_i^m(\boldsymbol{\xi}_i,\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))\right], \\
\mathcal{H}_{\boldsymbol{\psi}\,\boldsymbol{\xi}_i} &= -\mathrm{E}\left[\nabla_{\boldsymbol{\psi}\boldsymbol{\xi}_i}\boldsymbol{\ell}^c(\boldsymbol{\xi},\boldsymbol{\psi},\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))\right] \text{ and} \\
\mathcal{H}_{\boldsymbol{\psi}\,\boldsymbol{\psi}} &= -\mathrm{E}\left[\nabla_{\boldsymbol{\psi}\boldsymbol{\psi}}\boldsymbol{\ell}^c(\boldsymbol{\xi},\boldsymbol{\psi},\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))\right].
\end{aligned}$$

The matrix $\mathcal{H}_{\boldsymbol{\xi}_i\boldsymbol{\xi}_j} \neq 0$ only if $i = j$ and $\mathcal{H}_{\boldsymbol{\xi}_i\boldsymbol{\psi}} = 0$ for any $i = 1, \ldots, N$. Therefore, the probability limit of $\mathcal{B}_T$ is $\mathbf{H}$. Moreover this term is not affected by the presence of curve $\phi(z_t)$ (Newey and McFadden (1994)).

As $T \to \infty$, $\mathcal{A}_T(\boldsymbol{\xi},\boldsymbol{\psi})$ converges to a normal distribution with zero mean and a $(5N + p_\psi) \times (5N + p_\psi)$ variance covariance matrix $\mathbf{V}(\mathcal{A}_T(\boldsymbol{\xi},\boldsymbol{\psi}))$. In the case where there is no curve $\phi(z_t)$

$$\mathbf{V}(\mathcal{A}_T(\boldsymbol{\xi},\boldsymbol{\psi})) = \mathrm{E}\left[\mathbf{g}(\boldsymbol{\xi},\boldsymbol{\psi},\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))\mathbf{g}(\boldsymbol{\xi},\boldsymbol{\psi},\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t))'\right].$$

Given (B-10) and (B-11) in Lemma 1, $\mathbf{V}(\mathcal{A}_T(\boldsymbol{\xi},\boldsymbol{\psi}))$ is block diagonal and is equal to the Fisher information matrix in the fully parametric case $\mathbf{V}(\mathcal{A}_T(\boldsymbol{\xi},\boldsymbol{\psi})) = \mathbf{I}_{\boldsymbol{\xi}\,\boldsymbol{\psi}}$. In the presence of the curve $\phi(z_t)$ a correction term is needed and it is denoted by $\mathbf{W}_\phi$:

$$\mathcal{A}_T(\boldsymbol{\xi},\boldsymbol{\psi}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}}\right) \text{ where } \mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}} = \mathbf{I}_{\boldsymbol{\xi}\,\boldsymbol{\psi}} - \mathbf{W}_\phi \tag{B-16}$$

as $T \to \infty$. Combining (B-15) and (B-16), from (B-14) we obtain the asymptotic distribution of the parametric estimates as $T \to \infty$:

$$\sqrt{T}\left(\begin{pmatrix}\hat{\boldsymbol{\xi}}_T \\ \hat{\boldsymbol{\psi}}_T\end{pmatrix} - \begin{pmatrix}\boldsymbol{\xi} \\ \boldsymbol{\psi}\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{H}^{-1}\mathbf{I}^*_{\boldsymbol{\xi}\,\boldsymbol{\psi}}\mathbf{H}^{-1}\right).$$

This completes the proof. □

Proof of Proposition 2

We proof is done without loss of generality for $N = 2$.

The Fisher information when maximizing $\sum_{t=1}^T \boldsymbol{\ell}(\boldsymbol{\eta},\hat{\phi}_{T\,\boldsymbol{\eta}}(z_t))$ (see Theorem 2) can be written as

$$\mathbf{I}^{*\,f}_{\boldsymbol{\eta}} = \begin{pmatrix} \mathcal{I}^{*\,f}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_1} & \mathcal{I}^{*\,f}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2} & \mathcal{I}^{*\,f}_{\boldsymbol{\xi}_1\boldsymbol{\psi}} \\ \mathcal{I}^{*\,f}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_1} & \mathcal{I}^{*\,f}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_2} & \mathcal{I}^{*\,f}_{\boldsymbol{\xi}_2\boldsymbol{\psi}} \\ \mathcal{I}^{*\,f}_{\boldsymbol{\psi}\boldsymbol{\xi}_1} & \mathcal{I}^{*\,f}_{\boldsymbol{\psi}\boldsymbol{\xi}_2} & \mathcal{I}^{*\,f}_{\boldsymbol{\psi}\boldsymbol{\psi}} \end{pmatrix}$$

where $\mathcal{I}^{*\,f}_{\boldsymbol{\xi}_i\boldsymbol{\xi}_j} = \mathrm{E}\left[\left(\nabla_{\boldsymbol{\xi}_i}\boldsymbol{\ell}(\boldsymbol{\eta},\hat{\phi}_{T\,\boldsymbol{\eta}}(z_t))\right)\left(\nabla_{\boldsymbol{\xi}_j}\boldsymbol{\ell}(\boldsymbol{\eta},\hat{\phi}_{T\,\boldsymbol{\eta}}(z_t))\right)'\right]$ and the superscript f stands for the full information case.

If we estimate $\boldsymbol{\xi}$ by maximizing the marginal densities and $\boldsymbol{\psi}$ by maximizing the copula density (as in Theorem 4), the Fisher information matrix is $\mathbf{I}^*_{\boldsymbol{\xi}\boldsymbol{\psi}}$ and can be written as

$$\mathbf{I}^*_{\boldsymbol{\xi}\boldsymbol{\psi}} = \begin{pmatrix} \mathcal{I}^{*m}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_1} & \mathcal{I}^{*m}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2} & \mathbf{0} \\ \mathcal{I}^{*m}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_1} & \mathcal{I}^{*m}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}^{*c}_{\boldsymbol{\psi}\boldsymbol{\psi}} \end{pmatrix},$$

where the generic elements of th matrix are given in Theorem 4. We add the superscripts m when using only the marginals and c when using only the copula. We define the upper left block of $\mathbf{I}^*_{\boldsymbol{\xi}\boldsymbol{\psi}}$ as

$$\mathbf{I}^{*m}_{\boldsymbol{\xi}} = \begin{pmatrix} \mathcal{I}^{*m}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_1} & \mathcal{I}^{*m}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2} \\ \mathcal{I}^{*m}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_1} & \mathcal{I}^{*m}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_2} \end{pmatrix}.$$

We can also estimate all the parameters by maximizing only the copula part. In this case the Fisher information matrix is

$$\mathbf{I}^{*c}_{\boldsymbol{\xi}\boldsymbol{\psi}} = \begin{pmatrix} \mathcal{I}^{*c}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_1} & \mathcal{I}^{*c}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2} & \mathcal{I}^{*c}_{\boldsymbol{\xi}_1\boldsymbol{\psi}} \\ \mathcal{I}^{*c}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_1} & \mathcal{I}^{*c}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_2} & \mathcal{I}^{*c}_{\boldsymbol{\xi}_2\boldsymbol{\psi}} \\ \mathcal{I}^{*c}_{\boldsymbol{\psi}\boldsymbol{\xi}_1} & \mathcal{I}^{*c}_{\boldsymbol{\psi}\boldsymbol{\xi}_2} & \mathcal{I}^{*c}_{\boldsymbol{\psi}\boldsymbol{\psi}} \end{pmatrix}.$$

From Property 3.9 in Gourieroux and Monfort (1995)

$$\mathbf{I}^{*f}_{\boldsymbol{\eta}} = \begin{pmatrix} \mathbf{I}^{*m}_{\boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \mathbf{I}^{*c}_{\boldsymbol{\xi}\boldsymbol{\psi}}.$$

Therefore

$$\mathbf{I}^{*f}_{\boldsymbol{\eta}} \succeq \begin{pmatrix} \mathbf{I}^{*m}_{\boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{B-17}$$

the difference with $\mathbf{I}^{*f}_{\boldsymbol{\eta}}$ being $\mathbf{I}^{*c}_{\boldsymbol{\xi}\boldsymbol{\psi}}$, which is by definition a positive semidefinite matrix.

If we consider the marginal Fisher information $\mathbf{I}^{*f}_{\boldsymbol{\eta}}$ relative to $\boldsymbol{\xi}$ and we indicate it as $\mathbf{I}^{*f}_{\boldsymbol{\eta}|\boldsymbol{\xi}}$, we have (see e.g. Property 7.18 in Gourieroux and Monfort (1995))

$$\mathbf{I}^{*f}_{\boldsymbol{\eta}|\boldsymbol{\xi}} = \begin{pmatrix} \mathcal{I}^{*f}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_1} & \mathcal{I}^{*f}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2} \\ \mathcal{I}^{*f}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_1} & \mathcal{I}^{*f}_{\boldsymbol{\xi}_2\boldsymbol{\xi}_2} \end{pmatrix} - \begin{pmatrix} \mathcal{I}^{*f}_{\boldsymbol{\xi}_1\boldsymbol{\psi}} \\ \mathcal{I}^{*f}_{\boldsymbol{\xi}_2\boldsymbol{\psi}} \end{pmatrix} \left(\mathcal{I}^{*f}_{\boldsymbol{\psi}\boldsymbol{\psi}}\right)^{-1} \begin{pmatrix} \mathcal{I}^{*f}_{\boldsymbol{\psi}\boldsymbol{\xi}_1} & \mathcal{I}^{*f}_{\boldsymbol{\psi}\boldsymbol{\xi}_2} \end{pmatrix}. \tag{B-18}$$

Given two matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A} \succeq \mathbf{B}$, then, for any vector $\mathbf{P}$, we have $\mathbf{PAP}' \succeq \mathbf{PBP}'$. By setting

$$\mathbf{A} = \mathbf{I}^{*f}_{\boldsymbol{\eta}}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{I}^{*m}_{\boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \text{and} \quad \mathbf{P} = \left(\mathbf{I}\ , \ -\begin{pmatrix} \mathcal{I}^{*f}_{\boldsymbol{\xi}_1\boldsymbol{\psi}} \\ \mathcal{I}^{*f}_{\boldsymbol{\xi}_2\boldsymbol{\psi}} \end{pmatrix} \left(\mathcal{I}^{*f}_{\boldsymbol{\psi}\boldsymbol{\psi}}\right)^{-1} \right), \tag{B-19}$$

and from (B-17) and (B-19)

$$\mathbf{I}^{*f}_{\boldsymbol{\eta}|\boldsymbol{\xi}} \succeq \mathbf{I}^{*m}_{\boldsymbol{\xi}}. \tag{B-20}$$

An analogous result can proved when estimating $\boldsymbol{\psi}$, namely $\mathbf{I}^{*f}_{\boldsymbol{\eta}|\boldsymbol{\psi}} \succeq \mathcal{I}^{*c}_{\boldsymbol{\psi}\boldsymbol{\psi}}$.

Finally, in terms of asymptotic variances the efficiency loss is

$$\mathbf{H}^{-1}\begin{pmatrix}\mathbf{I}_{\boldsymbol{\xi}}^{*m} & \mathbf{0}\\ \mathbf{0} & \mathcal{I}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{*c}\end{pmatrix}\mathbf{H}^{-1'} \succeq \begin{pmatrix}\mathbf{I}_{\boldsymbol{\xi}}^{*m\,-1} & \mathbf{0}\\ \mathbf{0} & \mathcal{I}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{*c\,-1}\end{pmatrix} \succeq \begin{pmatrix}\mathbf{I}_{\boldsymbol{\eta}|\boldsymbol{\xi}}^{*f\,-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{\boldsymbol{\eta}|\boldsymbol{\psi}}^{*f\,-1}\end{pmatrix}.$$

The first inequality is straightforward and is due to the two–step estimation procedure, and the second inequality is in (B-20). This completes the proof. □

References

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.

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, **57**, 1047–1091.

Amado, C. and Teräsvirta, T. (2008). Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure. Technical report, SSE/EFI Working Paper Series in Economics and Finance. No. 691.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 579–625.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics*, **89**, 701–720.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, **70**, 191–221.

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

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.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realised kernels in practice: trades and quotes. *Econometrics Journal*, **12**, 1–32.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra–high frequency: Data handling concerns. *Computational Statistics and Data Analysis*, **51**, 2232–2245.

Brownlees, C. T. and Gallo, G. M. (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*, **8**, 29–56.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, **51**, 1305–1324.

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

Chiriac, R. and Voev, V. (2010). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*. forthcoming.

Cipollini, F., Engle, R. F., and Gallo, G. M. (2007). A model for multivariate non-negative valued processes in financial econometrics. Technical Report 2007/16.

Colacito, R., Engle, R., and Ghysels, E. (2010). A component model for dynamic correlations. 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.

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**(1-2), 29–58.

Engle, R., Ghysels, E., and Sohn, B. (2008). On the economic sources of stock market volatility. Technical report.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**, 425–446.

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. F. and Rangel, J. G. (2008). The spline-GARCH model for low frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, **21**, 1187–1222.

Feng, Y. (2006). A local dynamic conditional correlation model. Technical report, MPRA. Paper No. 1592.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, **82**, 540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004). The generalized dynamic factor model: consistency and rates. *Journal of Econometrics*, **119**, 231–255.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, **100**, 830–840.

Gagliardini, P. and Gourieroux, C. (2009). Efficiency in large dynamic panel models with common factor. Technical report, Swiss Finance Institute. Research Paper Series 09–12.

Gourieroux, C. and Monfort, A. (1995). *Statistics and Econometric Models: General Concepts, Estimation, Prediction and Algorithms*, volume 1. Cambridge University Press.

Hafner, C. M. and Linton, O. (2009). Efficient estimation of a multivariate multiplicative volatility model. Technical report, Institut de Statistique. DP 0905.

Hansen, P. R. and Lunde, A. (2010). Estimating the persistence and the autocorrelation function of a time series that is measured with error. Technical report.

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.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, **94**, 401–419.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (2000). *Continuous Multivariate Distributions*, volume 1. John Wiley & Sons, New York.

Long, X., Su, L., and Ullah, A. (2009). Estimation and forecasting of dynamic conditional covariance: A semiparametric multivariate model. Technical report.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, **53**, 61–65.

Patton, A. (2010). Volatility Forecast Comparison using Imperfect Volatility Proxies. *Journal of Econometrics*. forthcoming.

Patton, A. J. and Sheppard, K. (2009). Good volatility, bad volatility: Signed jumps and the persistence of volatility. Technical report.

Rangel, J. G. and Engle, R. (2009). The factor-spline-garch model for high and low frequency correlations. Technical report, Banco de México. N 2009-03.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, **20**, 1768–1802.

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

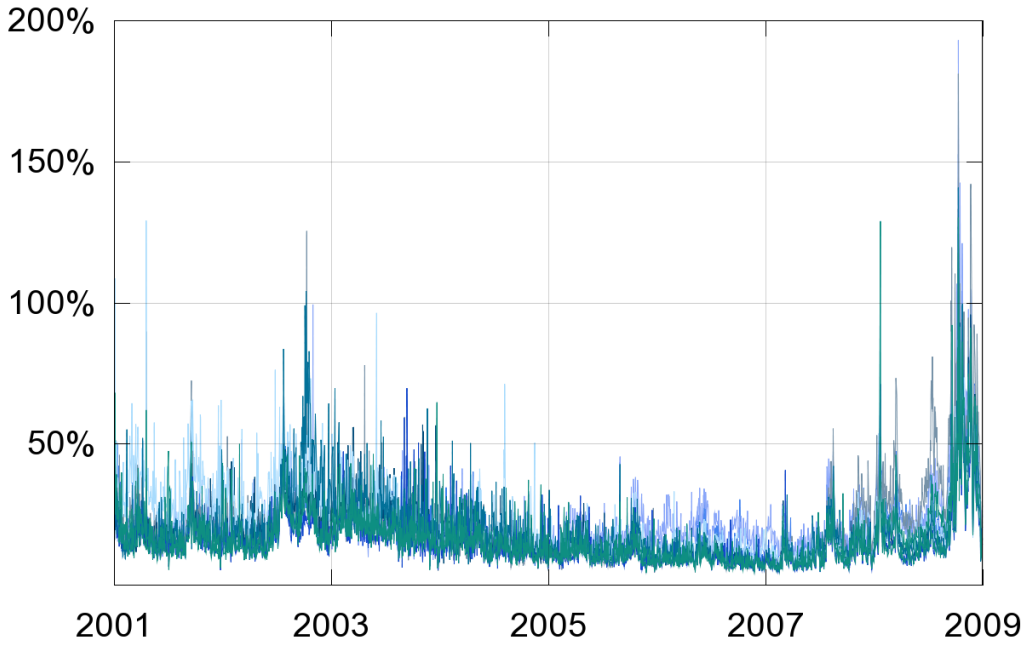
Staniswalis, J. G. (1987). A weighted likelihood formulation for kernel estimators of a regression function with biomedical applications. Technical report, Virginia Commonwealth University. Techincal Report n.5.

Staniswalis, J. G. (1989). On the kernel estimate of a regression function in likelihood based models. *Journal of the American Statistical Association*, **84**, 273–283.

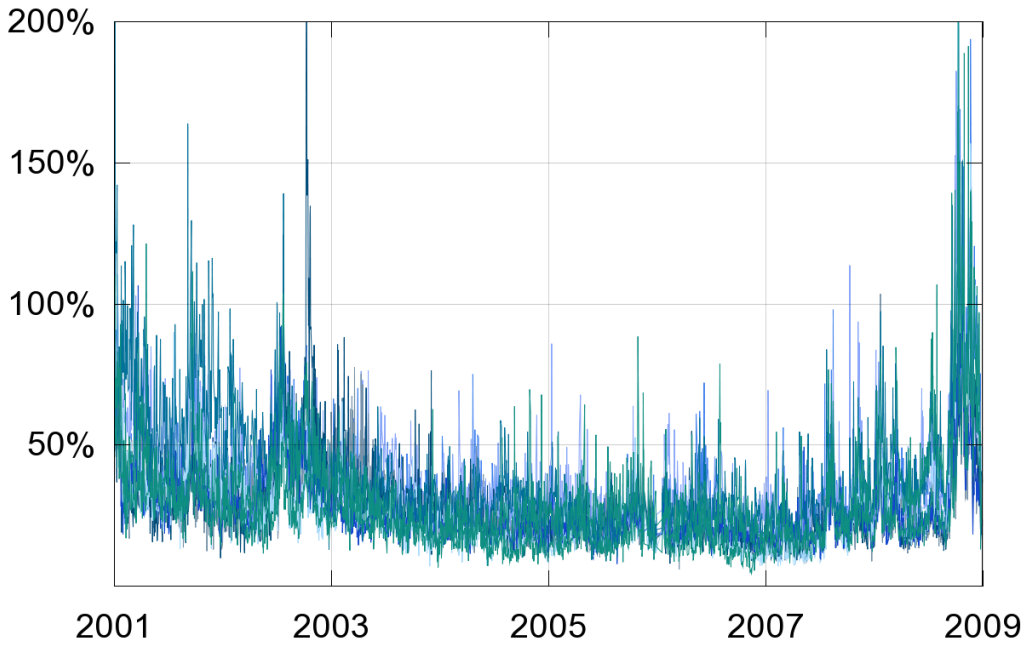
Veredas, D., Rodriguez-Poo, J., and A., E. (2007). *Semiparametric estimation for financial durations*, pages 204–208. Springer Verlag.

Tables and Figures

Figure 1: Realized volatilities



(a) SPDR



(b) S&P100

Annualized realized volatilities for the SPDR select sectors (top) and the 90 S&P100 constituents (bottom).

Table 1: S&P100 constituents

Ticker	Name	Sector
AA	Alcoa Inc	Materials
AAPL	Apple Inc.	Information Technology
ABT	Abbott Labs	Health Care
AEP	American Electric Power	Utilities
ALL	Allstate Corp.	Financials
AMGN	Amgen	Health Care
AMZN	Amazon Corp.	Consumer Discretionary
AVP	Avon Products	Consumer Staples
AXP	American Express	Financials
BA	Boeing Company	Industrials
BAC	Bank of America Corp.	Financials
BAX	Baxter International Inc.	Health Care
BHI	Baker Hughes	Energy
BK	Bank of New York Mellon Corp.	Financials
BMY	Bristol-Myers Squibb	Health Care
BNI	Burlington Northern Santa Fe C	Industrials
CAT	Caterpillar Inc.	Industrials
C	Citigroup Inc.	Financials
CL	Colgate-Palmolive	Consumer Staples
CMCSA	Comcast Corp.	Consumer Discretionary
COF	Capital One Financial	Financials
COST	Costco Co.	Consumer Staples
CPB	Campbell Soup	Consumer Staples
CSCO	Cisco Systems	Information Technology
CVS	CVS Caremark Corp.	Consumer Staples
CVX	Chevron Corp.	Energy
DD	Du Pont (E.I.)	Materials
DELL	Dell Inc.	Information Technology
DIS	Walt Disney Co.	Consumer Discretionary
DOW	Dow Chemical	Materials
DVN	Devon Energy Corp.	Energy
EMC	EMC Corp.	Information Technology
ETR	Entergy Corp.	Utilities
EXC	Exelon Corp.	Utilities
FDX	FedEx Corporation	Industrials
F	Ford Motor	Consumer Discretionary
GD	General Dynamics	Industrials
GE	General Electric	Industrials
GILD	Gilead Sciences	Health Care
GS	Goldman Sachs Group	Financials
HAL	Halliburton Co.	Energy
HD	Home Depot	Consumer Discretionary
HNZ	Heinz (H.J.)	Consumer Staples
HON	Honeywell Int'l Inc.	Industrials
HPQ	Hewlett-Packard	Information Technology

(cont.)

(cont.)

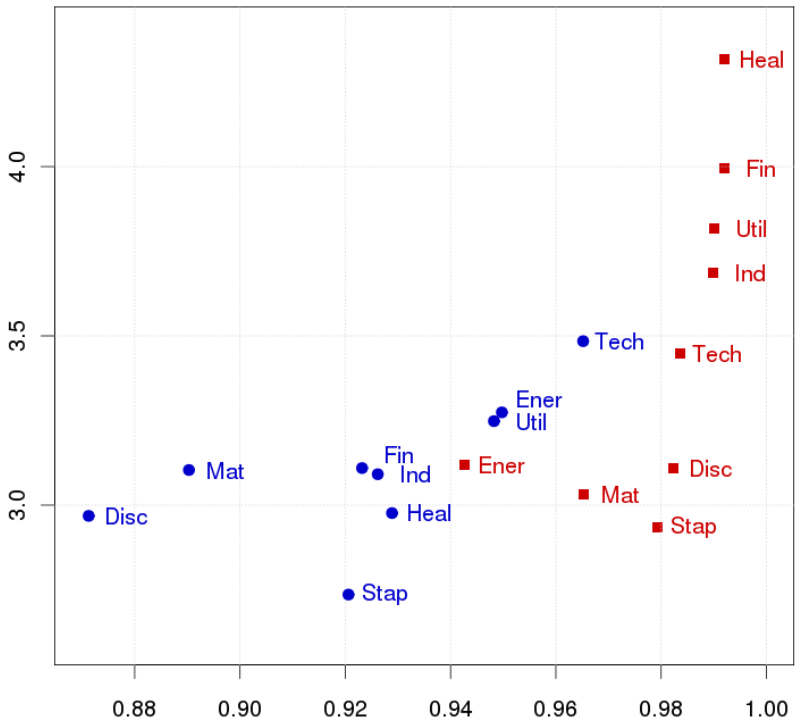
IBM	International Bus. Machines	Information Technology
INTC	Intel Corp.	Information Technology
JNJ	Johnson & Johnson	Health Care
JPM	JPMorgan Chase & Co.	Financials
KO	Coca Cola Co.	Consumer Staples
LMT	Lockheed Martin Corp.	Industrials
LOW	Lowe's Cos.	Consumer Discretionary
MCD	McDonald's Corp.	Consumer Discretionary
MDT	Medtronic Inc.	Health Care
MMM	3M Company	Industrials
MO	Altria Group, Inc.	Consumer Staples
MRK	Merck & Co.	Health Care
MSFT	Microsoft Corp.	Information Technology
MS	Morgan Stanley	Financials
NKE	NIKE Inc.	Consumer Discretionary
NSC	Norfolk Southern Corp.	Industrials
ORCL	Oracle Corp.	Information Technology
OXY	Occidental Petroleum	Energy
PEP	PepsiCo Inc.	Consumer Staples
PFE	Pfizer, Inc.	Health Care
PG	Procter & Gamble	Consumer Staples
QCOM	QUALCOMM Inc.	Information Technology
RF	Regions Financial Corp.	Financials
SGP	Schering-Plough	Health Care
SLB	Schlumberger Ltd.	Energy
SLE	Sara Lee Corp.	Consumer Staples
SO	Southern Co.	Utilities
S	Sprint Nextel Corp.	Telecommunications Services
T	AT&T Inc.	Telecommunications Services
TGT	Target Corp.	Consumer Discretionary
TWX	Time Warner Inc.	Consumer Discretionary
TXN	Texas Instruments	Information Technology
TYC	Tyco International	Industrials
UNH	UnitedHealth Group Inc.	Health Care
UPS	United Parcel Service	Industrials
USB	U.S. Bancorp	Financials
UTX	United Technologies	Industrials
VZ	Verizon Communications	Telecommunications Services
WAG	Walgreen Co.	Consumer Staples
WFC	Wells Fargo	Financials
WMB	Williams Cos.	Energy
WMT	Wal-Mart Stores	Consumer Staples
WY	Weyerhaeuser Corp.	Materials
XOM	Exxon Mobil Corp.	Energy
XRX	Xerox Corp.	Information Technology

Table 2: Descriptive statistics

		vol	vov	$\hat{\rho}_{\mathsf{day}}$	$\hat{\rho}_{\mathsf{week}}$	$\hat{\rho}_{\mathsf{month}}$	$\bar{\rho}$	PC_1
				SPDR				
Mat		23.34	11.83	0.72	0.63	0.39	0.82	0.91
Ener		25.68	12.78	0.65	0.61	0.35	0.79	0.87
Fin		25.85	15.71	0.69	0.50	0.35	0.77	0.84
Ind		21.85	11.29	0.66	0.57	0.37	0.81	0.88
Tech		26.82	13.34	0.49	0.41	0.27	0.66	0.58
Stap		16.93	8.21	0.45	0.36	0.19	0.77	0.77
Util		24.12	12.73	0.65	0.55	0.35	0.72	0.70
Heal		17.27	8.61	0.34	0.27	0.17	0.73	0.68
Disc		20.56	10.57	0.64	0.55	0.35	0.82	0.90
				S&P100				
Mat	$q_{0.25}$	33.80	14.49	0.66	0.55	0.31	0.63	0.62
	$q_{0.50}$	34.73	15.17	0.69	0.56	0.34	0.67	0.72
	$q_{0.75}$	36.42	16.59	0.73	0.60	0.37	0.72	0.83
Ener	$q_{0.25}$	34.00	15.04	0.57	0.50	0.26	0.61	0.61
	$q_{0.50}$	37.98	16.36	0.66	0.57	0.32	0.68	0.72
	$q_{0.75}$	44.28	18.50	0.71	0.66	0.38	0.69	0.74
Fin	$q_{0.25}$	36.45	21.01	0.63	0.27	0.17	0.52	0.49
	$q_{0.50}$	39.77	23.77	0.66	0.47	0.28	0.64	0.70
	$q_{0.75}$	42.93	26.59	0.74	0.56	0.34	0.68	0.76
Ind	$q_{0.25}$	29.25	12.42	0.61	0.49	0.31	0.65	0.64
	$q_{0.50}$	30.61	13.44	0.64	0.54	0.33	0.69	0.73
	$q_{0.75}$	33.19	17.33	0.70	0.56	0.36	0.71	0.80
Tech	$q_{0.25}$	34.96	16.10	0.64	0.52	0.34	0.56	0.49
	$q_{0.50}$	38.53	17.85	0.67	0.58	0.37	0.62	0.58
	$q_{0.75}$	46.20	23.26	0.72	0.61	0.40	0.72	0.78
Util	$q_{0.25}$	28.24	13.58	0.67	0.45	0.25	0.60	0.54
	$q_{0.50}$	29.61	14.09	0.70	0.52	0.30	0.67	0.66
	$q_{0.75}$	31.42	15.20	0.72	0.60	0.34	0.70	0.73
Stap	$q_{0.25}$	25.46	11.43	0.46	0.36	0.22	0.60	0.52
	$q_{0.50}$	28.32	12.45	0.53	0.43	0.24	0.68	0.69
	$q_{0.75}$	29.80	13.30	0.58	0.47	0.30	0.71	0.76
Heal	$q_{0.25}$	29.19	13.02	0.48	0.34	0.22	0.64	0.62
	$q_{0.50}$	30.06	13.55	0.59	0.46	0.29	0.65	0.63
	$q_{0.75}$	32.95	15.00	0.61	0.50	0.34	0.66	0.67
Disc	$q_{0.25}$	32.98	15.49	0.58	0.47	0.30	0.63	0.60
	$q_{0.50}$	35.76	17.41	0.64	0.54	0.36	0.68	0.68
	$q_{0.75}$	41.13	21.01	0.67	0.58	0.41	0.70	0.76

The top part of the table shows descriptive statistics for SPDR. The bottom part of the table shows the same descriptive statistics for S&P100 with the assets grouped according to the same sectors as SPDR. For each group the table shows the 25, 50 and 75 quantiles. The columns report the average annualized volatility (vol), standard deviation of volatility a.k.a. volatility of volatility (vov), the autocorrelations of order 1, 5 and 22 ($\hat{\rho}_{\mathsf{day}}$, $\hat{\rho}_{\mathsf{week}}$ and $\hat{\rho}_{\mathsf{month}}$), the average cross-correlation with the other series in the dataset ($\bar{\rho}$), and the percentage of the variance explained by the first principal component (PC_1).

Figure 2: SPDR Persistence versus volatility



Scatter plots of persistence (X-axis) versus unconditional volatility (Y-axis). The circles are for our model and the squares for univariate MEMs

Table 3: Detailed iterative procedure

Initialization

For any $i = 1, \dots, N$ set $(\boldsymbol{\xi}_1^0, \dots, \boldsymbol{\xi}_N^0)$ where $\boldsymbol{\xi}_i^0 = (a_i^0, \alpha_i^0, \beta_i^0, \gamma_i^0, \nu_i^0)$, and $\mathrm{K}(x) = \frac{15}{16}(1-x^2)^2 \mathbf{1}_{|x|\leq 1}(x)$.

For any $t_0, t \in [0,T]$, define $z_0 = t_0/T$ and $z_t = t/T$ and compute an initial estimate of the systematic risk

$$\phi^0(z_0) = \frac{\sum_{t=1}^T \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \frac{1}{N} \sum_{i=1}^N \frac{x_{i\,t}}{\bar{x}_i} \frac{1/s_i^2}{\sum_{i=1}^N 1/s_k^2}}{\sum_{t=1}^T \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right)}, \qquad z_0 \in [0,1].$$

where $\bar{x}_i$ and s_i^2 are, respectively, sample mean and variance of the i-th series.
For any $i = 1, \dots, N$ compute the initial idiosyncratic risks
$\mu_{i\,t}^0 = \left(1 - \alpha_i^0 - \beta_i^0 - \frac{\gamma_i^0}{2}\right) + \alpha_i^0 \frac{x_{i\,t}}{a_i^0 \phi^0(z_{t-1})} + \beta_i^0 \mu_{i\,t-1}^0 + \gamma_i^0 \frac{x_{i\,t}}{a_i^0 \phi^0(z_{t-1})} \mathbf{1}_{r_{,i\,t-1}<0}$.

For iteration $q \geq 0$

1 For any $i = 1, \dots, N$ maximize the N un–smoothed log-likelihoods

$\max_{\boldsymbol{\xi}_i} \sum_{t=1}^T \log f_{x_i}(x_{i\,t} | \mathcal{F}_{t-1}; \boldsymbol{\xi}_i, \phi^q(z_t))$ and get $\boldsymbol{\xi}_i^{q+1} = (a_i^{q+1}, \alpha_i^{q+1}, \beta_i^{q+1}, \gamma_i^{q+1}, \nu_i^{q+1})$

2.1 **Intermediate risks** Compute the N idiosyncratic risks evaluated at the systematic risk of the previous iteration:
$\tilde{\mu}_{i\,t}^{q+1} = \left(1 - \alpha_i^{q+1} - \beta_i^{q+1} - \frac{\gamma_i^{q+1}}{2}\right) + \alpha_i^{q+1} \frac{x_{i\,t}}{a_i^{q+1} \phi^q(z_{t-1})} + \beta_i^{q+1} \tilde{\mu}_{i\,t-1}^{q+1} + \gamma_i^{q+1} \frac{x_{i\,t}}{a_i^{q+1} \phi^q(z_{t-1})} \mathbf{1}_{r_{i\,t-1}<0}$.

and, for any $z_0 \in [0,1]$, compute the systematic risk given $\tilde{\mu}_{i\,t}^{q+1}$: $\tilde{\phi}^{q+1}(z_0) = \frac{\sum_{t=1}^T \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \sum_{i=1}^N \frac{x_{i\,t}}{a_i^{q+1} \tilde{\mu}_{i\,t}^{q+1}} \frac{\nu_i}{\sum_{i=1}^N \nu_i}}{\sum_{t=1}^T \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right)}$.

2.2 **Updated risks** Compute the N idiosyncratic risks evaluated at the systematic risk $\tilde{\phi}^{q+1}(z_t)$:
$\mu_{i\,t}^{q+1} = \left(1 - \alpha_i^{q+1} - \beta_i^{q+1} - \frac{\gamma_i^{q+1}}{2}\right) + \alpha_i^{q+1} \frac{x_{i\,t}}{a_i^{q+1} \tilde{\phi}^{q+1}(z_{t-1})} + \beta_i^{q+1} \mu_{i\,t-1}^{q+1} + \gamma_i^{q+1} \frac{x_{i\,t}}{a_i^{q+1} \tilde{\phi}^{q+1}(z_{t-1})} \mathbf{1}_{r_{i\,t-1}<0}$.

and, for any $z_0 \in [0,1]$, compute the systematic risk given $\mu_{i\,t}^{q+1}$: $\phi^{q+1}(z_0) = \frac{\sum_{t=1}^T \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right) \sum_{i=1}^N \frac{x_{i\,t}}{a_i^{q+1} \mu_{i\,t}^{q+1}} \frac{\nu_i}{\sum_{i=1}^N \nu_i}}{\sum_{t=1}^T \mathrm{K}\left(\frac{z_0 - z_t}{h_T}\right)}$.

3 $q = q + 1$ and go back to 1.

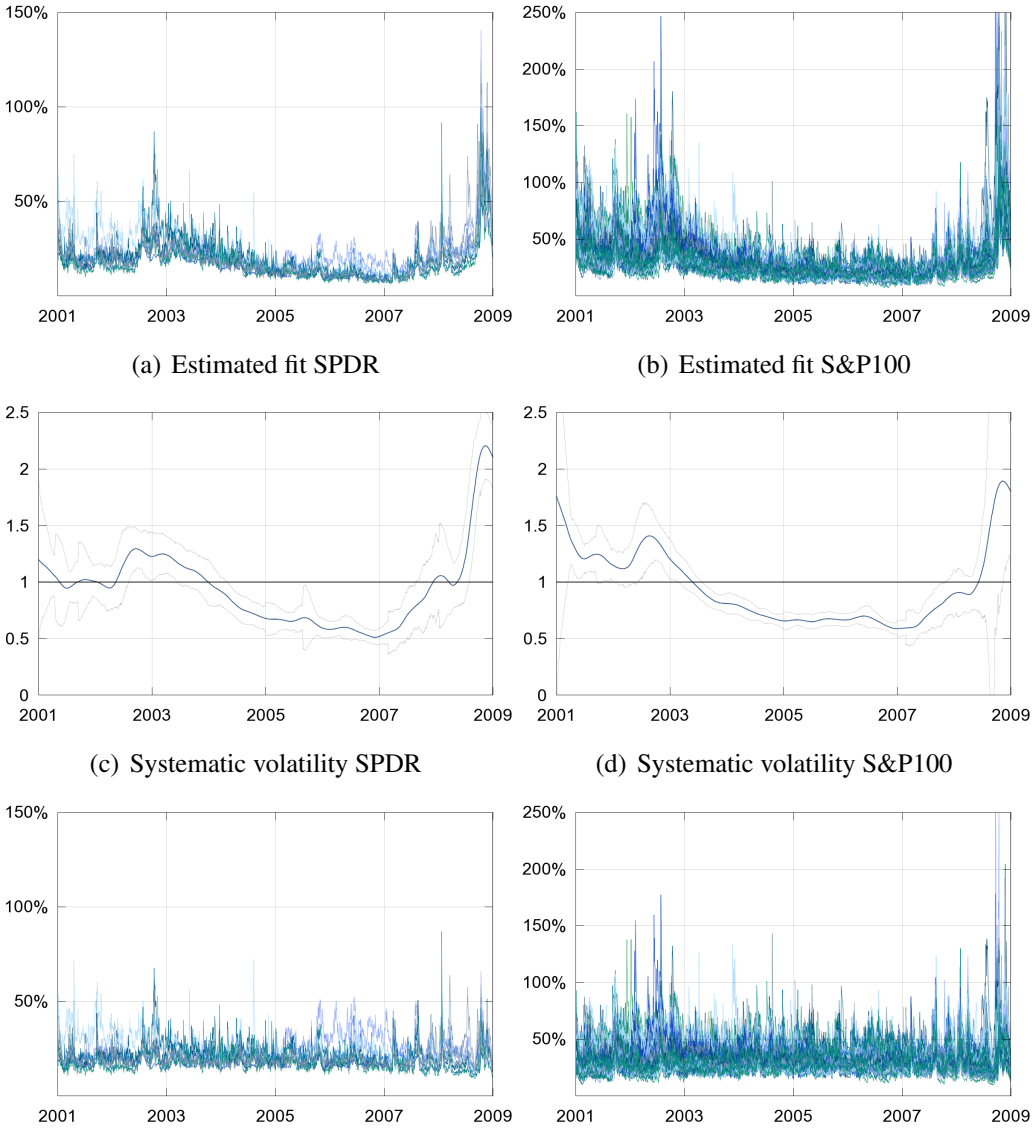
Stop when convergence of $\boldsymbol{\xi}_i^{q+1}$ is reached and run 2.1 and 2.2 for a last time.

Table 4: Estimated parameters

		Our Model							MEM						
		α_i	β_i	γ_i	π_i	a_i	ν_i	QLL	α_i	β_i	γ_i	π_i	a_i	ν_i	QLL
							SPDR								
Mat		0.24	0.59	0.15	0.90	22.3	0.34	1.61	0.25	0.66	0.13	0.97	20.7	0.33	1.72
Ener		0.26	0.64	0.12	0.96	27.0	0.34	1.85	0.25	0.64	0.12	0.95	22.6	0.34	1.97
Fin		0.30	0.56	0.14	0.93	22.5	0.63	1.52	0.21	0.73	0.13	0.99	54.3	0.59	1.61
Ind		0.27	0.60	0.13	0.93	22.3	0.47	1.44	0.28	0.66	0.10	0.99	39.9	0.48	1.54
Tech		0.28	0.62	0.15	0.97	33.4	0.71	1.90	0.25	0.68	0.12	0.99	31.4	0.69	1.99
Stap		0.10	0.81	0.03	0.93	15.5	0.95	1.08	0.19	0.74	0.10	0.99	18.8	0.75	1.15
Util		0.26	0.64	0.11	0.96	26.1	0.57	1.65	0.27	0.69	0.07	0.99	45.5	0.66	1.73
Heal		0.36	0.51	0.13	0.94	20.0	1.21	1.07	0.37	0.57	0.12	0.99	74.9	1.24	1.16
Disc		0.20	0.60	0.15	0.88	19.5	0.47	1.34	0.24	0.68	0.12	0.99	22.4	0.47	1.44
							S&P100								
Mat	$q_{0.25}$	0.29	0.57	0.09	0.92	34.10	0.26	2.42	0.27	0.61	0.09	0.97	30.93	0.25	2.57
	$q_{0.50}$	0.30	0.58	0.10	0.92	35.95	0.28	2.50	0.29	0.64	0.10	0.97	33.55	0.26	2.65
	$q_{0.75}$	0.31	0.59	0.11	0.93	37.62	0.31	2.59	0.31	0.67	0.10	0.97	35.40	0.30	2.74
Ener	$q_{0.25}$	0.24	0.61	0.09	0.95	32.93	0.22	2.42	0.24	0.65	0.08	0.96	30.30	0.21	2.57
	$q_{0.50}$	0.26	0.64	0.10	0.96	40.75	0.27	2.72	0.25	0.68	0.08	0.97	35.30	0.25	2.87
	$q_{0.75}$	0.28	0.68	0.12	0.98	46.35	0.45	3.00	0.28	0.69	0.09	0.98	41.90	0.41	3.16
Fin	$q_{0.25}$	0.30	0.50	0.13	0.94	32.38	0.28	2.21	0.31	0.56	0.10	0.99	33.50	0.26	2.33
	$q_{0.50}$	0.34	0.54	0.15	0.95	35.05	0.31	2.30	0.34	0.59	0.12	0.99	40.95	0.30	2.44
	$q_{0.75}$	0.37	0.60	0.16	0.96	41.52	0.45	2.50	0.37	0.64	0.13	0.99	48.23	0.44	2.65
Ind	$q_{0.25}$	0.29	0.55	0.08	0.93	30.10	0.25	2.13	0.28	0.59	0.08	0.97	29.30	0.25	2.28
	$q_{0.50}$	0.32	0.58	0.10	0.94	31.00	0.28	2.22	0.33	0.62	0.09	0.98	30.20	0.28	2.37
	$q_{0.75}$	0.34	0.61	0.12	0.95	37.40	0.35	2.48	0.34	0.64	0.11	0.98	32.90	0.32	2.62
Tech	$q_{0.25}$	0.27	0.48	0.08	0.92	33.25	0.27	2.43	0.26	0.56	0.08	0.97	31.75	0.22	2.59
	$q_{0.50}$	0.30	0.60	0.10	0.94	39.60	0.31	2.62	0.29	0.64	0.09	0.98	36.70	0.25	2.77
	$q_{0.75}$	0.36	0.64	0.13	0.95	45.15	0.47	2.98	0.37	0.68	0.11	0.99	43.50	0.41	3.11
Stap	$q_{0.25}$	0.18	0.57	0.05	0.91	25.80	0.33	1.93	0.23	0.61	0.05	0.97	25.50	0.33	2.06
	$q_{0.50}$	0.30	0.59	0.06	0.93	29.80	0.42	2.00	0.29	0.65	0.06	0.98	28.10	0.42	2.15
	$q_{0.75}$	0.32	0.77	0.09	0.97	32.50	0.69	2.31	0.32	0.72	0.08	0.98	29.80	0.56	2.41
Util	$q_{0.25}$	0.27	0.60	0.10	0.95	30.05	0.32	2.00	0.27	0.63	0.08	0.98	29.30	0.32	2.14
	$q_{0.50}$	0.28	0.62	0.10	0.96	31.25	0.35	2.11	0.28	0.66	0.09	0.98	31.00	0.35	2.25
	$q_{0.75}$	0.30	0.65	0.11	0.97	33.20	0.37	2.23	0.29	0.68	0.10	0.99	31.97	0.37	2.36
Heal	$q_{0.25}$	0.25	0.56	0.05	0.93	31.35	0.35	2.16	0.27	0.57	0.04	0.97	30.10	0.36	2.30
	$q_{0.50}$	0.28	0.63	0.07	0.96	33.40	0.47	2.22	0.31	0.64	0.08	0.98	31.90	0.45	2.36
	$q_{0.75}$	0.34	0.67	0.11	0.97	38.30	0.75	2.39	0.35	0.68	0.09	0.99	35.85	0.66	2.52
Disc	$q_{0.25}$	0.29	0.52	0.09	0.90	33.88	0.29	2.32	0.25	0.63	0.07	0.98	31.60	0.28	2.46
	$q_{0.50}$	0.31	0.57	0.10	0.92	35.25	0.37	2.49	0.28	0.66	0.09	0.98	33.35	0.33	2.63
	$q_{0.75}$	0.34	0.61	0.10	0.95	39.10	0.41	2.69	0.30	0.68	0.09	0.99	37.38	0.37	2.83

Estimated parameters of our model (left) and the univariate MEM (right). $\pi_i = \alpha_i + \beta_i + \frac{1}{2}\gamma_i$ is a measure of persistence. QLL stands for the Quasi Log-likelihood Loss function.

Figure 3: Volatility decomposition



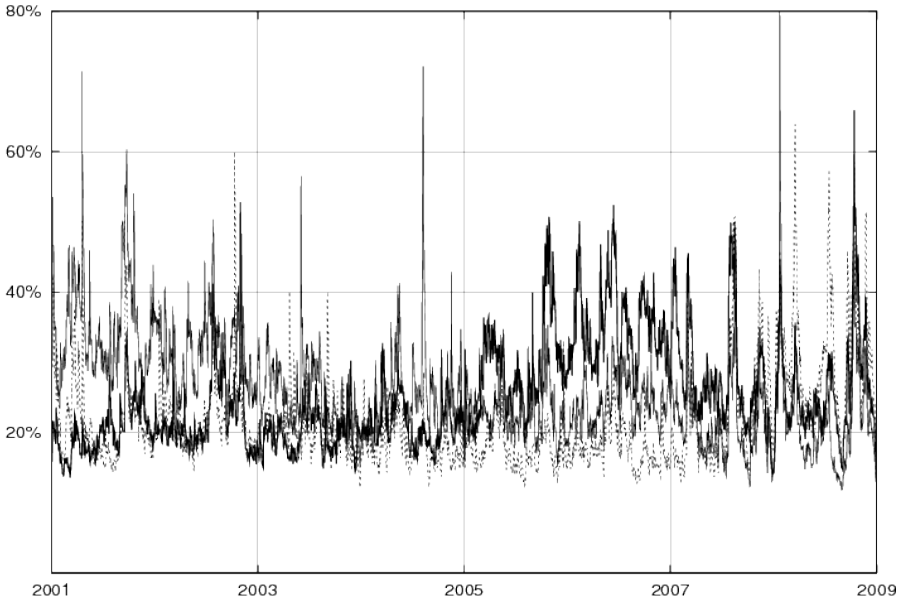
(a) Estimated fit SPDR (b) Estimated fit S&P100

(c) Systematic volatility SPDR (d) Systematic volatility S&P100

(e) Idiosyncratic volatilities SPDR (f) Idiosyncratic volatilities S&P100

The top row shows the estimated fit of the annualized volatilities entailed by the model $\sqrt{252\ \hat{a}_{T\,i}\ \hat{\phi}_{T\,\boldsymbol{\xi}}(z_t)\ \hat{\mu}_{i\,t}}$, the middle row shows the systematic volatility $\sqrt{\hat{\phi}_{T\,\boldsymbol{\xi}}(z_t)}$, and the bottom row shows the annualized idiosyncratic volatilities $\sqrt{252\ \hat{a}_{T\,i}\ \hat{\mu}_{i\,t}}$. Left column is for SPDR and right column for S&P100.

Figure 4: Idiosyncratic volatilities for Energy, Financial and Technology



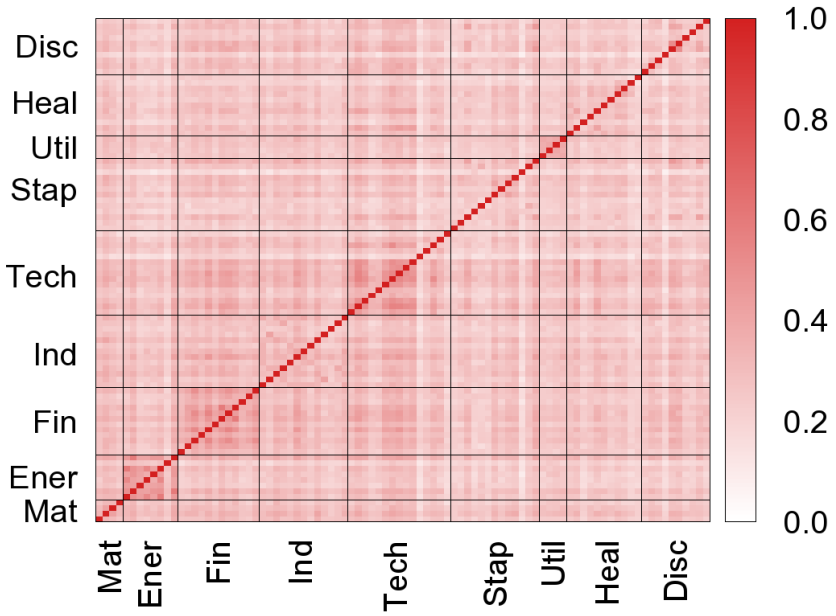
The thick solid line shows the idiosyncratic volatility for Energy, the dashed for Financial and the thin solid for Technology.

Table 5: SPDR Depependence among innovations

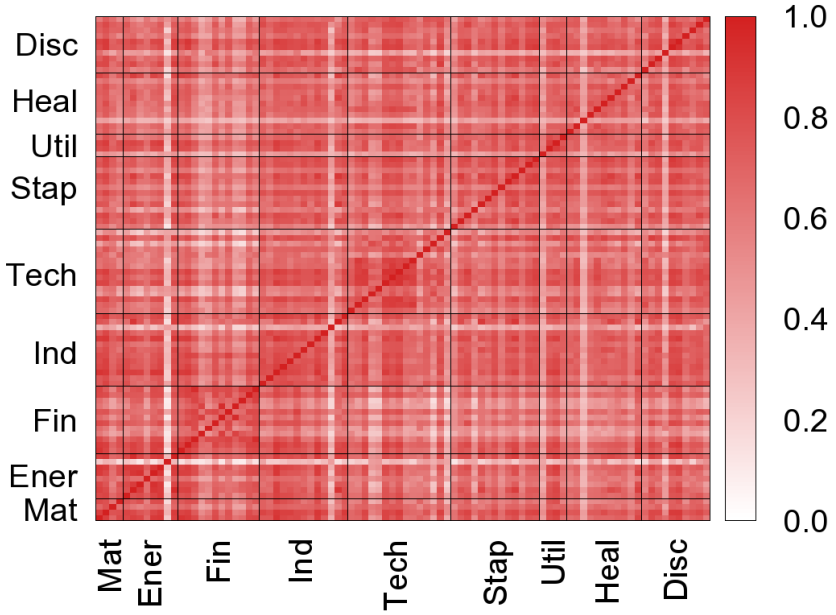
	Mat	Ener	Fin	Ind	Tech	Stap	Util	Heal	Disc
Mat	1.00								
Ener	0.41	1.00							
Fin	0.40	0.39	1.00						
Ind	0.43	0.40	0.47	1.00					
Tech	0.41	0.35	0.44	0.44	1.00				
Stap	0.39	0.31	0.40	0.40	0.37	1.00			
Util	0.34	0.33	0.36	0.38	0.33	0.33	1.00		
Heal	0.38	0.35	0.41	0.45	0.44	0.42	0.37	1.00	
Disc	0.44	0.39	0.50	0.47	0.43	0.43	0.36	0.46	1.00
$\bar{\mathbf{R}}_i$	0.40	0.37	0.42	0.43	0.40	0.38	0.35	0.41	0.43
$\%\bar{\mathbf{R}}_i/\bar{\rho}_i$	48.7	46.8	54.5	53.1	60.6	49.3	48.6	56.1	52.4

The top panel shows the dependencies among variance innovations estimated as the correlation matrix $\mathbf{R}$. As for the bottom panel, the row $\bar{\mathbf{R}}_i$ shows the cross–sectional average dependence among innovations. The row $\bar{\mathbf{R}}_i/\bar{\rho}_i$ shows the percentage difference between the cross–sectional average dependence and the cross–sectional average sample correlation of Table 2.

Figure 5: S&P100 Dependence among realized variances and their innovations



(a) Innovations



(b) Realized Variances

The heat map is a diagrammatic representation of the dependencies among innovations (top plot) and realized variances (bottom plot) for the 90 constituents. The assets are sorted according to the same sector classification as SPDR. The average value for the realized variances is 0.78 and 0.26 for the innovations.

Table 6: Explanatory power of the idiosyncratic volatilities

	SPDR								
	Mat	Ener	Fin	Ind	Tech	Stap	Util	Heal	Disc
	0.30	0.28	0.14	0.28	0.28	0.33	0.28	0.31	0.30
	S&P100								
	Mat	Ener	Fin	Ind	Tech	Stap	Util	Heal	Disc
$q_{0.25}$	0.48	0.46	0.17	0.48	0.41	0.49	0.45	0.50	0.44
$q_{0.50}$	0.49	0.48	0.28	0.49	0.44	0.50	0.46	0.50	0.46
$q_{0.75}$	0.49	0.50	0.35	0.49	0.46	0.52	0.46	0.51	0.48

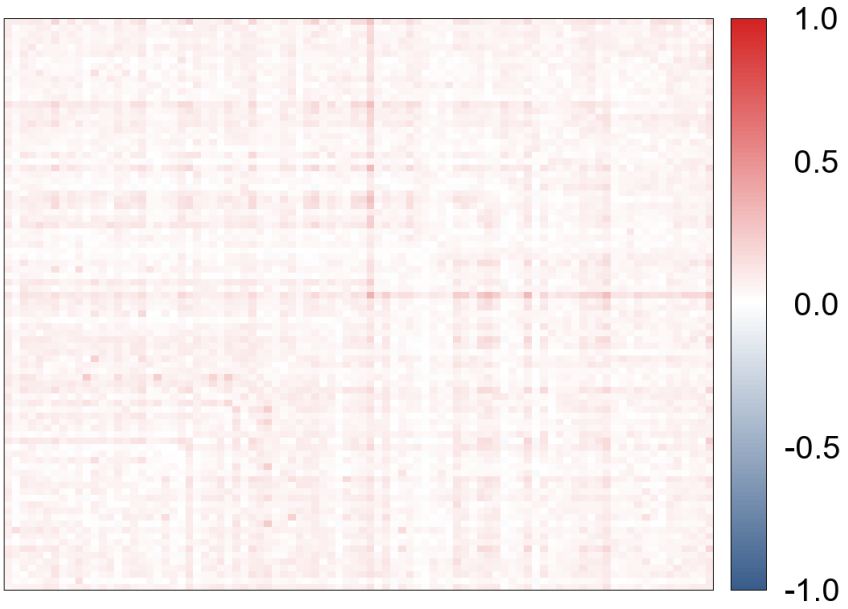
The top panel shows the explanatory power of the idiosyncratic volatilities for SPDR. The bottom shows the minimum and maximum explanatory powers of the idiosyncratic volatilities for S&P100 as well as the 25, 50 and 75 quantiles.

Table 7: SPDR residual diagnostics.

	$\hat{\rho}_{\mathsf{day}}$	$\hat{\rho}_{\mathsf{week}}$	$\hat{\rho}_{\mathsf{month}}$	Q_{day}	Q_{week}	Q_{month}	$\bar{\rho}_{\mathsf{day}}$	$\bar{\rho}_{\mathsf{week}}$	$\bar{\rho}_{\mathsf{month}}$
Mat	-0.00	0.00	0.03	0.86	0.86	0.20	0.01	0.00	0.00
Ener	-0.01	0.02	0.02	0.58	0.19	0.05	0.00	0.00	-0.00
Fin	-0.00	-0.02	0.01	0.92	0.01	0.19	0.00	-0.00	0.00
Ind	-0.03	-0.01	0.01	0.15	0.75	0.35	0.00	-0.00	0.00
Tech	-0.01	-0.06	-0.00	0.66	0.04	0.27	0.01	-0.00	-0.00
Stap	0.04	0.02	-0.02	0.05	0.11	0.09	0.00	0.00	0.00
Util	-0.01	-0.03	0.00	0.70	0.05	0.07	0.01	0.00	0.00
Heal	-0.03	-0.01	0.00	0.17	0.21	0.78	0.00	-0.00	0.00
Disc	-0.01	0.00	0.00	0.55	0.75	0.27	0.00	0.00	0.00
	Lag one correlation matrix								
	Mat	Ener	Fin	Ind	Tech	Stap	Util	Heal	Disc
Mat	-0.01								
Ener	0.02	-0.01							
Fin	-0.02	0.02	-0.01						
Ind	0.03	0.02	0.01	-0.03					
Tech	0.03	-0.02	0.01	0.01	-0.01				
Stap	0.04	0.04	0.02	0.05	0.05	0.03			
Util	0.02	0.01	0.01	0.03	0.01	0.04	0.01		
Heal	0.00	0.00	0.01	0.02	0.01	0.03	0.03	-0.04	
Disc	0.04	0.04	0.03	0.00	0.06	0.02	0.07	0.03	-0.02

The top part shows the autocorrelations of order 1, 5 and 22 ($\hat{\rho}_{\mathsf{day}}$, $\hat{\rho}_{\mathsf{week}}$ and $\hat{\rho}_{\mathsf{month}}$), the p-values of the Ljung-Box tests for autocorrelations of order 1, 5 and 22 (Q_{day}, Q_{week}, Q_{month}), and the average lagged cross-correlations of order 1, 5 and 22 ($\bar{\rho}_{\mathsf{day}}$, $\bar{\rho}_{\mathsf{week}}$, $\bar{\rho}_{\mathsf{month}}$). The bottom part of the table shows the lagged one correlation matrix of the residuals.

Figure 6: S&P100 lag one and lag five autocorrelation matrices of the residuals



(a) Lag 1

(b) Lag 5

The heat map is a diagrammatic representation of the lagged correlations between the residuals of S&P100. The cell ij correspond to the correlation between $\hat{\epsilon}_{i\,t}$ and $\hat{\epsilon}_{j\,t-l}$, where l equals one (top plot) or five (bottom plot).