Série des Documents de Travail

n° 2021-05

Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas

Stefano GRASSI[1]
Francesco VIOLANTE[2]

[1] University of Rome 'Tor Vergata' and CREATES - Aarhus University. Phone: +390672595914 – E-mail:stefano.grassi@uniroma2.it

[2] CREST, GENES, ENSAE Paris, Institut Polytechnique de Paris, 5, avenue Henry Le Chatelier, TSA 96642, 91764 Palaiseau cedex, France and CREATES - Aarhus University. Phone:+33 (0) 1 70 26 67 50 - E-mail:francesco.violante@ensae.fr (correspondingauthor)

Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas

Stefano Grassi[1] and Francesco Violante[2]

February 25, 2021

Abstract

Starting from the Cholesky-GARCH model, recently proposed by Darolles, Francq, and Laurent (2018), the paper introduces the Block-Cholesky GARCH (BC-GARCH). This new model adapts in a natural way to the asset pricing framework. After deriving conditions for stationarity, uniform invertibility and beta tracking, we investigate the finite sample properties of a variety of maximum likelihood estimators suited for the BC-GARCH by means of an extensive Monte Carlo experiment.

We illustrate the usefulness of the BC-GARCH in two empirical applications. The first tests for the presence of beta spillovers in a bivariate system in the context of the Fama and French (1993) three factor framework. The second empirical application consists of a large scale exercise exploring the cross-sectional variation of expected returns for 40 industry portfolios.

Keywords: Cholesky decomposition, Multivariate GARCH, Asset Pricing, Time Varying Beta, Two Pass Regression.

JEL Classification: C12, C22, C58, G12, G13

[1]University of Rome 'Tor Vergata' and CREATES - Aarhus University. Phone: +390672595914 - E-mail:stefano.grassi@uniroma2.it

[2]CREST, GENES, ENSAE Paris, Institut Polytechnique de Paris, 5, avenue Henry Le Chatelier, TSA 96642, 91764 Palaiseau cedex, France and CREATES - Aarhus University. Phone: +33 (0) 1 70 26 67 50 - E-mail: francesco.violante@ensae.fr (corresponding author)

The authors would like to thank Gianluca Cubadda, Christian Francq, Edoardo Otranto, Christian Yann Robert, and Jean-Michel Zakoian for useful comments.

Stefano Grassi gratefully acknowledges financial support from the University of Rome 'Tor Vergata' under Grant "Beyond Borders" (CUP: E84I20000900005). Francesco Violante acknowledges financial support by the grant of the French National Research Agency (ANR), Investissements d'Avenir (LabEx Ecodec/ANR-11-LABX-0047), and of the Danish Council for Independent Research (1329-00011A). The funding organizations had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

1 Introduction

The drawbacks of testing asset pricing models with constant parameters have been long established in the literature. Wrongly assuming parameters constancy may lead to model misspecification, inefficiency and bias in the measurements of risk exposure and premia, potentially resulting in the rejection of the model.

Due to its intuitive appeal and ease of implementation, the 2-Pass Cross-Sectional Regression (2PR) approach, introduced by Fama and MacBeth (1973), has represented a standard in the asset pricing literature to assess the cross-sectional variation of expected assets returns. Time variation in the beta coefficients is obtained by least square estimation on a rolling sample, resulting in dynamic properties of the conditional betas heavily determined by the estimation window size. Also, the rolling least square approach makes it difficult to distinguish between sampling variability and actual time variation in the beta coefficients. Furthermore, the multi-step estimation approach is subject to the well known error-in-variables problem which renders the inference on the risk parameters challenging.

To overcome these issues González, Nave, and Rubio (2012) propose a linear beta pricing model with time-varying risk exposures based on a mixed frequency approach. The conditional betas are estimated using a kernel-based weighted realized variance estimator, which exploits high-frequency information. The model does not require specifying the dynamics of the conditional betas, making it unsuited for forecasting and, most importantly,assuming orthogonality of the risk factors, which is often rejected in practice.

Hansen, Lunde, and Voev (2014) construct a model based on mixed frequencies which exploit additional information coming from non-parametric realized measures of variances and correlations. However, because they opt for modeling correlations directly, individually, and independently from the variances, they cannot explicitly enforce restrictions ensuring positive definiteness of the covariance matrix of the factors-assets system beyond the bivariate dimension.

Recently Engle (2016) propose to model the dynamic conditional betas using a dynamic conditional correlation GARCH (DCB-DCC). In the DCB-DCC model, the betas are recovered as a transformation of the conditional covariance matrix of the joint factors-assets system. While easy to implement, this indirect specification does not allow to identify the relevant drivers of the betas' evolution, and, it makes the coexistence of dynamic and constant betas challenging. Indeed, the constancy constraint is imposed a priori rather than through parameter restrictions on the conditional covariance dynamics.

In a recent paper, Darolles, Francq, and Laurent (2018) propose a model for time-varying betas based on the Cholesky decomposition of the conditional covariance matrix of the factors-assets system. This model, dubbed Cholesky-GARCH (CHAR), explicitly specifies the dynamics of the conditional betas, and simplyfy the coexistence of constant and time-varying betas. The authors show that the CHAR proves superior to the Engle (2016) model in terms of forecasting and beta hedging. However, in asset pricing applications, the CHAR model shows limitations stemming from the strictly triangular Cholesky decomposition and more so when the cross-sectional dimension of the number of assets is large.

Inspired by the CHAR model and borrowing from the decomposition used in the DCB-DCC, we introduce a new model, the Block-Cholesky GARCH (BC-GARCH). The BC-GARCH achieves orthogonalization between two sets of variables, namely risk factors and investment assets via a Block-Cholesky decomposition. This provides several direct advantages over the CHAR.

First, the BC-GARCH preserves the natural hierarchy existing between risk factors and assets without requiring any sorting within the subset of the factors and that of the assets. Consequently, the conditional beta dynamics, stemming from conditioning the assets to the risk factors and driven by the block-orthogonalized innovations, are order independent.

Second, in the asset pricing framework, the BC-GARCH allows for a coherent and parsimonious multivariate evaluation of a multi-asset system. In the CHAR, a capital asset pricing model (CAPM) or Fama-French linear factor model interpretation of the conditional mean equation of the assets is straightforward when the system considered is composed by possibly multiple risk factors but only one asset. When multiple assets are considered, the triangular structure of the CHAR model introduces nuisance conditional betas, whose number increases quadratically with the number of assets.[1] The presence of nuisance betas inevitably rises a dimensionality problem in large systems. Also they do not have clear interpretation as they merely account for the corre-

[1]For instance, in a system with k risk factors and n assets, the CHAR requires $(k+n)[(k+n)-1]/2$ conditional betas, kn of which are "relevant" betas, i.e. measuring the exposure of assets to risk factors, while $n(n-1)/2$ are "nuisance" betas linking assets to assets (as well as $k(k-1)/2$ betas are nuisance conditional betas linking risk factors to risk factors).

lation between assets. The BC-GARCH only requires modeling the conditional betas measuring the exposure of assets to risk factors, whose number grows only linearly in the number of assets. Co-movements between variables in each block are instead modeled explicitly via time-varying covariances or correlations. Darolles, Francq, and Laurent (2018) avoid to some extent the problem by marginalizing the distribution of the multi-asset system iteratively, isolating risk factors and a single asset at each time. Although valid, this approach limits the type of hypotheses on the asset pricing model that can be tested, e.g. restrictions on the cross section of assets.

Third, a consequence of the triangular decomposition of the CHAR is that the magnitude of the orthogonalized innovations tends to vanish by construction as the size of the system increases and the higher the system's correlation. Since these are the innovations driving the conditional betas, they can potentially impact their dynamics in an undesirable way. The block orthogonalization of the BC-GARCH, instead, only underlies a reduction of the magnitude of the orthogonalizad innovations between the two blocks, while preserving it within each block independently of their size.

Fourth, our model introduces spillovers effects in the conditional beta dynamics in an explicit, symmetric and interpretable way. We isolate two different sources of spillovers. We define a factor spillover a shock in a risk factor affecting the exposure of a given asset to a different risk factor. Similarly, we define asset spillovers a shocks in one asset affecting the exposure of another asset to a risk factors. The latter case can be implemented symmetrically across assets because the block structure allows to model them jointly, bypassing the restriction imposed by the strictly triangular structure of the CHAR and thus its inherent sequentiality. Indeed, although the CHAR inherently allows for spillovers of shocks in the conditional betas, the Cholesky orthogonalization makes it impossible to identify the exact source of the shock. Several suitable parametric specifications for the conditional beta dynamics accounting for these features are discussed.

After deriving conditions for stationarity, uniform invertibility and beta tracking, we illustrate the finite sample properties of a variety of maximum likelihood estimators suited for the BC-GARCH by means of extensive Monte Carlo experiments.

In this paper we carry out two empirical applications. In the first, we test our model in the context of the Fama and French (1993) three factor framework (3FF). The aim of this empirical application is to benchmark a conditional beta specification driven only by idiosyncratic shocks against three alternative specifications accounting for beta spillovers. We consider a bivariate asset system composed by the Coal (C) and Petroleum-Natural Gas (P) value-weighted industry portfolios studied over a period spanning from January 1, 1927 to November 30, 2020. Beside time variation in all the conditional betas, we find compelling evidence of both factor and asset spillovers. More specifically, for the Coal industry portfolio, we find significant spillovers of the size factor on the exposure to the market factor as well as asset spillovers on exposure to the value factor. For the Pertroleum-Natural gas portfolio, the size factor significantly impacts the exposure to the market factor. Significant market and asset spillovers are found in the exposure to both the size and the value factors.

The second empirical application consists of a large scale exercise exploring the cross-sectional variation of expected returns for 40 industry portfolios. As standard in the literature, the exercise is carried out using data aggregated at a monthly frequency. We test the linear conditional beta pricing model used in Fama and MacBeth (1973) and estimate the risk premia associated with market factor in CAPM framework as well as those for the exposure to the three Fama-French factors. We benchmark our model against the 2-pass cross-sectional regression approach (2PR) of Fama and MacBeth (1973). Following the 2PR approach, the risk premia for the different factors exposures are estimated as time-averages from a sequence of cross-sectional regressions of the asset returns on the factor betas, where the betas are obtained from a time-series regression of the asset on the factors. We show that our approach provides a more accurate inference on the risk premia. This is because our model benefits from direct modeling of the time variation in the risk exposures and from the joint estimation. The latter, better exploits both the time series and cross sectional dimension of the data avoiding the compounding of estimation error.

In general, there is a compelling difference between the results obtained using the 2PR and those obtained under the BC-GARCH. Inference based on the 2PR always rejects the existence of non-zero risk premia. This result appears to be a direct consequence of the unfavorable signal-to-noise ratio struck by the rolling least square beta estimator, together with the compounding of estimation error of the multistep estimation method.

More precisely, we find a significantly positive market risk premium. This result validates the existence of a positive expected risk-return tradeoff. However, the estimated market risk premium is substantially smaller than the expected market return, violating the Sharpe-Lintner hypothesis and warning against misspecification of the pricing model. Estimates of the size and value risk premia are also significantly positive and comparable in size with the average return of the corresponding factors.

The rest of the paper is organized as follows. Section 2 briefly introduces the notation and operators used throughout the paper. Section 3 introduces the BC-GARCH model and discusses several conditional beta parameterizations, the model's theoretical properties, i.e. stationarity, invertibility and beta targeting, and several quasi maximum likelihood estimation approaches. Monte Carlo simulations results are reported in Section 4 and the empirical applications in Sections 5. Finally, Section 6 concludes.

2 Notation

We make use of the following matrix notation and operators. For any matrix (or vector) $\mathbf{A}$ partitioned in blocks, we denote $\mathbf{A}_{ij}$ the ij-th block of $\mathbf{A}$ and, $a_{ij,[kl]}$ the kl element of the block ij of $\mathbf{A}$. We denote $\mathbf{I}_{(c)}$, the identity matrix of size c, $\mathbf{e}_{(r)}$ the $r \times 1$ unit vector, $\mathbf{0}_{(r\times c)}$ the null matrix of size $r \times c$ and similarly $\mathbf{0}_{(r)}$ denotes the null vector of size $r \times 1$. When specifically needed, we denote $\mathbf{A}_{(c)}$ a square matrix of size c.

We also define $\mathbf{a} = \text{vec}(\mathbf{A})$, the $rc \times 1$ vectorization of an $r \times c$ matrix $\mathbf{A}$ obtained by stacking its columns, and denote $\mathbf{A} = \text{vec}^{-1}_{(r\times c)}(\mathbf{a})$ its inverse. For any two vectors $\mathbf{a}$ of size $r \times 1$ and $\mathbf{b}$ of size $c \times 1$, $\mathbf{a} \otimes \mathbf{b} = \text{vec}(\mathbf{a}\mathbf{b}')$, where $\otimes$ denotes the Kroneker product. Using standard notation, $\odot$ denotes the Hadamard entry-wise product, with $\mathbf{A}^{\odot k}$ its k-th power and $\mathbf{A}^{\odot -1}$ its inverse satisfying $\mathbf{A} \odot \mathbf{A}^{\odot -1} = \mathbf{e}_{(r)}\mathbf{e}'_{(c)}$. For $\mathbf{A}_{(c)}$, $\mathbf{a} = \text{vech}(\mathbf{A})$ defines the $c(c+1)/2 \times 1$ vector that stacks the lower triangular portion, including the main diagonal, of $\mathbf{A}$. We define $\mathbf{a} = \text{diag}(\mathbf{A})$ the $c \times 1$ vector holding the main diagonal of $\mathbf{A}$ and, denote $(\mathbf{A} \odot \mathbf{I}_{(c)}) = \text{diag}^{-1}(\mathbf{a})$ its inverse.

Also, for any two random vectors $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$, $\mathbf{x}_{1,t} \perp \mathbf{x}_{2,t}$ denotes $\text{E}(\mathbf{x}_{1,t}\mathbf{x}_{2,t}) = 0$, i.e. $\mathbf{x}_{1,t}$ is statistically orthogonal to $\mathbf{x}_{2,t}$. Finally, L denotes the lag operator, $L\mathbf{x}_t = \mathbf{x}_{t-1}$ and $\mathcal{I}_{t-1}$ denotes the information set generated by past values of $\mathbf{x}_t$.

3 The Block-Cholesky Garch Model

Consider a vector $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\epsilon}_{t-1};\boldsymbol{\theta})\boldsymbol{\eta}_t$, of possibly correlated and heteroskedastic random variables, with $\boldsymbol{\Sigma}_t(\boldsymbol{\epsilon}_{t-1};\boldsymbol{\theta})$ a, symmetric, positive-definite, $\mathcal{I}_{t-1}$-measurable conditional covariance matrix depending upon a vector of parameters $\boldsymbol{\theta}$ and, $\boldsymbol{\eta}_t$ a sequence of independent and identically distributed (i.i.d) random variables with zero mean and identity covariance matrix. Let us partition the vector $\boldsymbol{\epsilon}_t$ in two blocks, i.e. $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{1,t}, \boldsymbol{\epsilon}_{2,t})'$, of size of sizes k and n respectively.[2]
The conditional covariance matrix $\boldsymbol{\Sigma}_t$ (the arguments have been dropped to simplify the notation) is partitioned accordingly as:

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \boldsymbol{\Sigma}_{11,t} & \boldsymbol{\Sigma}'_{12,t} \\ \boldsymbol{\Sigma}_{12,t} & \boldsymbol{\Sigma}_{22,t} \end{bmatrix}, \tag{1}$$

with diagonal blocks square, symmetric and positive definite.
Since the upper left block is invertible, we can further factor $\boldsymbol{\Sigma}_t$ by means of a block LDL$'$ Cholesky decomposition as:

$$\begin{bmatrix} \mathbf{I}_{(k)} & \mathbf{0}_{(k\times n)} \\ \mathbf{B}_t & \mathbf{I}_{(n)} \end{bmatrix} \begin{bmatrix} \mathcal{S}_{11,t} & \mathbf{0}_{(k\times n)} \\ \mathbf{0}_{(n\times k)} & \mathcal{S}_{22,t} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{(k)} & \mathbf{B}'_t \\ \mathbf{0}_{(n\times k)} & \mathbf{I}_{(n)} \end{bmatrix}. \tag{2}$$

This decomposition entails $\mathcal{S}_{11,t} = \boldsymbol{\Sigma}_{11,t}$, $\mathcal{S}_{22,t} = \boldsymbol{\Sigma}_{22,t} - \boldsymbol{\Sigma}_{12,t}\boldsymbol{\Sigma}_{11,t}^{-1}\boldsymbol{\Sigma}'_{12,t}$, i.e. the Schur complement of $\boldsymbol{\Sigma}_{22,t}$ and, $\mathbf{B}_t = \boldsymbol{\Sigma}_{12,t}\boldsymbol{\Sigma}_{11,t}^{-1}$, i.e. the $n \times k$ matrix of time-varying coefficients of the regressions of $\epsilon_{2,[i],t}$ on $\boldsymbol{\epsilon}_{1,t}$, $i = 1, \ldots, n$. In fact, noting that

$$\begin{bmatrix} \mathbf{I}_{(k)} & \mathbf{0}_{(k\times n)} \\ \mathbf{B}_t & \mathbf{I}_{(n)} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_{(k)} & \mathbf{0}_{(k\times n)} \\ -\mathbf{B}_t & \mathbf{I}_{(n)}, \end{bmatrix}, \tag{3}$$

holds because the lower uni-triangular matrix on the right hand side of equation (3) fulfills the identity on the diagonal blocks, then using (2) we can write:

$$\begin{bmatrix} \mathbf{I}_{(k)} & \mathbf{0}_{(k\times n)} \\ -\mathbf{B}_t & \mathbf{I}_{(n)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}_{1,t} \\ \boldsymbol{\epsilon}_{2,t} \end{bmatrix} = \begin{bmatrix} \mathcal{S}_{11,t} & \mathbf{0}_{(k\times n)} \\ \mathbf{0}_{(n\times k)} & \mathcal{S}_{22,t} \end{bmatrix}^{1/2} \begin{bmatrix} \boldsymbol{\eta}_{1,t} \\ \boldsymbol{\eta}_{2,t} \end{bmatrix}. \tag{4}$$

Note that, provided $\boldsymbol{\epsilon}_t$ is ergodic and second order stationary, by the same arguments, $\mathbf{B}_{\mathrm{OLS}} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}$ obtained using the block LDL$'$ Cholesky decomposition of the unconditional covariance

[2] In asset pricing applications, e.g. CAPM, linear factors models, etc., this structure reflects the partition in k risk factors $\boldsymbol{\epsilon}_{1,t}$ and n investment assets $\boldsymbol{\epsilon}_{2,t}$.

matrix of $\boldsymbol{\epsilon}_t$, represents the matrix of unconditional (least square) betas.

The conditional beta $\mathbf{B}_t$ in equation (4) is assumed to be a $\mathcal{I}_{t-1}$-measurable dynamic process which, under suitable conditions discussed below, is centered in $\mathbf{B}_{\text{OLS}}$. The model is completed with a suitable choice of the dynamics of $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$ in the multivariate GARCH family, see Bauwens, Laurent, and Rombouts (2006) for a survey.

The decomposition in equation (2) relates to the DCB-DCC model of Engle (2016). In the Gaussian framework the block orthogonalization above naturally stems from the marginalization of the distribution of $\boldsymbol{\epsilon}_{1,t}$ and conditioning of the distribution of $\boldsymbol{\epsilon}_{2,t}$ to $\boldsymbol{\epsilon}_{1,t}$. Engle (2016) models $\boldsymbol{\Sigma}_t$ by means of dynamic conditional correlations (DCC), recovering indirectly the time-varying betas as $\mathbf{B}_t = \boldsymbol{\Sigma}_{12,t}\boldsymbol{\Sigma}_{11,t}^{-1}$. This approach suffers the same internal inconsistency characterizing non linear dynamics as it implies $\mathrm{E}(\mathbf{B}_t) = \mathrm{E}\left(\boldsymbol{\Sigma}_{12,t}\boldsymbol{\Sigma}_{11,t}^{-1}\right) \neq \mathrm{E}\left(\boldsymbol{\Sigma}_{12,t}\right)\mathrm{E}\left(\boldsymbol{\Sigma}_{11,t}\right)^{-1} = \mathbf{B}_{\text{OLS}}$. Also, being the betas a non-linear function of the elements of $\boldsymbol{\Sigma}_t$, the coexistence of dynamic and constant betas becomes rather convoluted. This is because it requires imposing constraints on the structure of the conditional covariance matrix itself, not simply on the parameters governing its dynamics, while ensuring its positive definiteness. Although Engle (2016) considers the case when some of the elements of $\mathbf{B}_t$ are constant, this constraint is imposed a priori, and then statistically validated ex-post, rather then via testable parameter restrictions.

The BC-GARCH mixes the block decomposition of the DCB-DCC with the modeling strategy of the CHAR, which specifies directly the dynamics of the elements of $\mathbf{B}_t$. However, unlike the CHAR, which is subject to dependence on the ordering imposed by the standard Cholesky factorization, the BC-GARCH only requires ordering between blocks without imposing restrictions on the ordering within each block[3]. This translates into a smaller number of conditional betas to be modeled, i.e. only those linking first block to second block variables, regardless of their number or order, while co-movements between variables in each block are modeled explicitly via time varying covariances or correlations.

[3]In a linear asset pricing framework this approach conforms to the natural ordering between factors (first block) and assets (second block), while treating variables in each block on equal ground.

3.1 Conditional Beta Specifications

The dynamics of $\mathbf{B}_t$ can be specified in a variety of ways. We explicitly focus on a family of specifications where $\mathbf{B}_t$ depends on products of block-orthogonal innovations. The latter are defined as $\mathbf{v}_{1,t} = \boldsymbol{\epsilon}_{1,t}$ and $\mathbf{v}_{2,t} = \boldsymbol{\epsilon}_{2,t} - \mathbf{B}_t\boldsymbol{\epsilon}_{1,t}$, respectively. This choice is motivated by the fact that Darolles, Francq, and Laurent (2018) find these dynamics to be superior, in empirical applications, to many competing specifications.

A general specification for $\mathbf{B}_t$ is

$$\mathbf{B}_t = \boldsymbol{\Psi} + \left[\text{vec}^{-1}_{(k\times n)}\left(\sum_{p=1}^{P}\boldsymbol{\Omega}_p\left(\mathbf{v}_{2,t-p}\otimes\mathbf{v}_{1,t-p}\right) + \sum_{q=1}^{Q}\boldsymbol{\Gamma}_q\text{vec}\left(\mathbf{B}'_{t-q}\right)\right)\right]', \tag{5}$$

where $\boldsymbol{\Psi}$ is an $n \times k$ matrix, while $\boldsymbol{\Omega}_p$ and $\boldsymbol{\Gamma}_q$ are $nk \times nk$ matrices.

It is important to stress that positive definiteness of $\boldsymbol{\Sigma}_t$ in equation (1) only requires semi positive definiteness of $\mathcal{S}_{22,t}$ while $\mathbf{B}_t \in \mathrm{R}^{n\times k}$, $\forall\, t$. Thus no sign restriction on the parameters in equation (5) needs to be imposed.

The formulation in equation (5) allows individual betas to be constant. Denoting $\boldsymbol{\omega}_{s,p}$ ($\boldsymbol{\gamma}_{s,p}$) the s-th row of $\boldsymbol{\Omega}_p$ ($\boldsymbol{\Gamma}_p$), under $\boldsymbol{\omega}_{s,p} = 0 \cup \boldsymbol{\gamma}_{s,p} = 0$ for some $s \in [1, kn]$ and $\forall p$, then $\beta_{ij,t} = \beta_{ij}$. Similarly, it nests the constant beta regression, that is all betas in the regression of $\epsilon_{2,[i],t}$ on $\mathbf{v}_{1,t}$, for some $i \in [1, n]$, are constant. If the i-th block of k rows of $\boldsymbol{\Omega}_p$ and $\boldsymbol{\Gamma}_p$, i.e. from row $k(i-1)+1$ to ik, $i \in [1, n]$, are equal to zero $\forall p$, then $\epsilon_{2,[i],t} = \sum_{j=1}^{k}\beta_{ij}\epsilon_{1,[j],t} + v_{2,[i],t}$.

Though very general, the inclusion of all cross-products of $\mathbf{v}_{1,t}$ and $\mathbf{v}_{2,t}$, as well as lagged betas, makes equation (5), with its $nk\,(1 + 2nk(P + Q))$ parameters, computationally cumbersome. However, by imposing suitable restrictions on $\boldsymbol{\Omega}_p$ and $\boldsymbol{\Gamma}_q$ we can identify several specific parameterisations relevant for financial applications.

Equation (5) encompasses dynamics driven exclusively by idiosyncratic shocks as well as dynamics allowing for different types and degrees of spillovers. The idea of beta spillovers relates closely to that of covariance and correlation spillovers in the multivariate GARCH literature. Our model mimics the same transmission mechanism. Equation (5) incorporates beta spillovers in an explicit, easily interpretable and analytically tractable manner. This is because each block of $\boldsymbol{\epsilon}_t$ retains the multivariate nature as well as the property of invariance to the ordering of its elements.

Beta spillovers can arise from either, or both, partitions $\boldsymbol{\epsilon}_{1,t}$ and $\boldsymbol{\epsilon}_{2,t}$. In the asset pricing framework, where $\boldsymbol{\epsilon}_{1,t}$ represents a set of market-wide risk factors to which a set of investment assets $\boldsymbol{\epsilon}_{2,t}$ are exposed to, asset and factor spillovers can be defined accordingly as the change in the exposure of a certain asset to a given factor stemming from shocks in competing risk factors and/or assets. Without loss of generality and to simplify the notation, we focus on specifications of equation (5) with $P = Q = 1$.

i) Idiosyncratic shocks (direct effect). This is a simple dynamics obtained by imposing diagonality of $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$. This parameterization represents our baseline specification. Under these restrictions $\beta_{ij,t}$ linking the j-th element of $\boldsymbol{\epsilon}_{1,t}$ to the i-th element of $\boldsymbol{\epsilon}_{2,t}$ depends solely on the idiosyncratic shock constructed as the product of the corresponding i-th and j-th innovations.[4] For $i \in [1, n]$ and $j \in [1, k]$, denote $\omega_{ii,[jj]}$ and $\gamma_{ii,[jj]}$ the j-th diagonal element of the i-th $(k \times k)$ diagonal block of $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ respectively. The typical element of $\mathbf{B}_t$ is then:

$$\beta_{ij,t} = \psi_{ij} + \omega_{ii,[jj]} v_{2,[i],t-1} v_{1,[j],t-1} + \gamma_{ii,[jj]} \beta_{ij,t-1}, \quad i = 1, \ldots, n; \ j = 1, \ldots, k.$$

ii) Factor spillovers. Spillovers generated by $\boldsymbol{\epsilon}_{1,t}$ on $\beta_{ij,t}$, linking $\epsilon_{1,[j],t}$ to $\epsilon_{2,[i],t}$, are introduced by letting $\beta_{ij,t}$ depends on the set of products $\left\{v_{2,[i],t-1} v_{1,[s],t-1}\right\}_{s=1,\ldots,k \setminus s=j}$, as well as, lagged betas $\{\beta_{is,t-1}\}_{s=1,\ldots,k \setminus s=j}$. The terminology factor spillover is used here as an explicit reference to the aforementioned role of $\boldsymbol{\epsilon}_{1,t}$ in asset pricing applications. This parameterization requires $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ in equation (5) to be block diagonal with n blocks of size $k \times k$, i.e. $\boldsymbol{\Omega}_{ii}$ and $\boldsymbol{\Gamma}_{ii}$ $i = 1, \ldots, n$ are non-zero matrices. The typical element of $\mathbf{B}_t$, takes the form:

$$\beta_{ij,t} = \psi_{ij} + \left(\sum_{s=1}^{k} \omega_{ii,[js]} v_{1,[s],t-1} \right) v_{2,[i],t-1} + \sum_{s=1}^{k} \gamma_{ii,[js]} \beta_{is,t-1}, \quad i = 1, \ldots, n; \ j = 1, \ldots, k.$$

iii) Asset spillovers. The symmetric structure, with respect to $\boldsymbol{\epsilon}_{2,t}$, of the block orthogonalization can be exploited to incorporate the effect of spillovers stemming from the elements of $\boldsymbol{\epsilon}_{2,t}$, bypassing the restriction imposed by the usual Cholesky triangular representation as in Darolles, Francq, and Laurent (2018). Also in this case, the reference to the assets alludes to the

[4] By defining $\boldsymbol{\Omega}^* = \left(\text{vec}^{-1}_{(k \times n)}\left(\text{diag}(\boldsymbol{\Omega})\right)\right)'$ and $\boldsymbol{\Gamma}^* = \left(\text{vec}^{-1}_{(k \times n)}\left(\text{diag}(\boldsymbol{\Gamma})\right)\right)'$, equation (5) can be expressed alternatively as:

$$\mathbf{B}_t = \boldsymbol{\Psi} + \left(\boldsymbol{\Omega}^* \odot \mathbf{v}_{2,t-1} \mathbf{v}'_{1,t-1}\right) + \left(\boldsymbol{\Gamma}^* \odot \mathbf{B}_{t-1}\right), \tag{5.i}$$

where Ψ, $\boldsymbol{\Omega}^*$ and $\boldsymbol{\Gamma}^*$ are $n \times k$ parameters matrices.

asset pricing context. Spillovers generated by $\boldsymbol{\epsilon}_{2,t}$ on $\beta_{ij,t}$ are introduced by adding the products $\left\{v_{2,[s],t-1}v_{1,[j],t-1}\right\}_{s=1,\ldots n\backslash s=j}$, as well as, lagged betas $\{\beta_{sj,t-1}\}_{s=1,\ldots n\backslash s=j}$. This parameterization requires $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ in equation (5) to be $n\times n$ block matrices with diagonal blocks each of size $k\times k$, i.e. $\mathbf{\Omega}_{ij}$ and $\mathbf{\Gamma}_{ij}$ $i,j=1,\ldots,n$ are diagonal matrices.[5] The typical element of the matrix of conditional betas $\mathbf{B}_t$ is expressed as:

$$\beta_{ij,t}=\psi_{ij}+\left(\sum_{s=1}^{n}\omega_{is,[jj]}v_{2,[s],t-1}\right)v_{1,[j],t-1}+\sum_{s=1}^{n}\gamma_{is,[jj]}\beta_{sj,t-1},\quad i=1,\ldots,n;\ j=1,\ldots,k.$$

iv) Factor and asset spillovers. Let $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ in equation (5) to be $n\times n$ block matrices with blocks $\mathbf{\Omega}_{ij}$ and $\mathbf{\Gamma}_{ij}$ of size $k\times k$. These blocks are full matrices if $i=j$ and diagonal if $i\neq j$, $i,j=1,\ldots,n$. This empirically relevant parameterization allows for both types of spillovers described above. The typical element of the $\mathbf{B}_t$ is:

$$\begin{aligned}\beta_{ij,t} &= \psi_{ij}+\left(\sum_{s=1}^{k}\omega_{ii,[js]}v_{1,[s],t-1}\right)v_{2,[i],t-1}+\left(\sum_{s\neq i,s=1}^{n}\omega_{is,[jj]}v_{2,[s],t-1}\right)v_{1,[j],t-1}+\\ &\qquad \sum_{s=1}^{k}\gamma_{ii,[js]}\beta_{is,t-1}+\sum_{s\neq i,s=1}^{n}\gamma_{is,[jj]}\beta_{sj,t-1},\quad i=1,\ldots,n;\ j=1,\ldots,k.\end{aligned}$$

Conditions for stationarity, beta and covariance tracking, as well as invertibility are given in Sections 3.2 and 3.3.

3.2 Stationarity and targeting

We derive sufficient stationarity conditions under the general specification in (5), as well as explicit conditions under the restrictions in i), ii), iii) and iv) in Section 3.1.

Define $\boldsymbol{\beta}_t=\text{vec}\left(\mathbf{B}_t'\right)$ and $\mathbf{w}_t=\text{vec}(\mathbf{\Psi}')+\sum_{p=1}^{P}\mathbf{\Omega}_p\left(\mathbf{v}_{2,t-p}\otimes\mathbf{v}_{1,t-p}\right)$ where the parameter's matrices are defined according equation (5). Assume that the block orthogonal process $\mathbf{z}_t=(\mathbf{z}_{1,t},\mathbf{z}_{2,t})$ with $\mathbf{z}_{i,t}=(\mathbf{v}_{i,t},\text{vec}(\mathcal{S}_{ii,t}))'$, $i=1,2$ is strictly stationary and ergodic. Conditions for stationarity and ergodicity of $\mathbf{z}_t$ are specific to the GARCH parameterization of $\mathcal{S}_{ii,t}$ and, are given in Pedersen

[5] Since all blocks of $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ are diagonal, one can define a $nk\times nk$ permutation matrix $\mathbf{P}$ such that $\mathbf{\Omega}^*=\mathbf{P\Omega P}'$ is block diagonal. By reordering accordingly the terms in parentheses on the right hand side of equation (5), this conditional beta specification can be rewritten as:

$$\mathbf{B}_t=\mathbf{\Psi}+\text{vec}^{-1}_{(n\times k)}\Big(\mathbf{\Omega}^*\left(\mathbf{v}_{1,t-1}\otimes\mathbf{v}_{2,t-1}\right)+\mathbf{\Gamma}^*\text{vec}\left(\mathbf{B}_{t-1}\right)\Big),\qquad (5.\text{iii})$$

where $\mathbf{\Omega}^*$ and $\mathbf{\Gamma}^*$ are $(nk\times nk)$ block diagonal matrices with blocks size $n\times n$.

and Rahbek (2014), Boussama, Fuchs, and Stelzer (2011), Hafner and Preminger (2009), Engle and Kroner (1995), Fermanian and Malongo (2017), Francq and Zakoian (2012) and Francq and Zakoian (2016) among others.

By the ergodic theorem, $\mathbf{w}_t$ is stationary and ergodic. It follows that if $|\mathbf{\Gamma}_Q(z)| = \left|\mathbf{I}_{(nk\times nk)} - \sum_{q=1}^{Q} \mathbf{\Gamma}_q z^q\right| \neq 0$ for all $|z| \leq 1$ then $\boldsymbol{\beta}_t = \mathbf{\Gamma}_Q(L)^{-1}\mathbf{w}_t = \sum_{i=0}^{\infty} \mathcal{G}_i \mathbf{w}_{t-i}$, $\mathcal{G}_0 = \mathbf{I}_{(nk\times nk)}$ and $\sum_{i=0}^{\infty} \|\mathcal{G}_i\| < \infty$, is stationary and ergodic. The solution to the model is then defined by $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{1,t}, \boldsymbol{\epsilon}_{2,t})'$ with $\boldsymbol{\epsilon}_{1,t} = \mathbf{v}_{1,t}$ and $\boldsymbol{\epsilon}_{2,t} = \left(\mathbf{B}_t,\ \mathbf{I}_{(n)}\right)(\mathbf{v}_{1,t}, \mathbf{v}_{2,t})'$.

More explicit stationarity condition can be written for equation (5) under the parameter restrictions i), ii), iii) and iv) discussed in Section 3.1 with $P = Q = 1$:

i) *Idiosyncratic shocks:* $\max_s |\gamma_{ss}| < 1$, $s = 1, \ldots, nk$, i.e. the largest absolute non-zero element of the diagonal matrix Γ;

ii) *Factor spillovers:* $\max_i \lambda_i^{max+}(\mathbf{\Gamma}_{ii}) < 1$, $i = 1, \ldots, n$, where $\lambda_i^{max+}(\mathbf{\Gamma}_{ii})$ is the largest absolute eigenvalue of the i-th $k \times k$ diagonal block of $\mathbf{\Gamma}$;

iii) *Asset spillovers:* $\max_j \lambda_j^{max+}(\mathbf{\Gamma}_{jj}^*) < 1$, $j = 1, \ldots, k$, where $\mathbf{\Gamma}^* = \mathbf{P}\mathbf{\Gamma}\mathbf{P}'$ is the block diagonalization of $\mathbf{\Gamma}$, i.e. is a block diagonal matrix with k non-zero blocks of size $n \times n$, obtained by rotating the rows and columns of $\mathbf{\Gamma}$ via the permutation matrix $\mathbf{P}$ and, $\lambda_j^{max+}(\mathbf{\Gamma}_{jj}^*)$ is the largest absolute eigenvalue of the j-th diagonal block of $\mathbf{\Gamma}^*$.

iv) *Factor and Asset spillovers:* $\max_s \lambda_s^{max+}(\mathbf{\Gamma}) < 1$, $s = 1, \ldots, nk$, i.e. the largest absolute eigenvalue of Γ;

The condition above, together with $\mathbf{v}_{1,t} \perp \mathbf{v}_{2,t}$, also implies:

$$\operatorname*{plim}_{\tau\to\infty} \mathrm{E}_{t-\tau}(\boldsymbol{\beta}_t) = \mathbf{\Gamma}_Q(1)^{-1}\mathrm{vec}(\mathbf{\Psi}) = \mathrm{E}(\boldsymbol{\beta}_t),$$

where $\mathrm{E}(\boldsymbol{\beta}_t) \equiv \mathrm{vec}(\mathbf{B}')$.

The unconditional variance of $\boldsymbol{\epsilon}_t$, is obtained by combining $\mathrm{E}\left(\boldsymbol{\epsilon}_{1,t}\boldsymbol{\epsilon}_{1,t}'\right) \equiv \mathbf{\Sigma}_{11} = \mathrm{E}\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right)$ and $\mathrm{E}\left(\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{1,t}'\right) \equiv \mathbf{\Sigma}_{21} = \mathrm{E}\left(\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right) = \mathrm{E}\left(\mathbf{B}_t\mathrm{E}_{t-1}(\mathbf{v}_{1,t}\mathbf{v}_{1,t}')\right) = \mathrm{E}\left(\mathbf{B}_t\mathbf{\Sigma}_{11,t}\right)$. It is worth noting that, under the condition $\mathrm{E}(\mathbf{B}_t\mathbf{\Sigma}_{11,t}) = \mathbf{B}\mathbf{\Sigma}_{11}$, i.e. $\mathbf{B}_t$ and $\mathbf{\Sigma}_{11,t}$ are unconditionally uncorrelated, then $\mathbf{\Sigma}_{21} = \mathbf{B}\mathbf{\Sigma}_{11}$ which implies $\mathbf{B} = \mathbf{B}_{\mathrm{OLS}}$. This condition requires additional distributional assumptions for $\boldsymbol{\epsilon}_t$ and, when satisfied, allows targeting of the unconditional beta to the least squares estimator.[6] Absence of correlation between $\mathbf{B}_t$ and $\boldsymbol{\Sigma}_{11,t}$ is satisfied, e.g. under Gaussianity of $\boldsymbol{\epsilon}_t$ together with $\mathbf{v}_{1,t} \perp \mathbf{v}_{2,t}$ (which in this case implies independence) and conditional beta defined by equation (5).[7] Finally, from equation (4):

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{2,t}'\right) \equiv \boldsymbol{\Sigma}_{22,t} &= \mathrm{E}\left(\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\mathbf{B}_t'\right) + \mathrm{E}\left(\mathbf{v}_{2,t}\mathbf{v}_{2,t}'\right) \\
&= \mathrm{E}\left(\mathbf{B}_t\mathrm{E}_{t-1}(\mathbf{v}_{1,t}\mathbf{v}_{1,t}')\mathbf{B}_t'\right) + \mathrm{E}(\mathcal{S}_{22,t}) \\
&= \mathrm{E}\left(\mathbf{B}_t\boldsymbol{\Sigma}_{11,t}\mathbf{B}_t'\right) + \mathcal{S}_{22}.
\end{aligned}$$

Knowledge of $\mathcal{S}_{22}$ would allow targeting the covariance of $\mathbf{v}_{2,t}$ to a sample estimator, which is particularly advantageous when the cross-sectional dimension of the second block n is large. However, even under the condition of zero correlation between $\mathbf{B}_t$ and $\boldsymbol{\Sigma}_{11,t}$ discussed above, the estimator $\boldsymbol{\Sigma}_{22} - \mathbf{B}_{\mathrm{OLS}}\boldsymbol{\Sigma}_{11}\mathbf{B}_{\mathrm{OLS}}'$ stemming from the unconditional equivalent of the decomposition in equation (2), is not in general unbiased for $\mathcal{S}_{22}$, i.e. $\mathrm{E}(\mathbf{B}_t\boldsymbol{\Sigma}_{11,t}\mathbf{B}_t') \neq \mathbf{B}_{\mathrm{OLS}}\boldsymbol{\Sigma}_{11}\mathbf{B}_{\mathrm{OLS}}'$, except for the trivial case of constant beta, as it ignores time variation in $\mathbf{B}_t$. This poses limits to covariance targeting of $\mathbf{v}_{2,t}$ to a sample estimator.[8]

---

[6] Unbiasedness and relative efficiency of the beta targeting based on the least square estimator are studied in a Monte Carlo simulation exercise in Section 4.

[7] As an example, let assume for simplicity that the conditional beta dynamics are given in equation (5.i), driven only by idiosyncratic shocks, then under stationarity we can write:

$$\mathbf{B}_t = \mathbf{B} + \boldsymbol{\Omega} \odot \sum_{i=0}^{\infty} \boldsymbol{\Gamma}^{\odot i} \odot \left(\mathbf{v}_{2,t-i-1}\mathbf{v}_{1,t-i-1}'\right),$$

with $\mathrm{E}(\mathbf{B}_t) \equiv \mathbf{B} = (\mathbf{e}_{(n)}\mathbf{e}_{(k)}' - \boldsymbol{\Gamma})^{\odot -1} \odot \boldsymbol{\Psi}$ because $\mathrm{E}(\mathbf{v}_{2,t-i-1}\mathbf{v}_{1,t-i-1}') = 0 \ \forall i$, i.e. $\mathbf{v}_{1,t} \perp \mathbf{v}_{2,t}$. Thus:

$$\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}' \quad = \quad \mathbf{B}\mathbf{v}_{1,t}\mathbf{v}_{1,t}' + \left(\boldsymbol{\Omega} \odot \sum_{s=0}^{\infty} \boldsymbol{\Gamma}^{\odot s} \odot (\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}')\right)\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right).$$

Expectations of the second term on the rhs are all null. They depend on linear combinations of k^2n 4th-order moments $\mathrm{E}(\mathrm{v}_{1,[m],t}\mathrm{v}_{1,[i],t}\mathrm{v}_{1,[i],t-s}\mathrm{v}_{2,[j],t-s})$, $i, m = 1, \ldots, k$ $j = 1, \ldots, n$ $s \in \mathbb{N}$ which, under Gaussianity of $\mathbf{v}_t$, absence of serial (cross) correlation and orthogonality of the two blocks, are all null. Hence, we obtain $\mathrm{E}(\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}') = \mathbf{B}\Sigma_{11}$, since $\mathrm{E}(\mathbf{v}_{1,t}\mathbf{v}_{1,t}') = \mathcal{S}_{11} = \Sigma_{11}$.

[8] $\mathcal{S}_{22}$ expressed in terms of the parameters of the model and sample moments of the data, is in general a complicated function depending on the specification of B_t and requiring additional assumptions on the distribution of $\boldsymbol{\epsilon}_t$ to be determined. This is because it depends on moments, up to the 6-th order, of the joint distribution of $\mathbf{v}_{1,t}$ and $\mathbf{v}_{2,t}$, where the latter is unobserved. The variance targeting estimator can thus be defined as semi-nonparametric. Continuing the example based on equation (5.i), and defining $\mathbf{A}_s = \boldsymbol{\Omega} \odot \boldsymbol{\Gamma}^{\odot s}$, $s = 1, \ldots, \infty$, with typical element $\omega_{ij}\gamma_{ij}^s$, $i = 1, \ldots, n$ and $j = 1, \ldots, k$, we can write:

$$\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\mathbf{B}_t' \quad = \quad \mathbf{B}\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\mathbf{B}' + \left(\sum_{s=0}^{\infty} \mathbf{A}_s \odot \left(\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}'\right)\right)\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right)\mathbf{B}' +$$

3.3 Uniform Invertibility

We now derive invertibility conditions, i.e. asymptotic irrelevance of the initial values for equation (5). Conditions for parameterizations under the restrictions decribed in Section 3.1 i), ii) and iii) are derived explicitly for completeness at the end of this section.

Let $\boldsymbol{\theta} = (\text{vec}(\boldsymbol{\Psi}),\ \text{vec}(\boldsymbol{\Omega}),\ \text{vec}(\boldsymbol{\Gamma}))'$ be a generic element of the convex parameter space $\boldsymbol{\Theta}$. Endowed with any arbitrary set of initial values $\widetilde{\boldsymbol{\epsilon}}_0 = (\widetilde{\boldsymbol{\epsilon}}_{1,0}, \widetilde{\boldsymbol{\epsilon}}_{2,0})'$ and $\mathbf{B}_0$ at $t = 0$, define $\widetilde{\mathbf{B}}_t(\boldsymbol{\theta}) = \mathbf{B}_t(\boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_1, \widetilde{\boldsymbol{\epsilon}}_0; \boldsymbol{\theta})$. Recalling that $\widetilde{\mathbf{v}}_{1,t}(\boldsymbol{\theta}) = \boldsymbol{\epsilon}_{1,t}$, $\widetilde{\mathbf{v}}_{2,t}(\boldsymbol{\theta}) = -\widetilde{\mathbf{B}}_t(\boldsymbol{\theta})\boldsymbol{\epsilon}_{1,t} + \mathbf{I}_{(n)}\boldsymbol{\epsilon}_{2,t}$ and using equation (5), assuming without loss of generality, $P = Q = 1$, we can define $\widetilde{\mathbf{B}}_t(\boldsymbol{\theta})$ for all $t > 0$ and for any value of $\boldsymbol{\theta}$. If $\widetilde{\mathbf{B}}_t(\boldsymbol{\theta})$ and $\widetilde{\mathbf{v}}_t$ are consistently approximated by the $\mathcal{I}_{t-1}$-measurable functions $\mathbf{B}_t(\boldsymbol{\theta})$ and $\mathbf{v}_t$, then the model is invertible.

The process $\widetilde{\mathbf{B}}_t(\boldsymbol{\theta})$ can be expressed in vector form in terms of the observables, $(\boldsymbol{\epsilon}_{1,t},\ \boldsymbol{\epsilon}_{2,t})'$, as a

---

$$+\mathbf{B}(\mathbf{v}_{1,t}\mathbf{v}_{1,t}') \left(\sum_{s=0}^{\infty} \left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{2,t-s-1}'\right) \odot \mathbf{A}_s' \right) +$$

$$+\sum_{s=0}^{\infty}\sum_{r=0}^{\infty} \left(\mathbf{A}_s \odot \left(\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}'\right)\right) \left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right) \left(\left(\mathbf{v}_{1,t-r-1}\mathbf{v}_{2,t-r-1}'\right) \odot \mathbf{A}_r'\right).$$

The expectation of the first term on the right hand side is trivially $\mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}'$, while those of the second and third terms depend on 4-th order moments which are null under Gaussianity of $\mathbf{v}_t$, absence of serial (cross) correlation and block orthogonality (or in any case where absence of correlation implies independence), see footnote 7. For the last term on the right hand side, expectations of each term in the double summation are linear combinations of $(kn)^2$ 6th-order moments of the form:

$$\text{E}(\text{v}_{1,[m],t}\text{v}_{1,[i],t}\text{v}_{1,[m],t-s}\text{v}_{1,[i],t-r}\text{v}_{2,[j],t-s}\text{v}_{2,[l],t-r}), \qquad i, m = 1, \ldots, k \quad j, l = 1, \ldots, n \quad r, s \in \mathbb{N},$$

which, under the distributional assumptions stated above, are null for $r \neq s$, while for $r = s$ it holds

$$\text{E}\left(\left(\mathbf{A}_s \odot \left(\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}'\right)\right)\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right)\left(\left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{2,t-s-1}'\right) \odot \mathbf{A}_s'\right)\right) = \mathcal{A}_s \odot \mathcal{S}_{22},$$

where $\mathcal{A}_s$ is a $n \times n$ matrix with typical element $\alpha_{ij,s} = \mathbf{a}_{i,s}\text{E}\left(\left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{1,t-s-1}'\right) \odot \left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right)\right)\mathbf{a}_{j,s}'$, $i, j = 1, \ldots, n$, and $\mathbf{a}_i$ is the i-th row of the matrix of parameters $\mathbf{A}_s$. Thus:

$$\text{E}(\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\mathbf{B}_t') = \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}' + \sum_{s=0}^{\infty} \mathcal{A}_s \odot \mathcal{S}_{22},$$

and $\mathcal{S}_{22}$, as a function of sample moments and model parameters, can be determined as the solution of

$$\boldsymbol{\Sigma}_{22} = \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}' + \sum_{s=0}^{\infty} \mathcal{A}_s \odot \mathcal{S}_{22} + \mathcal{S}_{22},$$

noting that the elements of the sequence of matrices $\mathcal{A}_s$ are an exponentially decaying power function of the elements of $\boldsymbol{\Gamma}$. Finally, it is worth noting that under the restriction $\boldsymbol{\Omega} = \boldsymbol{\Gamma} = \mathbf{0}$, then $\mathcal{A}_s = \mathbf{0}\ \forall s$, and the Schur complement of $\boldsymbol{\Sigma}_{22}$, i.e. $\mathcal{S}_{22} = \boldsymbol{\Sigma}_{22} - \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}'$, under constancy of the betas is recovered.

random coefficients recursion:

$$\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta}) = \mathbf{w}_{t-1} + \mathbf{S}_{t-1}\widetilde{\boldsymbol{\beta}}_{t-1}(\boldsymbol{\theta}), \tag{6}$$

where $\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta}) = \text{vec}\left(\widetilde{\mathbf{B}}_t(\boldsymbol{\theta})'\right)$, $\mathbf{w}_t = \text{vec}(\boldsymbol{\Psi}') + \boldsymbol{\Omega}(\boldsymbol{\epsilon}_{2,t} \otimes \boldsymbol{\epsilon}_{1,t})$ and, $\mathbf{S}_t = \boldsymbol{\Gamma} - \boldsymbol{\Omega}\left(\mathbf{I}_{(n)} \otimes \left(\boldsymbol{\epsilon}_{1,t}\boldsymbol{\epsilon}'_{1,t}\right)\right)$. From equation (6), by recursive substitution we get:

$$\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta}) = \mathbf{w}_{t-1} + \sum_{i=1}^{t-2}\left(\prod_{j=1}^{i}\mathbf{S}_{t-j}\right)\mathbf{w}_{t-i-1} + \left(\prod_{i=1}^{t-1}\mathbf{S}_{t-i}\right)\left(\mathbf{w}_0 + \widetilde{\boldsymbol{\beta}}_0(\boldsymbol{\theta})\right). \tag{7}$$

Following the arguments of Darolles, Francq, and Laurent (2018), by the Cauchy rule and for any multiplicative norm $\|\cdot\|$, holding the following conditions:

$$\text{E}(\log \sup_{\theta} \|\mathbf{w}_1\|) < \infty \tag{c.1}$$

$$\limsup_{s\to\infty}\left(\frac{1}{s}\log\sup_{\theta}\left\|\prod_{i=1}^{s}\mathbf{S}_{t-i}\right\|\right) < 0, \tag{c.2}$$

the series $\boldsymbol{\beta}_t(\boldsymbol{\theta}) = \mathbf{w}_{t-1} + \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\mathbf{S}_{t-j}\right)\mathbf{w}_{t-i-1}$ converges absolutely and uniformly to $\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta})$ with probability one. Using Jensen's inequality, by sub-additivity and sub-multiplicativity of $\|\cdot\|$ and, noting that $\left\|\boldsymbol{\epsilon}_{2,1} \otimes \boldsymbol{\epsilon}'_{1,1}\right\| = \left\|\text{vec}(\boldsymbol{\epsilon}_{2,1}\boldsymbol{\epsilon}'_{1,1})\right\|$, condition (c.1) holds if $\text{E}(\left\|\text{vec}(\boldsymbol{\epsilon}_{2,1}\boldsymbol{\epsilon}'_{1,1})\right\|^{s_0}) < \infty$ for some $s_0 > 0$. By similar arguments, (c.2) is satisfied if $\text{E}(\sup_{\theta}\|\mathbf{S}_1\|) < 1$. Using the spectral norm, $\lambda_A \equiv \|\mathbf{A}\| = \lambda^{max}(\mathbf{A}'\mathbf{A})^{1/2}$, where $\lambda^{max}(\cdot)$ is the largest eigenvalue of $\mathbf{A}'\mathbf{A}$ and, because $\left\|\mathbf{I}_{(n)} \otimes \left(\boldsymbol{\epsilon}_{1,1}\boldsymbol{\epsilon}'_{1,1}\right)\right\| = \left(\boldsymbol{\epsilon}'_{1,1}\boldsymbol{\epsilon}_{1,1}\right)$, then uniform invertibility holds if:[9]

$$\lambda_\Gamma + \lambda_\Omega \sum_{i=1}^{k}\text{E}\left(\epsilon^2_{1,[i],1}\right) < 1. \tag{8}$$

Under the restrictions discussed in Section 3.1, the equation (8) can be expressed more explicitly as follows:

i) *Idiosyncratic shocks:* $\lambda_\Gamma = \max_s|\gamma_{ss}|$, $\lambda_\Omega = \max_s|\omega_{ss}|$, $s = 1,\ldots,nk$, where γ_{ss} and ω_{ss} are the non-zero elements of the $nk \times nk$ diagonal parameters matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$;

ii) *Factor spillovers:* $\lambda_\Gamma = \max_i \lambda_i^{max}(\boldsymbol{\Gamma}'_{ii}\boldsymbol{\Gamma}_{ii})^{1/2}$ and $\lambda_\Omega = \max_i \lambda_i^{max}(\boldsymbol{\Omega}'_{ii}\boldsymbol{\Omega}_{ii})^{1/2}$, $i = 1,\ldots,n$, where $\boldsymbol{\Gamma}_{ii}$ and $\boldsymbol{\Omega}_{ii}$ denote the i-th $k \times k$ diagonal blocks of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$;

[9] Noting that $\mathbf{S}_1 = \mathbf{A} - \mathbf{BC}$, by sub-additivity and sub-multiplicativity of $\|\cdot\|$, then $\|\mathbf{S}_1\| = \|\mathbf{A} - \mathbf{BC}\| \leq \|\mathbf{A}\| + \|\mathbf{BC}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|\,\|\mathbf{C}\|$.

iii) *Asset spillovers:* $\lambda_\Gamma = \max_j \lambda_j^{max}(\mathbf{\Gamma}_{jj}^{*\prime}\mathbf{\Gamma}_{jj}^*)^{1/2}$ and $\lambda_\Omega = \max_j \lambda_j^{max}(\mathbf{\Omega}_{jj}^{*\prime}\mathbf{\Omega}_{jj}^*)^{1/2}$, $j = 1, \ldots, k$, where $\mathbf{\Gamma}_{jj}^*$ and $\mathbf{\Omega}_{jj}^*$ denote the corresponding $n \times n$ diagonal blocks of $\mathbf{\Gamma}^*$ and $\mathbf{\Omega}^*$, defined in Section 3.2 and representing the diagonalizations of $\mathbf{\Gamma}$ and $\mathbf{\Omega}$.

Although more explicit than condition (c.2), the conditions above reveals more restrictive because it chains transformations and norm inequalities. In addition, it explicitly requires finite second order moments of $\boldsymbol{\epsilon}_{1,t}$.

3.4 Remarks on alternative parameterizations

As mentioned in the previous sections, in a typical asset pricing applications, $\boldsymbol{\epsilon}_{1,t}$ represents the vector of market-wide risk factors and $\boldsymbol{\epsilon}_{2,t}$ the set of assets. Although in most situations the number of risk factors (k) is small, the number of assets (n) may be large, rising a dimensionality problem. For instance, in the specifications discussed in Section 3.1 the number of parameters amounts to $3nk$ (idiosyncratic shocks), $nk(1+2k)$ (factor spillovers), $nk(1+2n)$ (asset spillovers) and $nk\,(2(n+k)-1)$ (factor and asset spillovers), respectively.

To mitigate the curse of dimensionality we propose the following alternative parameterizations of $\mathbf{B}_t$:

$$\mathbf{B}_t = \mathbf{\Psi} + \mathbf{\Omega}^L(\mathbf{v}_{2,t-1}\mathbf{v}_{1,t-1}')\mathbf{\Omega}^R + \mathbf{\Gamma}^L\mathbf{B}_{t-1}\mathbf{\Gamma}^R, \tag{9}$$

where the left parameters matrices, $\mathbf{\Omega}^L$ and $\mathbf{\Gamma}^L$ are of dimension $n \times n$ and the right parameters matrices $\mathbf{\Omega}^R$ and $\mathbf{\Gamma}^R$ are $k \times k$. Notice that equation (9) can be obtained from (5) by imposing suitable parameter restrictions. Equation (9) encompasses the following parameterizations:

a) $\mathbf{\Omega}^L$, $\mathbf{\Gamma}^L$, $\mathbf{\Omega}^R$ and $\mathbf{\Gamma}^R$ are full matrices. In this case the dynamics accounts for both risk factors and assets spillovers, but subject to parameters restrictions across equations. Further parameter reduction could be obtained by expressing the matrices above as rank one (outer) product of conformable vectors of parameters.

b) $\mathbf{\Omega}^L$, $\mathbf{\Gamma}^L$, $\mathbf{\Omega}^R$ and $\mathbf{\Gamma}^R$ are diagonal matrices. In this case the dynamics depends only on the product of asset and risk factor idiosyncratic innovations and own past betas. The model corresponds to imposing the following restrictions on equation (5): $\mathbf{\Omega} = \text{diag}^{-1}\left(\text{vec}\left(\boldsymbol{\omega}_{(k)}\boldsymbol{\omega}_{(n)}'\right)\right)$ and $\mathbf{\Gamma} = \text{diag}^{-1}\left(\text{vec}\left(\boldsymbol{\gamma}_{(k)}\boldsymbol{\gamma}_{(n)}'\right)\right)$.[10]

[10]This restriction could be limiting in that it impose that the dynamic of the beta associated to $\epsilon_{1,[j+1],t}$ in the

c) $\mathbf{\Omega}^L = \mathbf{\Gamma}^L = \mathbf{I}_{(n)}$, $\mathbf{\Omega}^R$ and $\mathbf{\Gamma}^R$ are diagonal matrices. This model imposes common dynamics for the exposures of all elements of $\boldsymbol{\epsilon}_{2,t}$ to each element of $\boldsymbol{\epsilon}_{1,t}$. More explicitly, $\beta_{1j,t}$, measuring the exposure of $\epsilon_{2,[1],t}$ to $\epsilon_{1,[j],t}$, shares the same dynamics parameters of $\beta_{sj,t}$, $s = 2, \ldots, n$. This model can be equivalently represented by (5) under the restrictions $\mathbf{\Omega} = \text{diag}^{-1}\left(\text{vec}\left(\boldsymbol{\omega}_{(k)}\mathbf{e}'_{(n)}\right)\right)$ and $\mathbf{\Gamma} = \text{diag}^{-1}\left(\text{vec}\left(\boldsymbol{\gamma}_{(k)}\mathbf{e}'_{(n)}\right)\right)$. We refer to this model as *semi-scalar*, in that is a scalar model in the columns of $\mathbf{B}_t$.

d) $\mathbf{\Omega}^L = \mathbf{\Gamma}^L = \mathbf{I}_{(n)}$, while $\mathbf{\Omega}^R = \omega\mathbf{I}_{(k)}$ and $\mathbf{\Gamma}^R = \gamma\mathbf{I}_{(k)}$. This is the most restricted specification, it imposes common dynamics for $\beta_{ij,t}$ for all i and j. This specification can be obtained from equation (5) by setting $\mathbf{\Omega} = \omega\mathbf{I}_{(nk)}$ and $\mathbf{\Gamma} = \gamma\mathbf{I}_{(nk)}$. We refer to this model as *scalar* dynamics.

Stationarity and invertibility conditions for these alterantive parameterizations follow directly from Section 3.2 and 3.3.

3.5 Remarks on devolatilized innovations

In all specifications discussed above, the innovation term in the conditional beta dynamics depends on the products of past orthogonal shocks, $\mathbf{v}_{1,t}$ and $\mathbf{v}_{2,t}$. However, in highly correlated systems, the relative scale of such shocks can turn out be strongly skewed, with the variance of $\mathbf{v}_{2,t}$ shrinking with the magnitude of the correlation between $\boldsymbol{\epsilon}_{2,t}$ and $\boldsymbol{\epsilon}_{1,t}$. To mitigate this effect, but also to reduce the impact of large realizations, which are common in empirical applications, we propose to let the conditional betas depend on products of devolatilized innovations, given by:

$$\begin{bmatrix} \boldsymbol{\xi}_{1,t} \\ \boldsymbol{\xi}_{2,t} \end{bmatrix} = \left(\mathbf{I}_{(nk)} \odot \begin{bmatrix} \mathcal{S}_{11,t} & \mathbf{0} \\ \mathbf{0} & \mathcal{S}_{22,t} \end{bmatrix}\right)^{-1/2} \begin{bmatrix} \mathcal{S}_{11,t} & \mathbf{0} \\ \mathbf{0} & \mathcal{S}_{22,t} \end{bmatrix}^{1/2} \begin{bmatrix} \boldsymbol{\eta}_{1,t} \\ \boldsymbol{\eta}_{2,t} \end{bmatrix}$$
$$= \left(\mathbf{I}_{(nk)} \odot \begin{bmatrix} \mathcal{S}_{11,t} & \mathbf{0} \\ \mathbf{0} & \mathcal{S}_{22,t} \end{bmatrix}\right)^{-1/2} \begin{bmatrix} \mathbf{v}_{1,t} \\ \mathbf{v}_{2,t} \end{bmatrix}, \qquad (10)$$

where $\boldsymbol{\xi}_{1,t}$ and $\boldsymbol{\xi}_{2,t}$ are $k \times 1$ and $n \times 1$ vectors of devolatilized innovations. Although the stationarity conditions discussed in Section 3.2 apply under the the redefinition of the innovations

conditional mean equation of $\epsilon_{2,[i],t}$ is the same as that of the beta associated to $\epsilon_{1,[j],t}$ in the conditional mean equation of $\epsilon_{2,[i+1],t}$. A similar type of restriction holds also in case a).

in $\mathbf{B}_t$, introducing a term of the form $(\boldsymbol{\xi}_{2,t} \otimes \boldsymbol{\xi}_{1,t})$ in (5) results in non explicit invertibility conditions. This is because these products are related to past observations through highly nonlinear recursions.

3.6 Estimation

Consistency and asymptotic normality of the quasi-maximum likelihood (QML) estimator stem directly from Darolles, Francq, and Laurent (2018). The model in equations (4)-(5), with conditional betas (eventually depending on innovations defined in equation (10)) and completed by suitable linear dynamics for $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$, constitutes a block representation of their smallest dimension system, i.e. bivariate. Trivially, in the limit case of a BC-GARCH with $n = k = 1$ the two models coincide.[11]

Endowed with this result, we provide four alternative multi-step estimators computationally more convenient than the full QML. Depending on the model formulation, which shall be discussed on a case by case basis, the likelihood function can be factored, marginalized and/or separated in a number of ways, allowing for different degrees of complexity of the estimation problem. Finite sample performances of these estimators are compared in Section 4 by means of Monte Carlo simulation studies.

Denote $\phi(\boldsymbol{\epsilon}_t; \boldsymbol{\Sigma}_t(\boldsymbol{\theta}), \boldsymbol{\theta})$ the probability density function of a multivariate normally distributed random vector $\boldsymbol{\epsilon}_t$ with covariance matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\theta})$. depending upon a vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$, partitioned according to the block orthogonalization adopted in Section 3. Let $\boldsymbol{\Theta}$ be a compact parameter space which contains the population value of $\boldsymbol{\theta}$.

1) Full quasi-maximum likelihood estimator ($\mathcal{M}_1$). The orthogonalization in equation (4), based on the decomposition of equation (2), entails the factorization of the density:

$$\phi(\boldsymbol{\epsilon}_t; \boldsymbol{\Sigma}_t(\boldsymbol{\theta}), \boldsymbol{\theta}) = \phi\left(\boldsymbol{\epsilon}_{1,t}; \mathcal{S}_{11,t}(\boldsymbol{\theta}_1), \boldsymbol{\theta}_1\right) \phi\left(\boldsymbol{\epsilon}_{2,t} - \mathbf{B}_t \boldsymbol{\epsilon}_{1,t}; \mathcal{S}_{22,t}(\boldsymbol{\theta}_2), \mathbf{B}_t(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2 | \boldsymbol{\epsilon}_{1,t}, \boldsymbol{\theta}_1\right). \tag{11}$$

The QML estimator is defined as the solution of:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{t=1}^{T} \phi\left(\mathbf{v}_{1,t}; \mathcal{S}_{11,t}(\boldsymbol{\theta}_1), \boldsymbol{\theta}_1\right) \phi\left(\mathbf{v}_{2,t}; \mathcal{S}_{22,t}(\boldsymbol{\theta}_2), \mathbf{B}_t(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2 | \boldsymbol{\epsilon}_{1,t}, \boldsymbol{\theta}_1\right). \tag{12}$$

[11] The two models relate in the following way. Starting from a bivariate system, to increase the cross-sectional dimension Darolles, Francq, and Laurent (2018) populate the system prior to the orthogonalization. Contrary, we first operate the orthogonalization of the bivariate system and, then we populate the two blocks.

2) Two-step block-by-block estimator ($\mathcal{M}_2$). The two step-estimator of equations (4)-(5) exploits the factorization in equation (11). Provided that $\mathcal{S}_{11,t}(\boldsymbol{\theta}_1) = \mathcal{S}_{11,t}(\mathbf{v}_{1,t-1}, \mathbf{v}_{1,t-2}, \ldots; \boldsymbol{\theta}_1)$, such that $\partial \mathcal{S}_{11,t}/\partial \boldsymbol{\theta}_2 = \mathbf{0}$ and similarly that $\partial \mathcal{S}_{22,t}/\partial \boldsymbol{\theta}_1 = \mathbf{0}$, then the two-step estimator, given by:

$$\hat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\Theta}}{\text{argmax}} \prod_{t=1}^{T} \phi\left(\mathbf{v}_{1,t}; \mathcal{S}_{11,t}(\boldsymbol{\theta}_1), \boldsymbol{\theta}_1\right), \tag{13}$$

$$\hat{\boldsymbol{\theta}}_2 = \underset{\boldsymbol{\Theta}}{\text{argmax}} \prod_{t=1}^{T} \phi\left(\mathbf{v}_{2,t}; \mathcal{S}_{22,t}(\boldsymbol{\theta}_2), \mathbf{B}_t(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2 | \boldsymbol{\epsilon}_{1,t}, \boldsymbol{\theta}_1\right). \tag{14}$$

coincides with that of $\mathcal{M}_1$.[12] However, if we let $\mathbf{B}_t$ depend on devolatilized innovations, defined in equation (10), then $\partial \mathbf{B}_t/\partial \boldsymbol{\theta}_1 \neq \mathbf{0}$ and the two estimators will be different[13], with the $\mathcal{M}_1$ estimator being in general more efficient.[14]

The following three approaches are based on further decomposition of the likelihood function inspired by Francq and Zakoian (2016). They are appealing when the correlations implied by $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$ are not directly of interest and can be treated as nuisance parameters. This largely reduces the dimension of the parameter space, increasing computational feasibility in large cross-sectional dimensions n. Alternatively, these approaches can be seen as multi-step estimators where, first, conditional means and variances are estimated (jointly or individually), and then, conditioned on the previous steps, correlations are estimated by means of a Gaussian copula.[15] It is important to point out that these approaches, to different extents, limit the type of dynamics that can be assumed for $\mathcal{S}_{11,t}$, $\mathcal{S}_{22,t}$ and $\mathbf{B}_t$ in terms of cross-sectional interactions and constraints. Let us denote $C(\boldsymbol{\xi}_{s,t}; \mathbf{R}_{s,t}(\boldsymbol{\theta}_s^{\mathbf{R}}), \boldsymbol{\theta}_s^{\mathbf{R}})$, $s = 1, 2$, a Gaussian copula, with $\boldsymbol{\xi}_{s,t}$ defined in equation (10), $\mathbf{R}_{s,t}(\boldsymbol{\theta}_s^{\mathbf{R}})$ the conditional correlation, implied by $\mathcal{S}_{ss,t}$, depending on parameters $\boldsymbol{\theta}_s^{\mathbf{R}} \in \boldsymbol{\Theta}$. Also, let us denote $\mathbf{b}_{j,t}$ $j = 1, \ldots, n$ the j-th row of the matrix of conditional betas $\mathbf{B}_t$. Finally, let us denote $\boldsymbol{\theta}_s^-$ the subset of parameters of the marginal distributions of $\mathbf{v}_{s,t}$, such that $\boldsymbol{\theta}_s = \boldsymbol{\theta}_s^- \cup \boldsymbol{\theta}_1^R$.

[12] The parameters of the second block can be estimated without any knowledge of the parameters of the first block, because $\mathbf{B}_t$ as defined in equation (5) depends on $\mathbf{v}_{1,t}$ which is observed.

[13] Notice that this is also the case when $\mathbf{B}_t$ follows the dynamics in equation (5), but the model for $\mathcal{S}_{22,t}$ is such that $\partial \mathcal{S}_{22,t}/\partial \boldsymbol{\theta}_1 \neq \mathbf{0}$.

[14] Examples in a simplified setup can be found in Darolles, Francq, and Laurent (2018).

[15] This approach can be cast in the constant/dynamic conditional correlation framework of Bollerslev (1990) and Engle (2002), where the logarithmic transformation of $C(\boldsymbol{\xi}_{s,t}; \mathbf{R}_{s,t}(\boldsymbol{\theta}_s^{\mathbf{R}}), \boldsymbol{\theta}_s^{\mathbf{R}})$ equals

$$-\frac{1}{2}\Big[\log|\mathbf{R}_{s,t}(\boldsymbol{\theta}_s^{\mathbf{R}})| + \boldsymbol{\xi}_{s,t}' \left(\mathbf{R}_{s,t}(\boldsymbol{\theta}_s^{\mathbf{R}})^{-1} - \mathrm{I}_{(n)}\right) \boldsymbol{\xi}_{s,t}\Big].$$

The joint density can be written as:

$$\phi(\boldsymbol{\epsilon}_t;\boldsymbol{\Sigma}_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \left(\prod_{i=1}^{k}\phi\left(\mathrm{v}_{1,[i],t};\mathcal{S}_{11,[ii],t}(\boldsymbol{\theta}_1^-),\boldsymbol{\theta}_1^-\right)\right) C\left(\boldsymbol{\xi}_{1,t};\mathbf{R}_{1,t}(\boldsymbol{\theta}_1^{\mathbf{R}}),\boldsymbol{\theta}_1^{\mathbf{R}}\right) \times$$

$$\left(\prod_{j=1}^{n}\phi\left(\mathrm{v}_{2,[j],t};\mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-),\mathbf{b}_{j,t}(\boldsymbol{\theta}_2^-),\boldsymbol{\theta}_2^-|\mathbf{v}_{1,t},\boldsymbol{\theta}_1^-\right)\right) C\left(\boldsymbol{\xi}_{2,t};\mathbf{R}_{2,t}(\boldsymbol{\theta}_2^{\mathbf{R}}),\boldsymbol{\theta}_2^{\mathbf{R}}\right). \quad .(15)$$

Note that the factorization above assumes that $\phi\left(\mathrm{v}_{2,[j],t};\mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-),\mathbf{b}_{j,t}(\boldsymbol{\theta}_2^-),\boldsymbol{\theta}_2^-|\mathbf{v}_{1,t},\boldsymbol{\theta}_1^-\right)$, i.e. the marginal distribution of $\mathrm{v}_{2,[j],t}$, $j=1,\ldots,n$, depends only on the parameters of the marginal distributions of $\mathrm{v}_{1,[i],t}$, $i=1,\ldots,k$, and not on $\boldsymbol{\theta}_1^{\mathbf{R}}$.[16]

3) Joint maximization of marginal distributions ($\mathcal{M}_3$). When the correlation parameters are not of interest, the problem can be reduced to:

$$\hat{\boldsymbol{\theta}}^- = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\prod_{t=1}^{T}\prod_{i=1}^{k}\phi\left(\mathrm{v}_{1,[i],t};\mathcal{S}_{11,[ii],t}(\boldsymbol{\theta}_1^-),\boldsymbol{\theta}_1^-\right)\prod_{j=1}^{n}\phi\left(\mathrm{v}_{2,[j],t};\mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-),\mathbf{b}_{j,t}(\boldsymbol{\theta}_2^-),\boldsymbol{\theta}_2^-|\mathbf{v}_{1,t},\boldsymbol{\theta}_1^-\right), \tag{16}$$

where $\boldsymbol{\theta}^-$ denotes a reduced dimension parameter vector with respect to $\boldsymbol{\theta}$.

4) Two-step estimation based on block-by-block joint maximization of marginals ($\mathcal{M}_4$). Provided that the conditions on $\mathcal{S}_{11,t}$, $\mathcal{S}_{22,t}$ and $\mathbf{B}_t$ discussed in $\mathcal{M}_2$ are satisfied, then we can define:

$$\hat{\boldsymbol{\theta}}_1^- = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\prod_{t=1}^{T}\prod_{i=1}^{k}\phi\left(\mathrm{v}_{1,[i],t};\mathcal{S}_{11,[ii],t}(\boldsymbol{\theta}_1^-),\boldsymbol{\theta}_1^-\right) \tag{17}$$

$$\hat{\boldsymbol{\theta}}_2^- = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\prod_{t=1}^{T}\prod_{j=1}^{n}\phi\left(\mathrm{v}_{2,[j],t};\mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-),\mathbf{b}_{j,t}(\boldsymbol{\theta}_2^-),\boldsymbol{\theta}_2^-|\mathbf{v}_{1,t},\boldsymbol{\theta}_1^-\right). \tag{18}$$

As previously discussed, $\mathcal{M}_4$ is equivalent to $\mathcal{M}_3$ whenever $\partial\mathbf{B}_t/\partial\boldsymbol{\theta}_1=\mathbf{0}$. It will differ when $\mathbf{B}_t$ depends on devolatilized innovations, in general entailing a loss of efficiency.

5) Estimation equation-by-equation via marginal distributions ($\mathcal{M}_5$). This method relies upon the maximal level of separation of the likelihood. It splits the maximization problem in kn individual optimizations:

$$\hat{\boldsymbol{\theta}}_i^- = \begin{cases} \operatorname{argmax}_{\boldsymbol{\Theta}}\prod_{t=1}^{T}\phi\left(\mathrm{v}_{i,t};\varsigma_{i,t}(\boldsymbol{\theta}_i^-),\boldsymbol{\theta}_i^-\right), & i=1,\ldots,k \\ \operatorname{argmax}_{\boldsymbol{\Theta}}\prod_{t=1}^{T}\phi\left(\mathrm{v}_{i,t};\varsigma_{i,t}(\boldsymbol{\theta}_i^-),\mathbf{b}_{i-k,t}(\boldsymbol{\theta}_i^-),\boldsymbol{\theta}_i^-|\mathbf{v}_{1,t},\boldsymbol{\theta}_1^-\right), & i=k+1,\ldots,k+n, \end{cases} \tag{19}$$

[16] This assumption allows for a reduction of the complexity of the estimation problem or a reduction of the parameter space without imposing limiting parameter constraints.

where $\mathrm{v}_{i,t}$ is the i-th element of the $kn \times 1$ vector $\mathbf{v}_t = (\mathbf{v}_{1,t}, \mathbf{v}_{2,t})'$ and $\varsigma_{i,t}$ is the i-th element of $(\mathrm{diag}(\mathcal{S}_{11,t}), \mathrm{diag}(\mathcal{S}_{22,t}))'$. This method is the most computationally efficient in large cross-sectional dimensions. By breaking the estimation in smaller problems it reduces the parameter space and avoids the inversion of large dimension matrices.

To be feasible, this approach requires that the dynamics of $\mathcal{S}_{11,t}$, $\mathcal{S}_{22,t}$ and $\mathbf{B}_t$ entail minimal or no cross-sectional parameter restrictions or interactions between elements. For instance, conditional betas driven by idiosyncratic shocks or factor spillovers can be accommodated, while dynamics allowing for asset spillovers cannot.[17] Similar arguments hold for the conditional covariance matrices $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$.

As mentioned earlier, $\mathcal{M}_3$ to $\mathcal{M}_5$ can be used as a preliminary step prior to the estimation of the conditional covariances (or correlations) implied by $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$. The off-diagonal elements of $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$ can be obtained by filtering, e.g. if $\mathcal{S}_{ii,t}$ is a scalar or diagonal BEKK, or they can be estimated, by means of a constant or dynamic correlation model, see Bollerslev (1990), Engle (2002) and Aielli (2013) among others.

It is also worth noting that, despite covariance targeting being unfeasible in general for $\mathcal{S}_{22,t}$, knowledge of $\mathrm{diag}(\mathcal{S}_{22})$, allows targeting its off-diagonal elements by means of sample estimators using first step residuals, e.g. $\hat{\mathcal{S}}_{22,[ij]} = T^{-1}\sum_{t=1}^{T} \hat{v}_{2,[i],t}\hat{v}_{2,[j],t}$, $i, j = 1, \ldots, n$ and $i \neq j$. Although these methods reduce the number of parameters[18] and the complexity of the estimation,[19] they come at the cost of a loss of efficiency. This point will be addressed in detail in Section 4 by means of a Monte Carlo simulation study.

4 Monte Carlo study

The finite sample performance of the Gaussian maximum likelihood estimator is evaluated by means of two Monte Carlo exercise. The model setup is the following.

[17]When $\mathbf{B}_t$ depends on devolatilized innovations, the estimation of the parameters of the marginal distributions of $\epsilon_{2,[j],t}$ $j = 1, \ldots, n$ requires knowledge of $\mathcal{S}_{11,[ii],t}$, $i = 1, \ldots, k$ which need to be estimated first, imposing sequentiality of the estimation between the two blocks (but not within).

[18]In the most parsimonious specification the number of parameters of multivariate GARCH is of order $O(n^2)$, which reduces to $O(n)$ in the case of separability and to $O(1)$ in the case of marginalization, where n is the dimension of the system being modeled.

[19]While the computation of the full likelihood requires inverting covariance matrices, that of likelihood under separability and marginalization requires computing only (scalar) fractions and sums.

The cross-sectional dimension of the system is set to five, with dimensions of the blocks $k = 3$ and $n = 2$ respectively. The sample sizes are $T = \{3750, 7500, 15000\}$.

The model specification is given by equation (4) with $\boldsymbol{\eta}_t \sim$ i.i.d $\mathrm{N}(\mathbf{0}_{(5)}, \mathbf{I}_{(5)})$. The matrix of conditional betas $\mathbf{B}_t$ follows the simple dynamics driven by idiosyncratic shocks expressed as in equation (5.i). The intercept term is reparameterized using the unconditional beta, $\boldsymbol{\Psi} = \mathbf{B} \odot (\mathbf{e}_{(2)}\mathbf{e}_{(3)}^{'} - \boldsymbol{\Gamma})$. To assess the use of devolatilized shocks, the driving innovation in the conditional betas depends on $(\boldsymbol{\xi}_{1,t}, \boldsymbol{\xi}_{2,t})'$ defined in equation (10). The dynamics of $\mathbf{B}_t$ is thus expressed as:

$$\mathbf{B}_t = \mathbf{B} \odot (\mathbf{e}_{(2)}\mathbf{e}_{(3)}^{'} - \boldsymbol{\Gamma}) + (\boldsymbol{\Omega} \odot \boldsymbol{\xi}_{2,t-1}\boldsymbol{\xi}_{1,t-1}') + (\boldsymbol{\Gamma} \odot \mathbf{B}_{t-1}). \tag{20}$$

The 2×3 matrices of parameters $\mathbf{B}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$, are set as follows:

$$\mathbf{B} = \begin{bmatrix} 1.00 & 0.50 & 0.70 \\ 1.00 & -0.50 & 0.00 \end{bmatrix}, \qquad \omega_{ij} = 0.04, \qquad \gamma_{ij} = 0.95, \qquad \forall\, i = 1,2 \text{ and } j = 1,2,3.$$

Although, this parameterization implies that all betas share common dynamics, the $2kn$ elements of $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ are estimated individually, without imposing equality restrictions. Note that for $\beta_{23,t}$, the dynamics above together with $\beta_{23} = 0$ imply that $\epsilon_{2,[2],t}$ and $\epsilon_{1,[3],t}$ are conditionally correlated but unconditionally orthogonal.

The model is completed by scalar BEKK dynamics for $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$, with intercept reparameterized via covariance tracking:

$$\mathcal{S}_{ii,t} = \mathcal{S}_{ii}(1 - \tau_i - \delta_i) + \tau_i(\mathbf{v}_{i,t}\mathbf{v}_{i,t}') + \delta_i \mathcal{S}_{ii,t-1}, \qquad i = 1,2, \tag{21}$$

with parameters

$$\mathcal{S}_{11} = \begin{bmatrix} 1.00 & 0.10 & -0.10 \\ 0.10 & 1.00 & 0.10 \\ -0.10 & 0.10 & 1.00 \end{bmatrix}, \quad \mathcal{S}_{22} = \begin{bmatrix} 1.00 & 0.50 \\ 0.50 & 1.00 \end{bmatrix}, \quad \tau_i = 0.04, \quad \delta_i = 0.95, \quad i = 1,2.$$

This choice of parameters implies the following unconditional covariance matrix of $\boldsymbol{\epsilon}_t$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1.00 & 0.10 & -0.10 & 0.98 & 0.95 \\ 0.10 & 1.00 & 0.10 & 0.67 & -0.4 \\ -0.10 & 0.10 & 1.00 & 0.65 & -0.15 \\ 0.98 & 0.67 & 0.65 & 2.77 & 1.15 \\ 0.95 & -0.40 & -0.15 & 1.15 & 2.15 \end{bmatrix}.$$

The scalar BEKK, a linear model with common dynamics and absence of spillovers,[20] ensures comparability of the estimator $\mathcal{M}_1$ with the alternative estimators based on different degrees of marginalization and decomposition of the likelihood discussed in Section 3.6.

Because for the BEKK model the properties of the maximum likelihood estimator under covariance targeting are well known, for the sake of feasibility, $\mathcal{S}_{11}$ is estimated prior to the other parameters by means of the plug-in sample variance estimator $T^{-1}(\mathbf{v}_{1,t}\mathbf{v}_{1,t}')$. To ensure positive definiteness of the second block variance, $\mathcal{S}_{22}$ is estimated via triangular decomposition $\mathbf{CC}'$ where $\mathbf{C}$ is a lower triangular matrix. Although the the elements of C are estimated by maximum likelihood, results are reported for $\mathcal{S}_{22} = \mathbf{CC}'$. The number of Monte Carlo replications is $s = 1500$.

In the first simulation we contrast $\mathcal{M}_1$ against the four multi-step estimators characterized by different layers of likelihood decomposition as discussed in Section 3.6. The competing estimators with their associated vector of parameters are:

$\mathcal{M}_1$*:(Full quasi-maximum likelihood)*

$$\boldsymbol{\theta} = (\tau_1, \delta_1, \text{vec}(\mathbf{B}), \text{vec}(\boldsymbol{\Omega}), \text{vec}(\boldsymbol{\Gamma}), \text{vech}(\mathcal{S}_{22}), \tau_2, \delta_2)'.$$

The parameter vector has dimension $(3kn + n(n+1)/2 + 4)$ which for $k = 3$ and $n = 2$ amounts to 25 parameters.

$\mathcal{M}_2$*:(Two-step block-by-block estimation)*: the parameter vector associated to the first block $\boldsymbol{\theta}_1 = (\tau_1, \delta_1)'$ has dimension 2, while

$$\boldsymbol{\theta}_2 = (\text{vec}(\mathbf{B}), \text{vec}(\boldsymbol{\Omega}), \text{vec}(\boldsymbol{\Gamma}), \text{vech}(\mathcal{S}_{22}), \tau_2, \delta_2)',$$

[20]The conditional variance of the i-th marginal distributions of the j-th block follows a GARCH(1,1) with parameters (τ_j, δ_j), $j = 1, 2$, common across series in each block.

which denotes the second block parameters, has dimension $3kn + n(n + 1)/2 + 2$. In our setup the latter amounts to 23 parameters.

$\mathcal{M}_3$*:(joint maximization of marginal distributions):*

$$\boldsymbol{\theta}^- = (\tau_1, \delta_1, \text{vec}(\mathbf{B}), \text{vec}(\boldsymbol{\Omega}), \text{vec}(\boldsymbol{\Gamma}), \text{diag}(\mathcal{S}_{22}), \tau_2, \delta_2)'.$$

The parameter vector has dimension $(3k+1)n+4$, which in our setup amounts to 24 parameters.

$\mathcal{M}_4$*:(Two-step estimation based on block-by-block joint maximization of marginals):* the first block parameter vector is $\boldsymbol{\theta}_1^- = (\tau_1, \delta_1)'$, while that of the second block is:

$$\boldsymbol{\theta}_2^- = \text{vec}(\mathbf{B}), \text{vec}(\boldsymbol{\Omega}), \text{vec}(\boldsymbol{\Gamma}), \text{diag}(\mathcal{S}_{22}), \tau_2, \delta_2)',$$

and has dimension $(3k + 1)n + 2$, which in our setup amounts to 22 parameters. Note that in $\mathcal{M}_3$ and $\mathcal{M}_4$ the equality restriction on the parameters governing the dynamics of the marginal variances $\mathcal{S}_{11,[ii],t}$ $i = 1, \ldots, k$ and $\mathcal{S}_{22,[jj],t}$ $j = 1, \ldots, n$, is binding.

$\mathcal{M}_5$*:(Estimation equation-by-equation via marginal distributions)* the parameter vector $\boldsymbol{\theta}_i^-$ equals to:

$$\boldsymbol{\theta}_i^- = \begin{cases} (\tau_{1,i}, \delta_{1,i})' & i = 1, \ldots, k, \\ (\mathbf{b}_{i-k}, \boldsymbol{\omega}_{i-k}, \boldsymbol{\gamma}_{i-k}, \varsigma_{ii}, \tau_{2,i}, \delta_{2,i})' & i = k + 1, \ldots, k + n. \end{cases}$$

Note that for $i > k$ the dimension of $\boldsymbol{\theta}_i^-$ is $(3k + 3)$, in our case 12 parameters for each marginal distribution of the second block. The total number of parameters is 30, which entails some redundancy due to the fact that this estimation approach does not impose a priori the equality restrictions on the parameters of the marginal variances, i.e. $(\tau_{1,i}, \delta_{1,i})' = (\tau_1, \delta_1)'$, $i = 1, \ldots, k$ and similarly $(\tau_{2,j}, \delta_{2,j})' = (\tau_2, \delta_2)'$, $j = 1, \ldots, n$.

For $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_5$ to be feasible and comparable to $\mathcal{M}_1$ and $\mathcal{M}_2$, the model must allow for separability of the likelihood. This is the case in our setting because of the absence of volatility/covariance and, beta spillovers.

Table 1 reports the the Monte Carlo results. To save space, element-wise Bias and Root Mean Squared Error (RMSE) for the first block plug-in estimator of the unconditional covariance $\mathcal{S}_{11}$, common to all simulations in this section, are reported only once in the top portion of Table 2. Due to the large number of parameters, we report average empirical bias and the square root of the average mean squared error across the number of conditional betas ($kn = 6$) for the elements

Table 1: *Comparison of estimation approaches: Bias and RMSE for the parameters in* $\mathbf{B}_t$ *and* $\mathcal{S}_{ii,t}$, $i = 1, 2$, *defined by equation (20) and equation (21) respectively. First block unconditional covariance* $\mathcal{S}_{11}$ *targeted to* $T^{-1}(\mathbf{v}_{1,t}\mathbf{v}'_{1,t})$.

	Bias					RMSE				
	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
					$T = 3750$					
τ_1	-0.0002	-0.0002	-0.0001	-0.0001	-0.0001	0.0026	0.0026	0.0036	0.0036	0.0064
δ_1	-0.0005	-0.0006	-0.0008	-0.0008	-0.0019	0.0035	0.0035	0.0049	0.0051	0.0092
β_{ij}	0.0007	0.0006	0.0006	0.0004	-0.0003	0.0285	0.0283	0.0292	0.0299	0.0296
ω_{ij}	0.0001	0.0002	0.0003	0.0005	0.0003	0.0071	0.0072	0.0081	0.0082	0.0080
γ_{ij}	-0.0043	-0.0047	-0.0055	-0.0058	-0.0050	0.0181	0.0183	0.0212	0.0214	0.0208
$\mathcal{S}_{22,[ii]}$	-0.0068	-0.0047	-0.0025	-0.0004	-0.0023	0.1085	0.1069	0.1149	0.1181	0.1237
$\mathcal{S}_{22,[12]}$	-0.0061	-0.0034				0.0793	0.0775			
τ_2	0.0003	0.0003	0.0004	0.0000	-0.0001	0.0037	0.0036	0.0048	0.0048	0.0063
δ_2	-0.0015	-0.0016	-0.0020	-0.0013	-0.0018	0.0052	0.0052	0.0068	0.0065	0.0088
					$T = 7500$					
τ_1	-0.0001	-0.0001	0.0000	0.0000	-0.0001	0.0019	0.0019	0.0026	0.0026	0.0044
δ_1	-0.0003	-0.0003	-0.0005	-0.0005	-0.0009	0.0025	0.0025	0.0034	0.0035	0.0061
β_{ij}	-0.0000	-0.0003	-0.0001	0.0001	0.0002	0.0202	0.0203	0.0207	0.0207	0.0209
ω_{ij}	0.0000	0.0001	-0.0001	0.0001	0.0000	0.0049	0.0049	0.0057	0.0056	0.0056
γ_{ij}	-0.0021	-0.0023	-0.0022	-0.0022	-0.0024	0.0116	0.0116	0.0130	0.0132	0.0132
$\mathcal{S}_{22,[ii]}$	-0.0009	-0.0047	-0.0010	0.0003	-0.0016	0.0779	0.0773	0.0818	0.0807	0.0872
$\mathcal{S}_{22,[12]}$	-0.0020	-0.0037				0.0555	0.0563			
τ_2	0.0002	-0.0000	-0.0000	0.0002	0.0000	0.0026	0.0027	0.0034	0.0034	0.0046
δ_2	-0.0008	-0.0006	-0.0006	-0.0008	-0.0009	0.0035	0.0036	0.0043	0.0046	0.0061
					$T = 15000$					
τ_1	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0013	0.0014	0.0019	0.0019	0.0032
δ_1	-0.0001	-0.0002	-0.0002	-0.0002	-0.0005	0.0017	0.0018	0.0024	0.0025	0.0043
β_{ij}	-0.0000	-0.0001	-0.0004	-0.0002	-0.0003	0.0142	0.0140	0.0144	0.0145	0.0146
ω_{ij}	0.0000	0.0000	0.0000	0.0001	0.0001	0.0034	0.0035	0.0039	0.004	0.0039
γ_{ij}	-0.0010	-0.0010	-0.0013	-0.0013	-0.0011	0.0076	0.0076	0.0092	0.0088	0.0088
$\mathcal{S}_{22,[ii]}$	0.0022	0.0000	0.0008	0.0008	0.0001	0.0574	0.0557	0.0597	0.0595	0.0617
$\mathcal{S}_{22,[12]}$	0.0009	0.0008				0.0408	0.0401			
τ_2	0.0001	0.0001	0.0001	-0.0001	-0.0001	0.0018	0.0019	0.0024	0.0023	0.0031
δ_2	-0.0004	-0.0005	-0.0005	-0.0003	-0.0004	0.0024	0.0024	0.0032	0.0031	0.0041

of $\mathbf{B}, \mathbf{\Omega}$ and $\mathbf{\Gamma}$ (denoted β_{ij}, ω_{ij} and γ_{ij}), as well as across the number of series ($n = 2$) for the diagonal elements of $\mathcal{S}_{22}$ (denoted $\mathcal{S}_{22,[ii]}$). For $\mathcal{M}_5$, average statistics across the number of series ($k = 3$ and $n = 2$ respectively) are reported also for the parameters $(\tau_{1,i}, \delta_{1,i})'$ and $(\tau_{2,j}, \delta_{2,j})'$. For all estimation methods, empirical biases are very small for all sample sizes and, RMSE decay at the appropriate square root rate for all parameters. Thus using devolatilized innovations to drive the conditional beta dynamics has no detrimental impact on the statistical properties of

Table 2: *$\mathcal{M}_2$ (first step): Bias and RMSE for the parameters in $\mathcal{S}_{11,t}$, defined by equation (21) with unconditional covariance $\mathcal{S}_{11}$ targeted to $T^{-1}(\mathbf{v}_{1,t}\mathbf{v}'_{1,t})$.*

	Bias			RMSE		
T	3750	7500	15000	3750	7500	15000
$\mathcal{S}_{11,[11]}$	0.0024	-0.0017	0.0007	0.1259	0.0890	0.0624
$\mathcal{S}_{11,[12]}$	0.0023	-0.0001	-0.0001	0.0879	0.0604	0.0444
$\mathcal{S}_{11,[13]}$	0.0028	-0.0000	-0.0001	0.0858	0.0605	0.0426
$\mathcal{S}_{11,[22]}$	-0.0030	-0.0016	0.0007	0.1219	0.0876	0.0615
$\mathcal{S}_{11,[23]}$	-0.0015	0.0011	-0.0006	0.0837	0.0617	0.0435
$\mathcal{S}_{11,[33]}$	-0.0023	-0.0010	0.0006	0.1240	0.0909	0.0627
τ_1	-0.0002	-0.0001	-0.0000	0.0026	0.0019	0.0013
δ_1	-0.0006	-0.0003	-0.0001	0.0035	0.0025	0.0017

the maximum likelihood estimator. The loss of efficiency between $\mathcal{M}_1$ and $\mathcal{M}_2$ is marginal, due to the small number of parameters estimated in the first step (τ_1 and δ_1) which are estimated to a high degree of accuracy. With a more parameterized first step and a larger dimension k, the loss of efficiency is likely to become more pronounced.[21]

The largest loss of efficiency is observed when estimating the model's parameter using only the marginal distributions, whether jointly or individually, i.e. $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_5$.

The parameters τ_1 and δ_1, show the largest drop of efficiency, namely between 32% and 40% under joint estimation ($\mathcal{M}_3$ and $\mathcal{M}_4$) and between 129% and 162% under the estimation equation-by-equation ($\mathcal{M}_5$), compared to $\mathcal{M}_1$ and $\mathcal{M}_2$. The parameters τ_2 and δ_2 show a loss of efficiency between 20% and 32% under joint estimation and 68% and 72% under the estimation equation-by-equation. The parameters $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ exhibit a loss of efficiency between 13% and 17% when using $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_5$. For the elements of $\mathbf{B}$ and $\mathcal{S}_{22}$ we find only marginal losses of efficiency. This is because these parameters represent unconditional levels, thus largely benefit from compensation of estimation error (compared to estimating intercepts) and from the fairly large sample sizes T.

Understanding the extent of the loss of efficiency when treating the conditional covariances/correlations as nuisance parameters is crucial. For the first block, being covariance targeting feasible, the gain

[21] In our setting, the loss of efficiency is likely to be hidden by sampling variability of the statistics reported, because of the limited cross-sectional dimension, the small number of parameters estimated in the first step and the contained number of Monte Carlo replications.

in efficiency stemming from estimating the conditional covariances/correlations comes at no or little cost in terms of computational feasibility.[22] However, as pointed out in Section 3.6, being the covariance targeting in general unfeasible for $\mathcal{S}_{22,t}$, the estimation via marginalization dramatically reduces the number of parameters, in this case from $O(n^2)$ to $O(n)$. This substantially increases the computational feasibility when the dimension of the second block n is large.

The aim of the second Monte Carlo simulation exercise is to contrast the performance of a fully unconstrained estimator against the case where the unconditional beta, $\mathbf{B}$, is targeted to the sample least square estimator. Building on the results of the previous Monte Carlo simulation, without loss of generality, parameters are estimated using $\mathcal{M}_2$. This method strikes a good balance between statistical and computational efficiency. Bias and RMSE for the first step parameters is reported in Table 2 for completeness.

Table 3 and 4 report bias, RMSE and relative efficiency for the second step parameters. The panel entitled MLE corresponds to the case where $\mathbf{B}$ is estimated by maximum likelihood jointly with the remaining parameters. The panel entitled OLS reports results for the case of beta tracking where the level of $\mathbf{B}_t$ is targeted to the sample least square estimate.

Empirical biases are close to zero for all parameters. Although unbiased for $\mathbf{B}$, the least square estimator reveals, on average, a 23% efficiency loss compared to the maximum likelihood estimator. The relative performance of the other parameters of the model remains unaffected by the beta targeting.

5 Data and empirical applications

In this section we carry out two empirical applications. The first tests the BC-GARCH in the linear three factor asset pricing model introduced by Fama and French (1992) (3FF). The aim is to benchmark a conditional beta specification driven only by idiosyncratic shocks, against three alternative specifications accounting for beta spillovers. The set of investment assets consist of the excess returns on two US industry portfolios, namely Coal (C) and Petroleum-Natural Gas

[22]This result holds provided that the assumed dynamics for $\mathcal{S}_{11,t}$ allows for unbiased and consistent targeting of the conditional covariance matrix to a sample variance estimator. This is the case for linear multivariate GARCH, VEC, and BEKK models and some non-linear models, the constant conditional correlation model of Bollerslev (1990) and the dynamic conditional correlation of Aielli (2013). In turn, for example, targeting is not feasible for the dynamic correlation model of Engle (2002), see Aielli (2013).

Table 3: *Bias for the parameters in* $\mathbf{B}_t$ *and* $\mathcal{S}_{22,t}$ *estimated using* $\mathcal{M}_2$ *(second step) without (MLE) and with (OLS) beta targeting. For* $\mathbf{\Omega}$ *and* $\mathbf{\Gamma}$ *average bias and square root of the average mean squared error over their elements are reported in the rows denoted* ω_{ij} *and* γ_{ij}*.*

Bias						
	MLE			OLS		
T	3750	7500	15000	3750	7500	15000
β_{11}	-0.0007	-0.0001	-0.0006	-0.0001	-0.0006	-0.0003
β_{21}	-0.0001	-0.0008	-0.0003	0.0004	-0.0003	0.0001
β_{12}	0.0015	-0.0005	0.0001	0.0015	-0.0012	-0.0004
β_{22}	0.0014	-0.0000	-0.0006	0.0017	-0.0010	0.0001
β_{13}	-0.0000	-0.0003	-0.0001	-0.0001	0.0012	-0.0003
β_{23}	0.0007	-0.0004	0.0000	0.0013	0.0008	0.0003
ω_{ij}	0.0001	0.0001	0.0000	0.0001	0.0001	-0.0001
γ_{ij}	-0.0042	-0.0012	-0.0010	-0.0041	-0.0017	-0.0008
$\mathcal{S}_{22,[11]}$	-0.0091	-0.0041	0.0008	-0.0095	-0.0034	0.0007
$\mathcal{S}_{22,[12]}$	-0.0027	-0.0007	0.0004	-0.0036	0.0001	0.0006
$\mathcal{S}_{22,[22]}$	-0.0001	-0.0001	-0.0012	-0.0017	-0.0005	0.0013
τ_2	0.0002	0.0000	0.0001	0.0002	0.0000	0.0000
δ_2	-0.0016	-0.0007	-0.0004	-0.0015	-0.0007	-0.0003

Table 4: *RMSE for the parameters in* $\mathbf{B}_t$ *and* $\mathcal{S}_{22,t}$ *estimated using* $\mathcal{M}_2$ *(second step) without (MLE) and with (OLS) beta targeting. For* $\mathbf{\Omega}$ *and* $\mathbf{\Gamma}$ *average bias and square root of the average mean squared error over their elements are reported in the rows denoted* ω_{ij} *and* γ_{ij}*.*

RMSE									
	MLE			OLS			Relative efficiency		
T	3750	7500	15000	3750	7500	15000	3750	7500	15000
β_{11}	0.0284	0.0201	0.0139	0.0353	0.0246	0.0176	0.25	0.22	0.26
β_{21}	0.0278	0.0202	0.0140	0.0342	0.0243	0.0172	0.23	0.20	0.23
β_{12}	0.0279	0.0203	0.0139	0.0343	0.0255	0.0171	0.23	0.26	0.23
β_{22}	0.0274	0.0205	0.0144	0.0339	0.0258	0.0171	0.23	0.26	0.19
β_{13}	0.0284	0.0196	0.0139	0.0341	0.0241	0.0177	0.20	0.23	0.27
β_{23}	0.0283	0.0201	0.0141	0.0344	0.0239	0.0177	0.21	0.19	0.26
ω_{ij}	0.0029	0.0020	0.0014	0.0029	0.0020	0.0014	-0.00	-0.01	0.00
γ_{ij}	0.0074	0.0045	0.0031	0.0074	0.0045	0.0031	-0.01	-0.01	-0.00
$\mathcal{S}_{22,[11]}$	0.1046	0.0769	0.0549	0.1046	0.0784	0.0544	0.00	0.02	-0.01
$\mathcal{S}_{22,[12]}$	0.0781	0.0567	0.0400	0.0779	0.0555	0.0392	-0.00	-0.02	-0.02
$\mathcal{S}_{22,[22]}$	0.1127	0.0790	0.0551	0.1124	0.0787	0.0546	-0.00	-0.00	-0.01
τ_2	0.0037	0.0027	0.0018	0.0037	0.0027	0.0018	0.00	-0.01	-0.00
δ_2	0.0052	0.0036	0.0024	0.0052	0.0036	0.0024	0.00	0.00	-0.01

(P). The three risk factors are: the market factor ($f_{mkt,t}$), given by the excess return of the value-weight portfolio formed by all US firms listed on the NYSE, AMEX, or NASDAQ; the size factor (small-minus-big, $f_{smb,t}$) given by a long-short self-financing portfolio of value weighted returns sorted by size; and the value factor (high-minus-low, $f_{hml,t}$) given by a long-short self-financing portfolio of value-weighted returns sorted by book-to-market ratio. The sample spans the period from January 1, 1927 to October, 30 2020, totaling 24682 daily observations. The data is obtained from Kenneth French's website data library.[23] The three factors and the two industry portfolios log-returns in excess of the risk-free rate are plotted in Figures 1 and 2, respectively.

In the figures are visible the long lasting volatility cluster of the Great depression (1929-1939), the effects of the oil shocks in the 70s, the drop of the Black Monday in October 1987, the turmoil of late 90s-early 2000s (1997 Asian financial crisis, 1998 Russian financial crisis, 2000 dot-com bubble burst, 2001 9/11 events, 2002 US stock market downturn), the 2007-8 global financial crisis followed by the European sovereign debt crisis, and, last, the 2020 COVID-19 pandemic. In Figure 2, it is also visible the 2015-16 Chinese stock market crash, which set off a global stock market sell-off and a steep drop in commodities prices.

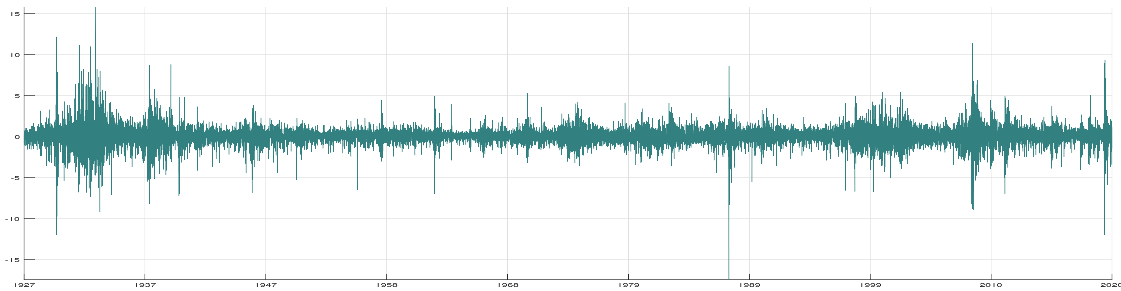
The second application uses the BC-GARCH to estimate the risk premia attached to risk factors exposures. We test the linear conditional beta pricing model used in Fama and MacBeth (1973) in the CAPM and 3FF frameworks. We benchmark the BC-GARCH against the 2PR approach of Fama and MacBeth (1973). The cross-section of assets consists of returns on 40 value-weighed industry portfolios.[24] For sake of comparability with the existing literature, we use a monthly aggregation frequency, totaling 1126 observations.

5.1 Beta Spillovers in the Coal and Petroleum-Natural Gas industry portfolios

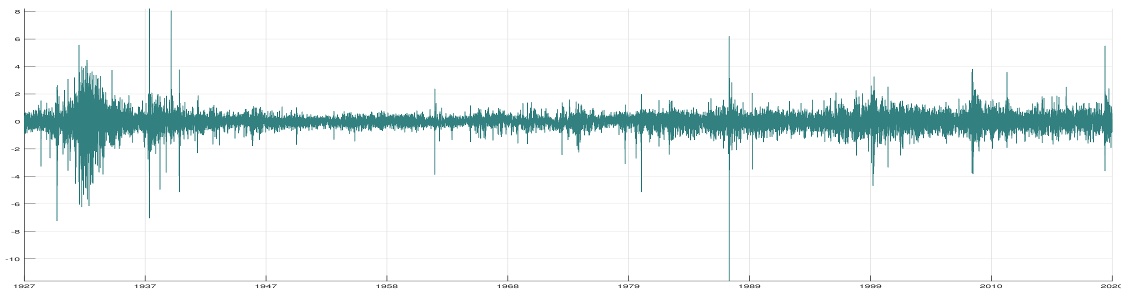
The two-blocks partition of equation (4) lends itself to the linear asset pricing context, naturally fitting the distinction between market-wide risk factors and investment assets.

[23]The composition of the two industry portfolios based on their four-digit SIC code is available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/Siccodes30.zip. The two portfolios are number 18 and 19 respectively of the list of 30 industry portfolios. The dataset containing daily returns on the industry portfolios and the the factors is available at http://mba.tuck.dartmouth.edu/pages/ faculty/ken.french/data_library.html

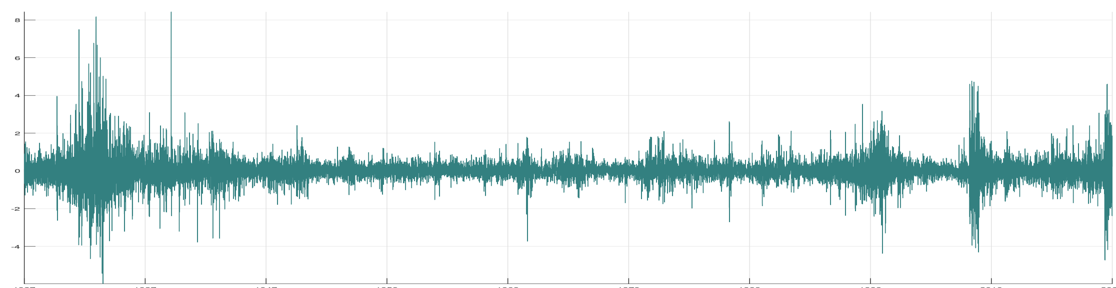
[24]The data used is a subset extracted from the collection of 48 industry portfolios available at Kenneth French's data library. We exclude 8 portfolios, 3, 11, 15, 20, 26, 27, 33, 38, for which returns are available only over a shorter time span.



(a) *Market factor* ($f_{mkt,t}$)



(b) *Size factor* ($f_{smb,t}$)



(c) *Value factor* ($f_{hml,t}$)

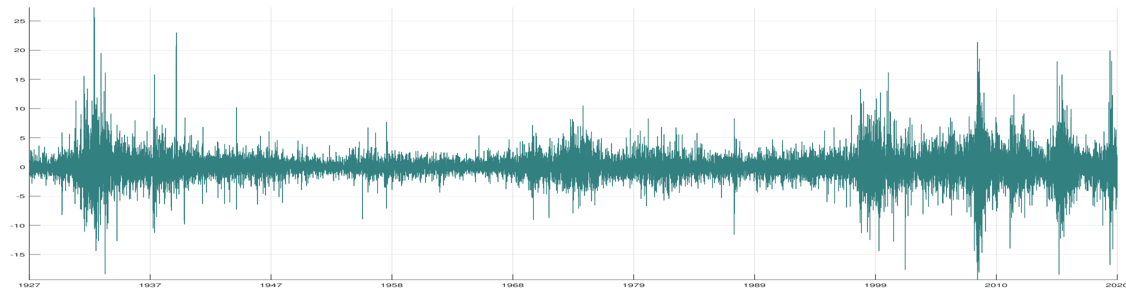
Figure 1: *Fama-French risk factors: (a) market factor* ($f_{mkt,t}$)*, (b) size factor* ($f_{smb,t}$) *and (c) value factor* ($f_{hml,t}$)*. Log-returns expressed in percentage.*

Let $(\mathbf{f}_t, \mathbf{r}_t)'$ be the vector of $k = 3$ factors, $(f_{mkt,t}, f_{smb,t}, f_{hml,t})'$ and $n = 2$ assets, $(r_{C,t}, r_{P,t})'$, with $\mathrm{E}(\mathbf{f}_t|\mathcal{I}_{t-1}) = \mathrm{E}(\mathbf{f}_t) = \boldsymbol{\mu}_f$ and $\mathrm{E}(\mathbf{r}_t|\mathcal{I}_{t-1}) = \mathrm{E}(\mathbf{r}_t) = \boldsymbol{\mu}_r$. According with equation (4), we consider the following model:

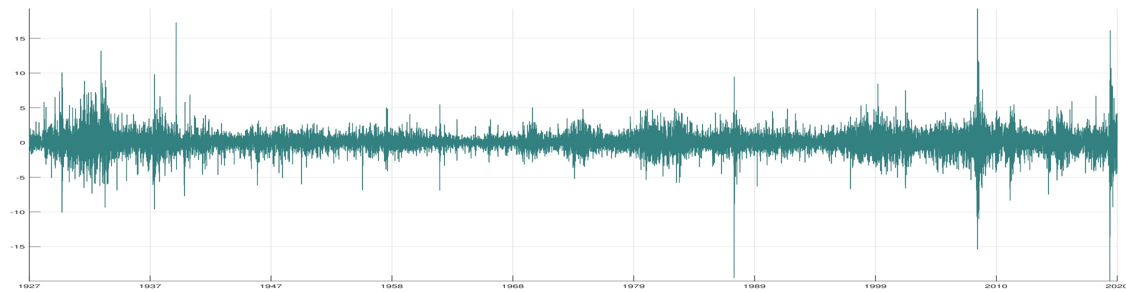
$$\mathbf{f}_t = \boldsymbol{\mu}_f + \mathbf{v}_{f,t} \tag{22}$$

$$\mathbf{r}_t = \boldsymbol{\mu}_r + \mathbf{B}_t(\mathbf{f}_t - \boldsymbol{\mu}_f) + \mathbf{v}_{r,t}. \tag{23}$$

Equation (23) represents the familiar 3FF regression but with time-varying coefficients $\mathbf{B}_t$, where the typical element $\beta^i_{j,t}$ $i = C, P$ and $j = mkt, smb, hml$ denotes the conditional beta measuring



(a) Coal industry portfolio excess return ($r_{C,t}$)



(b) Petroleum-Natural Gas industry portfolio excess return ($r_{P,t}$)

Figure 2: *Investment assets: (a) Coal ($r_{C,t}$) and (b) Petroleum-Natural gas ($r_{P,t}$). Log-returns in excess of the risk free rate expressed in percentage.*

the exposure of asset i to factor j.[25] Simple algebra shows that equation (23) can be equivalently reparameterized as:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}_t \mathbf{f}_t - (\mathbf{B}_t - \mathbf{B})\boldsymbol{\mu}_f + \mathbf{v}_{r,t}, \tag{23.b}$$

where $\boldsymbol{\alpha} = \boldsymbol{\mu}_r - \mathbf{B}\boldsymbol{\mu}_f$, represents the usual "unconditional" Jensen's alpha in the 3FF assets pricing model.[26] The innovations in equations (22)-(23) (or equivalently (23.b)) are decomposed as $\mathbf{v}_{f,t} = \mathcal{S}_{f,t}^{1/2}\boldsymbol{\eta}_{f,t}$ and $\mathbf{v}_{\mathbf{r},t} = \mathcal{S}_{r,t}^{1/2}\boldsymbol{\eta}_{r,t}$, where each of the conditional covariance matrices ($\mathcal{S}_{f,t}$ and $\mathcal{S}_{r,t}$) follow a distinct scalar BEKK process. In particular:

$$\mathcal{S}_{i,t} = \mathcal{S}_i\,(1 - \tau_i - \delta_i) + \tau_i\,(\mathcal{S}_{i,t-1}^{1/2}\boldsymbol{\eta}_{i,t-1}\boldsymbol{\eta}_{i,t-1}'\mathcal{S}_{i,t-1}^{1/2'}) + \delta_i\,\mathcal{S}_{i,t-1} \qquad i = f, r. \tag{24}$$

The unconditional covariance of the factors, $\mathcal{S}_f$ is targeted to the sample covariance, and positive definiteness of $\mathcal{S}_r$ is ensured via triangular decomposition, see Section 4.

To assess the type and extent of beta spillovers we estimate equation (5) under the four parame-

[25] The three factor model with time-varying slope coefficients in equation (23) implies time-varying intercepts $\boldsymbol{\alpha}_t = \boldsymbol{\mu}_r - \mathbf{B}_t\boldsymbol{\mu}_f$.

[26] Recall that $\mathrm{E}(\boldsymbol{\alpha}_t) = \boldsymbol{\mu}_r - \mathrm{E}(\mathbf{B}_t)\boldsymbol{\mu}_f = \boldsymbol{\mu}_r - \mathbf{B}\boldsymbol{\mu}_f = \boldsymbol{\alpha}$.

ter restrictions: *i)* idiosyncratic shocks, which constitutes the benchmark specification, *ii)* factor spillovers, *iii)* asset spillovers and, *iv)* both asset and factor spillovers. To contain the number of parameters and thus reduce the computational burden, we allow spillovers of shocks via the matrix $\mathbf{\Omega}$ (direct effect), while we restrict the contribution of the smoothing term (feedback) to be idiosyncratic. This is achieved by constraining the matrix of parameters $\mathbf{\Gamma}$ to be diagonal. For the sake of feasibility, we also target the unconditional level of $\mathbf{B}_t$ to the OLS estimate. This is done via reparameterization of the intercept in equation (5) as $\mathbf{\Psi} = \mathbf{B}_{\text{OLS}} \odot \left(\text{vec}^{-1}_{(k\times n)}\left(\text{diag}(\text{I}_{(nk)} - \mathbf{\Gamma})\right)\right)'$. Finally, to avoid the aforementioned problem of reduction in the scale of the innovations driving the dynamics of the conditional beta induced by the block orthogonalization, as well as to reduce the impact of abnormally large shocks, we let $\mathbf{B}_t$ depend on products of devolatilized innovations. The model's parameter are estimated by QML using the two-step approach, $\mathcal{M}_2$.

Table 5 reports the second step estimation results for the four conditional beta specifications and the covariance equation. Note that the inference on the elements of the unconditional beta matrix $\mathbf{B}$ are based on OLS. For the other parameters, QML standard errors are computed using the sandwich formula.[27] For the Coal industry portfolio, the unconditional exposure to the market factor amounts to 1.17 denoting a portfolio fairly correlated with the market, $(\text{Corr}(r_{C,t}, f_{mkt,t}) = 0.59)$ but sensibly more volatile $(\text{Var}(r_{C,t}) = 4.64$ against $\text{Var}(f_{mkt,t}) = 1.17)$. This portfolio appears to be also unconditionally exposed to the size factor $(\beta^C_{smb} = 0.45)$, as well as the value factor $(\beta^C_{hml} = 0.51)$. Turning to the Petroleum-Natural Gas industry portfolio, the unconditional beta associated to the market factor is slightly below one. This portfolio, despite being more correlated with the market $(\text{Corr}(r_{P,t}, f_{mkt,t}) = 0.78)$, exhibits a much less erratic behavior $(\text{Var}(r_{P,t}) = 1.76)$. Unconditional exposure to the size factor is negative but close to zero. Finally, with respect to the unconditional exposure to the value factor the Petroleum-Natural Gas industry portfolio shows levels comparable to what observed for the Coal industry portfolio $(\beta^P_{hml} = 0.54)$.

For all six conditional betas we find evidence of time variation characterized by highly persistent dynamics. The estimated smoothing coefficients are very close to one, with the $\gamma_{i,j}$ coefficients ranging from 0.9966 to 0.9996. These results are in line with the findings of Darolles, Francq,

[27] Parameter estimates (standard errors) for the first step are $\tau_f = 0.0641$ (0.0031) and $\delta_f = 0.9327$ (0.0033).

Table 5: *Estimation results: parameters (Par), source of spillover (SO), estimates (Est), standard errors (SE), p-values (pval). The last row reports the second step log-likelihood ($\log \boldsymbol{\mathcal{L}}$).*

		Idiosyncratic			Asset spillovers			Factor spillovers			Factor & Asset spill.		
Par	SO	Est	SE	pval	Est	SE	pval	Est	SE	pval	Est	SE	pval
Conditional mean: Coal													
α^C		-0.0208	0.0068	0.00	-0.0210	0.0070	0.00	-0.0207	0.0068	0.00	-0.0206	0.0068	0.00
β^C_{mkt}		1.1706	0.0102	0.00									
ω_{11}	mkt	0.0064	0.0011	0.00	0.0065	0.0012	0.00	0.0070	0.0011	0.00	0.0071	0.0012	0.00
ω_{12}	smb							0.0027	0.0013	0.04	0.0032	0.0014	0.02
ω_{13}	hml							0.0013	0.0018	0.45	0.0025	0.0022	0.26
ω_{14}	oil				0.0020	0.0018	0.28				0.0012	0.0016	0.47
γ_{11}		0.9996	0.0002	0.00	0.9996	0.0002	0.00	0.9995	0.0003	0.00	0.9995	0.0003	0.00
β^C_{smb}		0.4488	0.0186	0.00									
ω_{21}	mkt							-0.0012	0.0018	0.50	-0.0008	0.0020	0.68
ω_{22}	smb	0.0083	0.0024	0.00	0.0103	0.0028	0.00	0.0071	0.0032	0.03	0.0111	0.0045	0.01
ω_{23}	hml							0.0019	0.0019	0.32	-0.0011	0.0037	0.75
ω_{25}	oil				0.0010	0.0023	0.65				0.0005	0.0032	0.86
γ_{22}		0.9982	0.0007	0.00	0.9981	0.0006	0.00	0.9984	0.0006	0.00	0.9981	0.0007	0.00
β^C_{hml}		0.5089	0.0181	0.00									
ω_{31}	mkt							-0.0012	0.0025	0.62	-0.0017	0.0024	0.48
ω_{32}	smb							0.0024	0.0030	0.41	0.0008	0.0032	0.80
ω_{33}	hml	0.0252	0.0029	0.00	0.0252	0.0032	0.00	0.0248	0.0028	0.00	0.0251	0.0031	0.00
ω_{36}	oil				0.0087	0.0036	0.02				0.0096	0.0052	0.06
γ_{33}		0.9969	0.0006	0.00	0.9967	0.0006	0.00	0.9969	0.0006	0.00	0.9970	0.0006	0.00
Conditional mean: Petroleum-Natural Gas													
α^P		0.0032	0.0034	0.35	0.0027	0.0034	0.43	0.0035	0.0034	0.31	0.0030	0.0033	0.36
β^P_{mkt}		0.9296	0.0049	0.00									
ω_{41}	coal				-0.0001	0.0008	0.94				0.0001	0.0009	0.97
ω_{44}	mkt	0.0096	0.0015	0.00	0.0090	0.0012	0.00	0.0084	0.0014	0.00	0.0081	0.0010	0.00
ω_{45}	smb							-0.0027	0.0012	0.02	-0.0026	0.0011	0.02
ω_{46}	hml							0.0015	0.0013	0.24	0.0015	0.0011	0.19
γ_{44}		0.9975	0.0009	0.00	0.9980	0.0009	0.00	0.9972	0.0009	0.00	0.9973	0.0010	0.00
β^P_{smb}		-0.0515	0.0090	0.00									
ω_{52}	coal				0.0034	0.0009	0.00				0.0041	0.0012	0.00
ω_{54}	mkt							-0.0036	0.0022	0.11	-0.0052	0.0017	0.00
ω_{55}	smb	0.0166	0.0038	0.00	0.0121	0.0027	0.00	0.0134	0.0030	0.00	0.0112	0.0024	0.00
ω_{56}	hml							0.0031	0.0020	0.14	0.0015	0.0022	0.48
γ_{55}		0.9966	0.0013	0.00	0.9978	0.0009	0.00	0.9973	0.0009	0.00	0.9974	0.0009	0.00
β^P_{hml}		0.2354	0.0088	0.00									
ω_{63}	coal				0.0065	0.0021	0.00				0.0063	0.0023	0.01
ω_{64}	mkt							-0.0060	0.0025	0.02	-0.0057	0.0018	0.00
ω_{65}	smb							0.0085	0.0018	0.00	0.0029	0.0021	0.18
ω_{66}	hml	0.0209	0.0028	0.00	0.0196	0.0020	0.00	0.0223	0.0029	0.00	0.0197	0.0025	0.00
γ_{66}		0.9976	0.0005	0.00	0.9977	0.0005	0.00	0.9968	0.0006	0.00	0.9976	0.0006	0.00
Conditional covariance: $\mathcal{S}_{r,t}$													
$\mathcal{S}_{r,[11]}$		4.2683	0.9536	0.00	4.1794	0.8806	0.00	4.3576	0.2709	0.00	4.4304	0.2648	0.00
$\mathcal{S}_{r,[12]}$		-0.0101	0.0078	0.19	-0.0102	0.0083	0.22	-0.0150	0.0112	0.18	-0.0172	0.0118	0.14
$\mathcal{S}_{r,[22]}$		0.8455	0.0614	0.00	0.8208	0.0385	0.00	0.8609	0.0106	0.00	0.8614	0.0238	0.00
τ_r		0.0347	0.0041	0.00	0.0336	0.0040	0.00	0.0346	0.0041	0.00	0.0341	0.0039	0.00
δ_r		0.9640	0.0044	0.00	0.9651	0.0041	0.00	0.9641	0.0043	0.00	0.9648	0.0041	0.00
$\log \boldsymbol{\mathcal{L}}$			-62892.00			-62849.44			-62845.57			-62808.60	

and Laurent (2018) for a comparable modeling approach and the same sampling frequency and assets class.

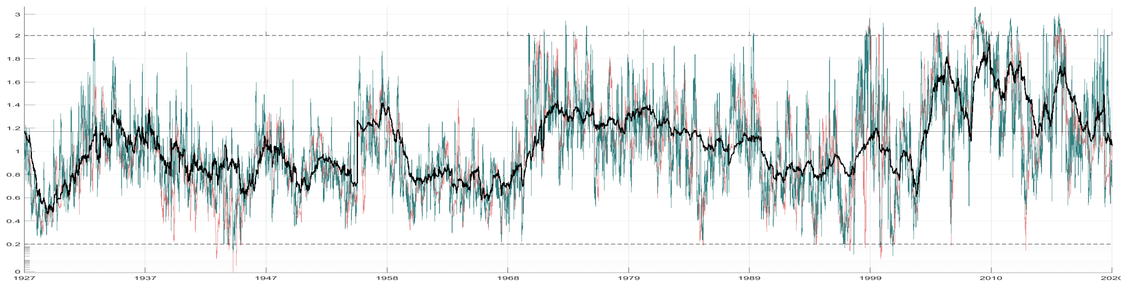
We also find evidence of both factor and asset spillovers. More precisely, with respect to factor spillovers, we find that shocks in the size factor ($f_{smb,t}$), impact the exposure of both portfolios to the market factor, albeit in opposite direction for $\beta^C_{mkt,t}$ and $\beta^P_{mkt,t}$. Evidence specific to each portfolio is found when looking at the exposure to the size and the value factors. For the Coal industry portfolio, we do not find any evidence of spillovers on $\beta^C_{smb,t}$, while we find statistically significant asset spillovers on β^C_{hml}. For the Petroleum-Natural Gas industry portfolio, the spillovers in the exposures to $f_{smb,t}$ and $f_{hml,t}$ involve to both asset and factor spillovers. The source of the spillover is, in both cases, limited to the market factor.

It is worth recalling that equation (5) under the restrictions in *iv)* nests the parameterizations in *i)*, *ii)* and *iii)*. Estimates of the parameters associated to the idiosyncratic shock, common to all specifications, are found to be close across specifications. This suggests that the inclusion of spillovers, although not independent nor orthogonal to the idiosyncratic shock, provide novel and relevant information. This is confirmed by the likelihood ratio test which finds the specification allowing for both assets and factors spillovers statistically superior to the others at standard confidence levels. We rely on this specification for all results in the reminder of this section.

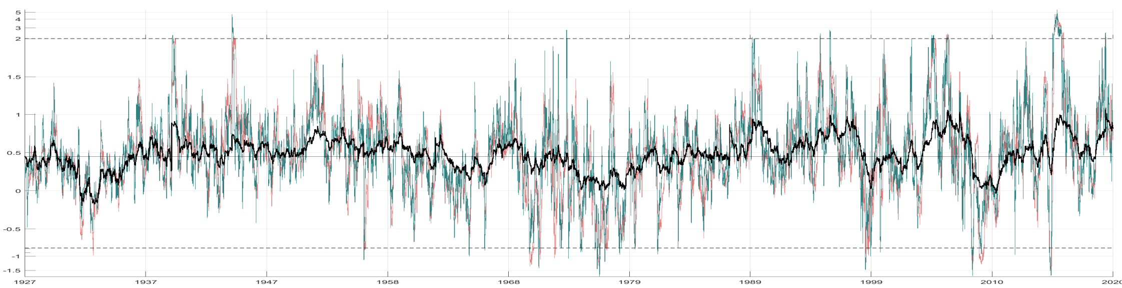
Finally, the 3FF model predicts that the risk factors are sufficient to price assets. Formally, the intercept in equation (23.b) ($\boldsymbol{\alpha} = (\alpha^C, \alpha^P)'$), is expected to vanish. When testing the null hypothesis $\alpha^n = 0$, $n = C, P$ individually, the intercept, α^P, is not statistically different from zero. The zero intercept hypothesis, is instead rejected for the Coal industry portfolio, suggesting possible misspecification of the pricing model.

Figures 3 and 4 report the six filtered conditional betas, $\beta^n_{j,t}$, $n = C, P$ and $j = mkt, smb, hml$ for the specification in equation (5) in the parameterisation iv). Each plot contains the conditional beta estimated by the BC-GARCH (black), as well as those obtained using two competing approaches: the indirect estimation of the DCB-DCC model (green) and the rolling OLS estimator with a window of 100 observations (red). The horizontal line represents the unconditional beta estimated by full-sample OLS.

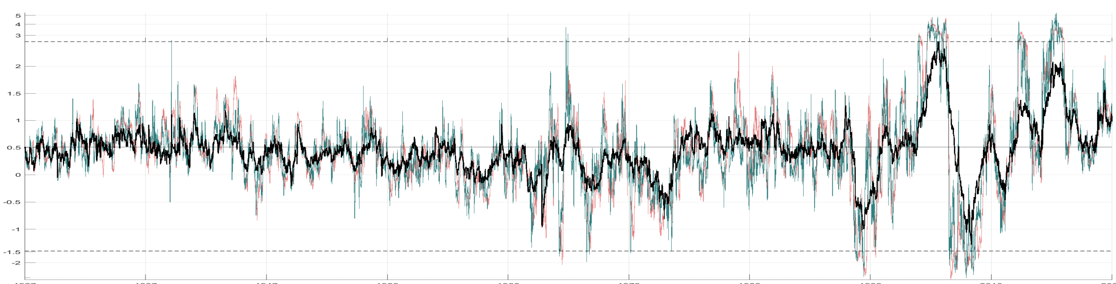
Note that for the rolling OLS estimator, smoother sample paths are to be expected for larger



(a) $\beta^{C}_{mkt,t}$



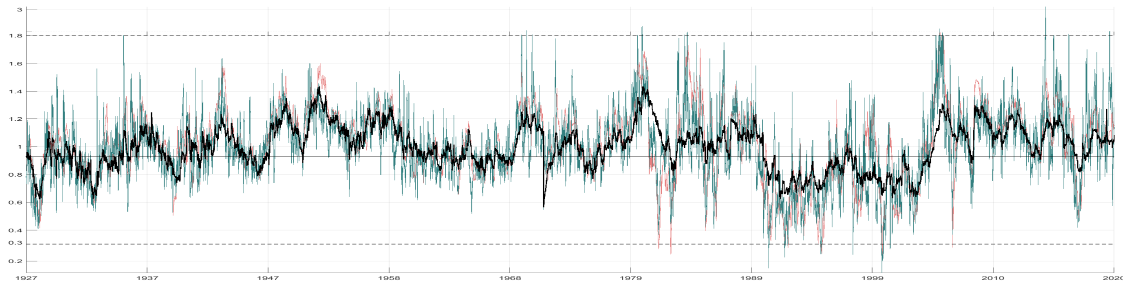
(b) $\beta^{C}_{smb,t}$



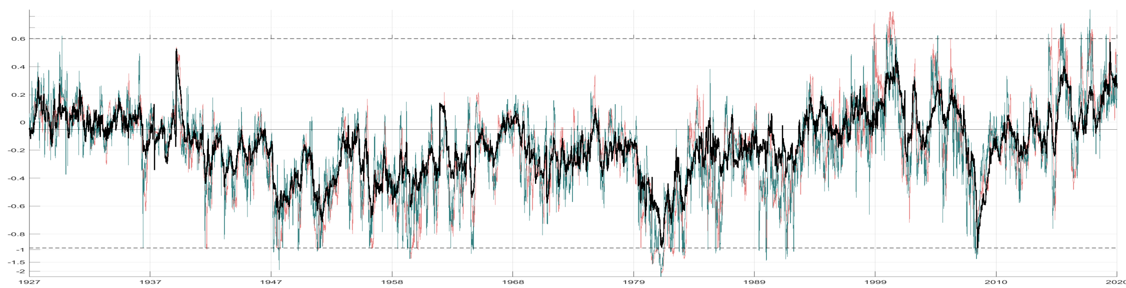
(c) $\beta^{C}_{hml,t}$

Figure 3: *Dynamic conditional betas for the Coal industry portfolio obtained using equation* (5) *in the parameterisation iv) (black) against the DCB-DCC (green) and the 100-obs. rolling OLS (red). Solid horizontal lines represent the unconditional betas. Values outside the dashed horizontal lines are reported in logarithmic scale.*

window sizes, up to the limit, in a stationary setting, of constant paths as the estimation windows grows to infinity. We include the rolling OLS estimator to the comparison because, since the seminal work of Fama and MacBeth (1973) it still represents a standard in the asset pricing literature to assess the time variation of the exposure to risk factors. However, if on the one hand, the rolling OLS has the advantage of an intuitive interpretation and easy implementation, on the other hand, it makes difficult to distinguish between actual time variation, i.e. the signal,



(a) *$\beta^{P}_{mkt,t}$ of the market factor.*



(b) *$\beta^{P}_{smb,t}$ of the small minus big factor.*

(c) *$\beta^{P}_{hml,t}$ of the hight-minus-low factor.*

Figure 4: *Dynamic conditional betas for the Petroleum and natural gas industry portfolio obtained using equation* (5) *in the parameterisation iv) (black) against the DCB-DCC (green) and the 100-obs. rolling OLS (red). Solid horizontal lines represent the unconditional betas. Values outside the dashed horizontal lines are reported in logarithmic scale.*

and estimator's sampling variability, i.e. the noise. In general, we observe a substantial time variation for all conditional betas, with the three methods tracking well each other.

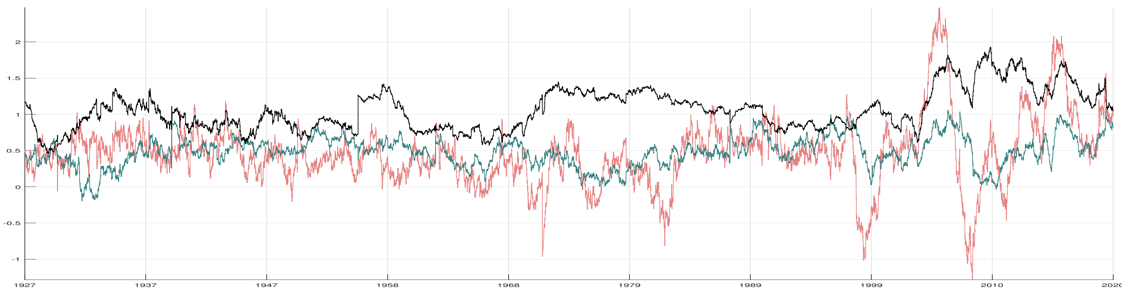
The conditional betas filtered with our model exhibit smoother sample paths than those obtained with the competing alternatives. Setting aside the rolling OLS approach, where the degree of smoothness directly depends on the window size, the DCB-DCC model yields comparatively very erratic conditional betas with short lived bursts of unrealistic size. Some examples are, $\beta^{C}_{mkt,t}$

Table 6: *Summary statistics of the conditional betas filtered using the BC-GARCH, the DCB-DCC and the rolling OLS.*

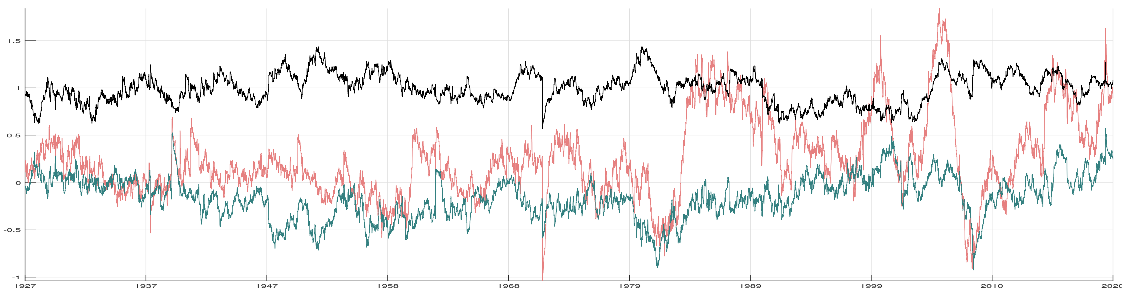
Beta	Model	Mean	Variance	Skew.	Kurt.	Min.	Max.
$\beta^C_{mkt,t}$	BC-GARCH	1.0538	0.0775	0.5679	2.8776	0.4577	1.9310
	DCB-DCC	1.0302	0.1716	0.9217	4.6702	0.0203	3.4301
	roll-OLS	1.0092	0.1767	0.6111	4.1824	-0.6869	2.8903
$\beta^C_{smb,t}$	BC-GARCH	0.4797	0.0441	-0.2128	3.1733	-0.1993	1.0904
	DCB-DCC	0.4871	0.3015	1.1229	9.5182	-1.8050	5.4688
	roll-OLS	0.4963	0.3528	1.0083	8.9596	-1.6652	4.7662
$\beta^C_{hml,t}$	BC-GARCH	0.4663	0.2078	0.6435	6.2719	-1.2849	2.4753
	DCB-DCC	0.4508	0.6118	1.0722	8.4339	-3.0003	5.3630
	roll-OLS	0.4664	0.7819	0.7057	6.8196	-3.1547	4.6908
$\beta^P_{mkt,t}$	BC-GARCH	0.9925	0.0240	0.0171	2.6388	0.5641	1.4350
	DCB-DCC	0.9914	0.0618	0.1446	4.0429	0.1547	3.1553
	roll-OLS	1.0166	0.0651	-0.1319	3.3598	0.2231	2.0801
$\beta^P_{smb,t}$	BC-GARCH	-0.1542	0.0497	-0.1647	3.0977	-0.9256	0.5753
	DCB-DCC	-0.1964	0.0946	-0.6967	5.1753	-2.4062	0.8690
	roll-OLS	-0.2072	0.1189	-0.5510	5.0423	-2.1478	0.8529
$\beta^P_{hml,t}$	BC-GARCH	0.2471	0.1747	0.7534	3.8998	-1.0375	1.8384
	DCB-DCC	0.2666	0.2924	1.5619	8.6156	-1.5077	4.4209
	roll-OLS	0.2728	0.4102	1.2401	6.5054	-1.4355	4.0686

peaking above 3.4 in January 2009, i.e. almost double than our model's estimate, or $\beta^C_{smb,t}$ which climbs rapidly from -1.8 to 5.4 between July and December 2015, while our estimate, over the same period, moves from 0.22 to 1.00. A similar behavior is easily noticeable for all the other betas and all along the sample.

For ease of comparison, Table 6 reports sample moments of the filtered betas plotted in Figure 3 and 4. Compared to our beta estimates, the two competing approaches show a variability that is twice to almost 10-fold. Our model seems to struggle only in presence of abrupt shifts caused by episodes of sharp reversion. When clusters of large unidirectional shocks occur the persistent dynamics of the BC-GARCH generates a rather slow transition. In our sample, we observe one of such instances in the period spanning between July and November 2006, though limited to $\beta^i_{hml,t}$, $i = C, P$. In such instance, the DCB-DCC and the rolling OLS exhibit a faster adjustment. Figure 5 plots the filtered paths of the three risk exposures for the two portfolios studied. With respect to the size factor, β^C_{smb} is only sporadically negative indicating that the Coal portfolio consistently trades like a small stock. For the Petroleum-Natural Gas portfolio, β^P_{smb} is systematically negative from the mid-40s to the mid-90s indicating that the sector in this



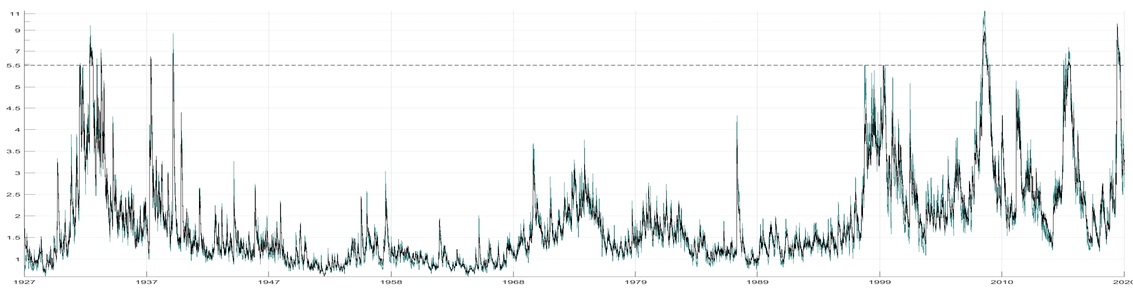
(a) Coal industry portfolio (C).



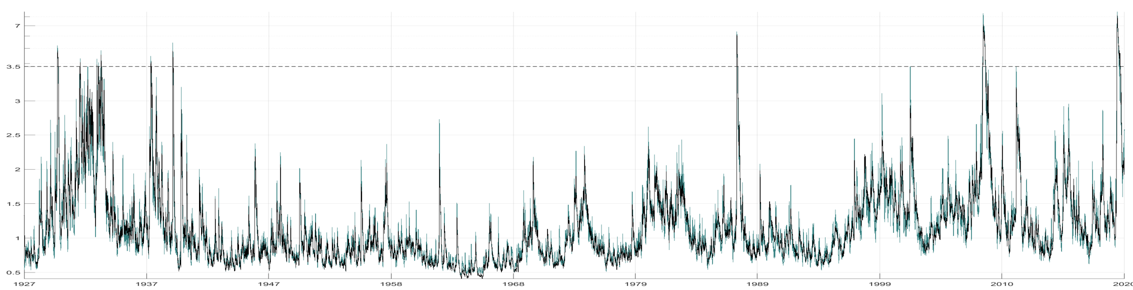
(b) Petroleum-Natural Gas industry portfolio (P).

Figure 5: *Time varying exposures $\beta^i_{mkt,t}$ (black), $\beta^i_{smb,t}$ (green), $\beta^i_{hml,t}$ (red), $i = C, P$.*

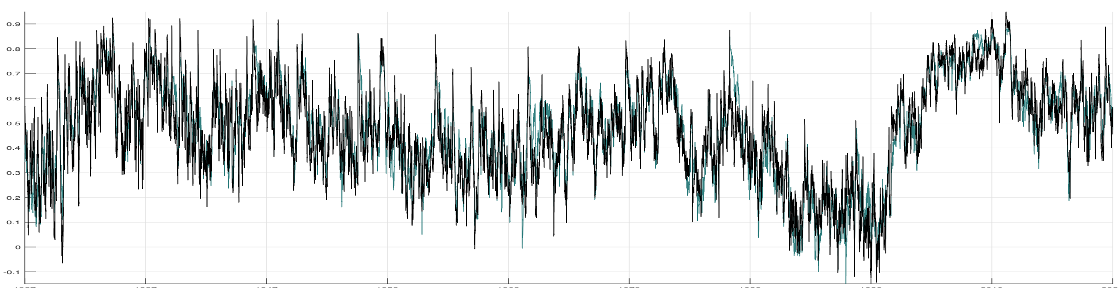
period moves like a large-cap stock. Outside this period, it hovers around zero except towards the inception and during the 2007 financial crisis. In line with the findings of Engle (2016), the exposure to the value factor is the most volatile. The exposure to this factor shows similar peaks and troughs for both portfolios, but with β^C_{hml} exhibiting a much larger amplitude. In both cases, the highest peak is observed in the period 2004-2006, when both portfolios trade as a value stock, before the deep dive over the following two years when both portfolios behave as a growth stock. One substantial difference is observed in the period from 1997 to 2000. While the coal sector shows a trough, trading as a growth stock, the oil sector shows a specular peak, trading in the same period as a value stock. The last question that we want to address is whether and to what extent the differences observed in the filtered betas obtained using different competing models translate into differences in the estimated $\mathcal{I}_{t-1}$-conditional covariance matrix of the cross-section of assets $\Sigma_{\mathbf{r},t}$. To this end, we limit the comparison to the DCB-DCC model. Note that, methodologically, the estimation of the conditional covariance $\Sigma_{\mathbf{r},t}$ is specular to that of the conditional betas in the two models. That is, $\Sigma_{\mathbf{r},t}$ it is estimated directly by the DCB-DCC and indirectly using (2) by the BC-GARCH model. Figures 6 plots the Coal and Petroleum-Natural Gas con-



(a) *Conditional standard deviation of the Coal industry portfolio.*



(b) *Conditional standard deviation of the Petroleum-Natural Gas industry portfolio.*



(c) *Conditional correlation between Coal and Petroleum-Natural Gas industry portfolios.*

Figure 6: *Volatility and correlation of Coal and Petroleum-Natural Gas portfolio returns from BC-GARCH (black) and DCB-DCC (green). Values above the dashed line are reported on a logarithmic scale.*

ditional standard deviations as well as the conditional correlation fileterd using the BC-GARCH and the DCB-DCC. We do not observe any substantial difference between the two models which track each other extremely closely. We remark, though, that sample paths estimated by the BCG model are slightly less noisy and characterized by a less extreme response of the conditional variance to isolated large shocks, which on the contrary is the typical reaction of GARCH-type dynamics. Table 7, which reports sample statistics for the individual elements of $\Sigma_{\mathbf{r},t}$ and the conditional correlation, confirms this conclusions. Such evidence suggests that the BC-GARCH represents a robust alternative for multivariate volatility modeling in presence of sporadic large

Table 7: *Summary statistics of the conditional variances, covariance and correlation filtered using the BC-GARCH and the DCB-DCC*

Variable	Model	Mean	Variance	Skew.	Kurt.	Min.	Max.
$\Sigma_{r,[11],t}$ (C)	BC-GARCH	4.5214	48.0623	5.0479	39.3491	0.3735	95.0693
	DCB-DCC	4.7733	71.0434	6.1755	59.1489	0.3469	129.8896
$\Sigma_{r,[12],t}$	BC-GARCH	1.5411	10.5895	8.3309	103.3036	-1.0674	64.1005
	DCB-DCC	1.6014	13.1995	9.6543	137.3450	-0.9023	74.3991
$\Sigma_{r,[22],t}$ (P)	BC-GARCH	1.7788	10.0684	9.2574	123.1644	0.1656	66.6386
	DCB-DCC	1.8082	11.5021	10.5299	157.4155	0.2110	74.1664
$\rho_{12,t}$	BC-GARCH	0.4734	0.0380	-0.1857	2.6211	-0.1426	0.9486
	DCB-DCC	0.4814	0.0378	-0.2984	2.5565	-0.1480	0.8947

Table 8: *In sample performance of the beta hedged portfolios*

	Mean		Variance		Abs. Sharpe ratio		Turnover	
Model	C	P	C	P	C	P	C	P
BC-GARCH	-0.0127	-0.0026	2.6451	0.5484	0.0078	0.0035	0.0281	0.0274
DCB-DCC	-0.0127	-0.0024	2.6796	0.5648	0.0077	0.0031	0.1666	0.0927
roll-OLS	-0.0099	-0.0037	2.6458	0.5496	0.0061	0.0050	0.1056	0.0502

shocks in risk management or portfolio allocation applications, as well as in any other situation where the dynamics of the conditional covariance of a cross section of assets, $\Sigma_{\mathbf{r},t}$, are of interest. In a beta hedging strategy, a more volatile beta implies more re-balancing and thus much larger transaction costs, as well as potentially a more volatile beta hedged portfolio. The performance of the beta hedged portfolios, though limited to an in-sample evaluation, is reported in Table 8. The most striking figure is observed for the turnover,[28] which for our model is up to six times smaller than the alternatives.

5.2 Inference on factors risk premia

In this empirical application we use the BC-GARCH to infer the risk premia associated to the risk factors exposures in a liner beta pricing specification. In the interest of brevity, we report results for two beta pricing models: the static CAPM and the Fama and French (1993) three-factor model (3FF).

The $k = 3$ factors' returns are described by equations (22)-(24), while the $n = 40$ assets returns,

[28]The turnover is calculated as $T^{-1} \sum_{k=1}^{3} \left|\beta^i_{k,t} - \beta^i_{k,t-1}\right|$, $i = C, P$ and $k = mkt, smb, hml$ indexing the risk factors.

conditional to the risk factors, are described by:

$$\mathbf{r}_t = \boldsymbol{\mu}_{r,t} + \mathbf{B}_t(\mathbf{f}_t - \boldsymbol{\mu}_f) + \mathbf{v}_{r,t}, \tag{25}$$

with $\boldsymbol{\mu}_{r,t} = \mathrm{E}(\mathbf{r}_t|\mathcal{I}_{t-1})$ and $\mathbf{v}_{r,t} = \mathcal{S}_{r,t}^{1/2}\boldsymbol{\eta}_{r,t}$. Assuming $\boldsymbol{\mu}_{r,t} = \boldsymbol{\mu}_r$, equation (25) boils down to equation (23). In this case the null hypothesis $\mathrm{H}_0 : \boldsymbol{\mu}_{r,t} = \mathbf{B}\boldsymbol{\mu}_f$ or equivalently $\mathrm{H}_0 : \alpha_i = 0$, $i = 1, \ldots, n$, in equation (23.b), represents a simple misspecification test for the pricing model. The test can be carried out individually or jointly, by exploiting the multivariate representation in equation (23.b).[29] Despite its usefulness, equation (23.b) cannot identify the source and the form of the risk-return trade-off.

Following Connor (1984), a more appealing testable hypothesis expresses the conditional expectation of the assets return as:

$$\mathrm{E}(\mathbf{r}_t|\mathcal{I}_{t-1}) = \delta + \mathbf{B}_t\boldsymbol{\lambda}, \tag{26}$$

that is a linear conditional beta pricing specification, where $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_k\}$, is the vector of risk premia attached to the factor exposure. Note that δ and $\boldsymbol{\lambda}$ are invariant in the cross-section. Our joint specification in equation (25) can easily integrate this restriction.[30]

The linear conditional pricing specification in equation (26) measures explicitly the size of the risk-return trade-off and represents a misspecification test of the pricing model, see Sharpe (1964) and Lintner (1965). For example, in the CAPM framework, assuming market and assets returns are both expressed in excess of the risk free rate, misspecification of the pricing model takes the form of the Sharpe-Lintner hypothesis $\mathrm{H}_0 : \delta = 0 \cap \lambda = \mu_{mkt}$.[31]

To explicitly account for misspecification of the asset return in the form of omitted priced factors, Kleibergen (2009) relaxes equation (26) by introducing a fixed idiosyncratic effect of the form:

$$\mathrm{E}(\mathbf{r}_t|\mathcal{I}_{t-1}) = \delta_0 + \boldsymbol{\delta} + \mathbf{B}_t\boldsymbol{\lambda}. \tag{27}$$

Under the suitable identification restriction $\delta_i = 0$, for some $i \in [1, n]$. The idiosyncratic intercepts make it possible to test the restriction $\boldsymbol{\delta} = 0$, rather than imposing it a priori. It also preserves the validity of the QML inference when the restriction is violated, by ensuring that the innovations

[29] Note that this is the standing hypothesis in a constant beta framework.

[30] Other specifications such as quadratic beta pricing, beta interactions or systematic effects of non-β risks could also be evaluated, see Fama and MacBeth (1973) for a variety of specifications.

[31] Such hypothesis is, in fact, equivalent to $\mathrm{H}_0 : \boldsymbol{\alpha} = 0$ in equation (23.b).

are centered over the time dimension. Thus, in a multi-factor asset pricing specification, the linear conditional pricing model can be expressed as:

$$\mathbf{f}_t = \boldsymbol{\mu}_f + \mathbf{v}_{f,t}, \tag{28}$$

$$\mathbf{r}_t = \delta_0 + \boldsymbol{\delta} + \mathbf{B}_t \left(\mathbf{f}_t - \boldsymbol{\mu}_{\mathbf{f}} + \boldsymbol{\lambda}\right) + \mathbf{v}_{r,t}. \tag{29}$$

A large cross-sectional dimension, as the one considered in our application, inevitably translates in a substantial number of parameters which in turn may hinder the computational feasibility. To attenuate the curse of dimensionality we impose the following restrictions. First, since our interest is on the risk premia $\boldsymbol{\lambda}$, we remove the individual effect by centering the returns around their unconditional expectation. Equation (29) is reparameterized as:

$$\mathbf{r}_t - \boldsymbol{\mu}_r = \left(\mathbf{B}_t - \mathbf{B}\right)\boldsymbol{\lambda} + \mathbf{B}_t \left(\mathbf{f}_t - \boldsymbol{\mu}_{\mathbf{f}}\right) + \mathbf{v}_{r,t}. \tag{30}$$

Second, we adopt a simple and parsimonious dynamics for the matrix of conditional betas $\mathbf{B}_t$ by imposing cross-sectional restrictions. To this end, we assume $\mathbf{B}_t$ follows the *semi-scalar* dynamics described in Section 3.4. Together with beta tracking, i.e. $\mathbf{B}$ is targeted to the OLS estimator, the conditional beta specification takes the form:

$$\mathbf{B}_t = \mathbf{B} \odot \left[\mathbf{e}_{(n)} \times \left(\mathbf{e}'_{(k)} - \boldsymbol{\gamma}'_{(k)}\right)\right] + \mathbf{e}_{(n)}\boldsymbol{\omega}'_{(k)} \odot \left(\boldsymbol{\xi}_{r,t}\boldsymbol{\xi}'_{f,t}\right) + \mathbf{e}_{(n)}\boldsymbol{\gamma}'_{(k)} \odot \mathbf{B}_{t-1}. \tag{31}$$

Third, similar considerations hold for the dynamics of the conditional covariance matrix $\mathcal{S}_{r,t}$. Because covariance targeting is unfeasible for $\mathcal{S}_{r,t}$ and since only its diagonal elements are required to standardize the innovations in equation (31), we rely on the marginalization of the scalar BEKK model. This allows us to reduce the number of parameters by $n(n-1)/2$, which in this case amounts to 780 nuisance unconditional covariances.

Fourth, compatibly with the choices described above, i.e. cross-sectional constraints together with marginalization, we estimate the model's parameters using $\mathcal{M}_4$ detailed in Section 3.6. The total number of parameters estimated by Gaussian QML, via log-transformation of equation (17) and (18), is 53.[32]

We benchmark our model against the 2-Pass Cross-Sectional Regression (2PR) framework of

[32] The total number of parameters is partitioned as follows: 2 for the first step (risk factors block) and 51 for the second step (assets block) of which 40 are the unconditional variances. of $\mathbf{v}_{r,t}$.

Fama and MacBeth (1973). The 2PR works as follows. The risk premia for the different factors exposures are estimated as time-averages from a sequence of cross-sectional regressions of the asset returns on the factor betas, where the betas are obtained from a time-series regression of the asset on the factors.

Following Fama and MacBeth (1973) we use a rolling window of 5 years of data (60 observations) to estimate the n first-pass time series regressions of each asset on the risk factors. Then in the second-pass, for each time period (month), $t = 61, \ldots, T$, the cross-section of asset returns is regressed on the cross-section of beta. The risk premia are then estimated by time averages of the sequences of slopes resulting from the second-pass regressions. To account for serial dependence in the sequence of slopes, standard errors of the last step are estimated using the Newey and West (1987) with lag-lenght $4(T/100)^{2/9}$.

Despite constituting a standard in the literature, the 2PR is subject to the well-known error-in-variables problem and the consequent compounding of estimation errors. Also, as pointed out by Kan and Zhang (1999), linear regression estimates of the risk premia are sensitive to collinearity of the betas which occurs when factors exposures are close to zero, or when the beta-matrix is almost of reduced rank, and the expected asset returns are non-zero. Furthermore, Kleibergen (2009) show that 2PR risk premia estimates are spurious whenever the betas are relatively small and their spuriousness is aggravated when the number of assets is large. The risk premia estimates typically turn too small and are characterized by an erroneous level of precision.

Figure 7 reports sample statistics (median, interquartile range and extremes) of the cross-sectional distribution of the unconditional betas (OLS) for the three risk factors. The market betas range between 1.32 and 0.66 and is distributed symmetrically about 1.01. Both the β^i_{smb} and β^i_{hml}, $i = 1, ..., 40$, exhibit a more dispersed cross-sectional distribution, centered about 0.18 and 0.078 and, right and left-skewed respectively.

Table 9 reports parameter estimates, standard errors and corresponding p-values for the model parameters of the BC-GARCH and the risk premia estimated using the 2PR. To save space, unconditional betas $\mathbf{B}$, unconditional variances diag$(\mathcal{S}_r)$ and, the parameters of equation (28) are not reported, but are available upon request.

The most striking result in Table 9 are the significantly positive risk premia estimated by the

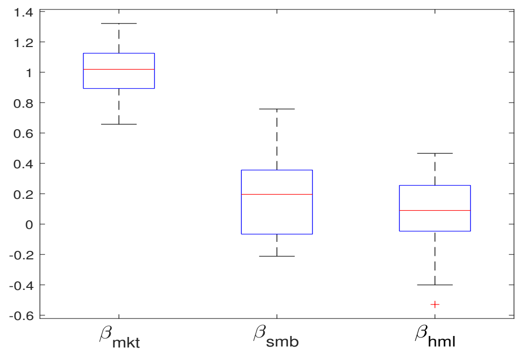


Figure 7: *Boxplot of the unconditional exposures to the 3FF factors for the 40 industry portfolios.*

Table 9: *Estimation results: parameter estimates (Est), standard errors (SE) and p-values (pval). The last row of each Panel reports the second step log likelihood ($\log \boldsymbol{\mathcal{L}}$) for the BC-GARCH.*

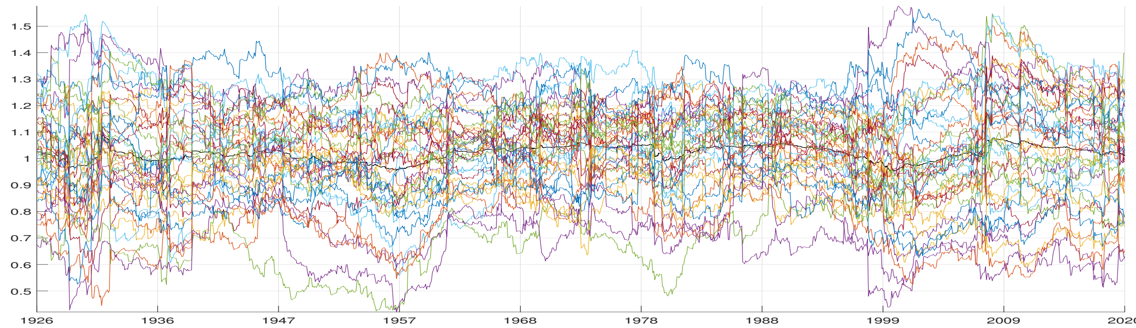
Panel A: CAPM							
			BC-GARCH			2PR	
		Est	SE	p-val	Est	SE	pval
$\mu_{r,t}$	λ_1 (*mkt*)	0.1458	0.0511	0.00	0.2489	0.2153	0.25
$\beta^i_{mkt,t}$	ω_1	0.0219	0.0026	0.00			
	γ_1	0.9848	0.0029	0.00			
$\boldsymbol{\Sigma}_{r,t}$	τ_r	0.0994	0.0090	0.00			
	δ_r	0.8831	0.0108	0.00			
$\log \boldsymbol{\mathcal{L}}$			-123768.25				
Panel B: 3FF							
$\mu_{r,t}$	λ_1 (*mkt*)	0.2856	0.1101	0.01	0.1824	0.1946	0.34
	λ_2 (*smb*)	0.1926	0.0236	0.00	0.1801	0.1215	0.13
	λ_3 (*hml*)	0.2651	0.0373	0.00	0.1475	0.1375	0.28
$\beta^i_{mkt,t}$	ω_1	0.0154	0.0036	0.00			
	γ_1	0.9854	0.0033	0.00			
$\beta^i_{smb,t}$	ω_2	0.0146	0.0047	0.00			
	γ_2	0.9909	0.0047	0.00			
$\beta^i_{hml,t}$	ω_3	0.0188	0.0053	0.00			
	γ_3	0.9890	0.0038	0.00			
$\boldsymbol{\Sigma}_{r,t}$	τ_r	0.0965	0.0207	0.00			
	δ_r	0.8910	0.0251	0.00			
$\log \boldsymbol{\mathcal{L}}$			-121144.33				

BC-GARCH, compared to the insignificant 2PR estimates. The market risk premium is 0.15% and 0.29% on a monthly basis, depending on the specification of the pricing model. This result validates the existence of a positive expected risk-return trade-off. However, our results show violation of the Sharpe-Lintner hypothesis, i.e. $\lambda_1 = \mathrm{E}(f_{mkt,t})$. Estimates of λ_1 are, under the CAPM and 3FF respectively, 1/4 and 1/2 the sample average of the market excess return, which equals 0.60%. This result conforms with the findings of Fama and MacBeth (1973), which for a shorter though overlapping sample, show evidence of a market risk premium substantially less than the expected market excess return, warning against misspecification of the pricing model.

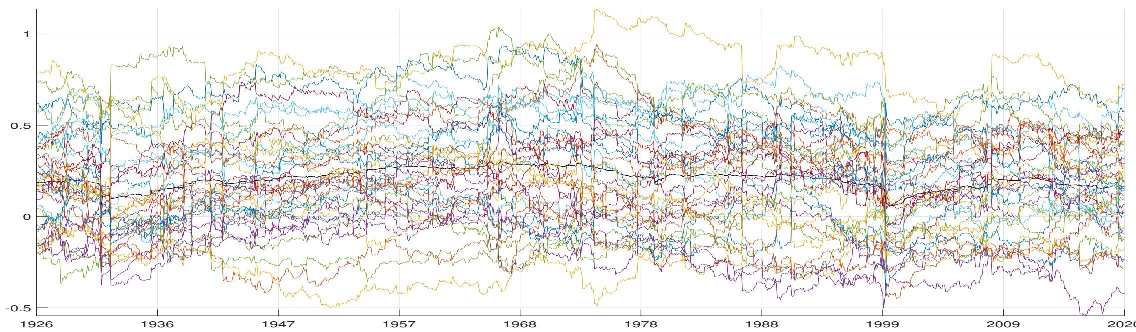
In the 3FF specification, the rejection of $\mathrm{H}_0 : \lambda_j = 0,\ j = 2, 3$, provides strong evidence of systematic effects of non-β_{mkt} risk. The parameter estimates of the *smb* and *hml* risk premia, 0.19% and 0.27%, are very close in size to the corresponding factors sample averages, 0.20% and 0.25%.

In general, there is a compelling difference in the results obtained using the 2PR and the BC-GARCH, namely the level of parameters accuracy. Inference based on the 2PR regression always rejects the existence of positive risk premia. This result appears to be a direct consequence of the unfavorable signal-to-noise ratio struck by the rolling OLS risk exposure estimator, as discussed in Section 5.1. With respect to the conditional betas, Table 9 shows significant time variation of all factors exposures, characterized by persistent dynamics. Estimates of the smoothing parameter γ_k, $k = 1, 2, 3$ range between 0.985 and 0.991. The conditional variances also show rather persistent dynamics. Given the type of data, the sampling frequency and the cross-sectional dimension, this evidence aligns with the vast literature on multivariate GARCH models.

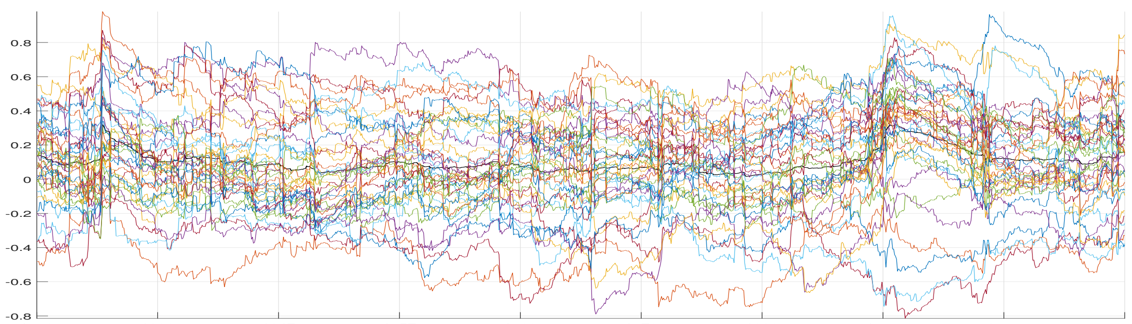
Figure 8 plots the filtered conditional exposures to the 3FF factors. The black line indicates the cross-sectional average of each time period. As expected, the exposure to the market factor ($\beta^i_{mkt,t}$, $i = 1, ..., 40$) is fairly symmetrically distributed around one homogeneously in time, though more dispersed in the first third of the sample and in the period 1998-2010. The exposures to the size ($\beta^i_{smb,t}$) and value ($\beta^i_{hml,t}$) factors are more dispersed and their cross-sectional averages hover about 0.2 and 0.08 respectively, with some pronounced deviations coinciding with the Great Depression and the first decade of the 2000s.



(a) $\beta_{mkt,t}$



(b) $\beta_{smb,t}$



(c) $\beta_{hml,t}$

Figure 8: *Cross-sections of conditional exposures to the 3FF factors for the 40 industry portfolios. The black line indicates the cross-sectional average.*

6 Conclusions

This paper introduces a new model, the BC-GARCH, to estimate time-varying coefficients regression. The BC-GARCH mixes the block orthogonalization of DCB-DCC model with the direct modeling of the conditional betas of the CHAR model.

This approach provides several direct advantages over the existing models. We propose several conditional beta specifications and derive their theoretical properties (stationarity, invertibility and, beta tracking). We also propose computationally efficient altrernative QML estimators. The finite sample properties of the BC-GARCH are studies by means of extensive Monte Carlo simulations.

We illustrate the usefulness of the BC-GARCH in two empirical applications. The first tests our model in the context of the linear three factors (market, size and, value) asset pricing framework. The aim of this application is to benchmark a conditional beta specification driven only by idiosyncratic shocks against three alternatives accounting for different types of beta spillovers. We consider a bivariate asset system composed by the Coal and Petroleum-Natural Gas value-weighted industry portfolios tracked over a period spanning from January 1, 1927 to November 30, 2020. Beside time variation in all the conditional betas, we find compelling evidence of both factor and asset spillovers.

The second empirical application consists of a large scale exercise exploring the cross-sectional variation of expected returns for 40 industry portfolios. We test a linear conditional beta pricing model and estimate the risk premia attached the market, size and, value exposures. We benchmark our model against the 2-pass cross-sectional regression approach. We show that our approach provides a more accurate inference on the risk premia. This is because our model benefits from direct modeling of the time variation in the risk exposures and from the joint estimation of risk factors exposures and risk premia, avoiding compounding of estimation errors.

References

Aielli, G. P. (2013): “Dynamic Conditional Correlation: On Properties and Estimation,” *Journal of Business and Economic Statistics*, 31, 282–299.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): “Multivariate GARCH Models: A Survey,” *Journal of Applied Econometrics*, 21, 79–109.

Bollerslev, T. (1990): “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model,” *Review of Economics and Statistics*, 72, 498–505.

Boussama, F., F. Fuchs, and R. Stelzer (2011): “Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models," *Stochastic Processes and their Applications*, 121, 2331–2360.

Connor, G. (1984): "A Unified Beta Pricing Theory," *Journal of Economic Theory*, 34, 13–31.

Darolles, S., C. Francq, and S. Laurent (2018): "Asymptotics of Cholesky GARCH Models and Time-Varying Conditional Betas," *Journal of Econometrics*, 204, 223–247.

Engle, R. F. (2002): "Dynamic Conditional Correlation: A Simple Class of Multivariate Feneralized Autoregressive Conditional Heteroscedasticity Models," *Journal of Business and Economic Statistics*, 20, 339–350.

——— (2016): "Dynamic Conditional Beta," *Journal of Financial Econometrics*, 14, 643–667.

Engle, R. F., and K. F. Kroner (1995): "Multivariate Simultaneous Generalized Arch," *Econometric Theory*, 11, 122–150.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Fama, F. F., and R. K. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fermanian, J.-D., and H. Malongo (2017): "On the Stationarity of Dynamic Conditional Correlation Models," *Econometric Theory*, 33, 636–663.

Francq, C., and J.-M. Zakoian (2012): "QML Estimation of a Class of Multivariate Asymmetric GARCH Models," *Econometric Theory*, 28, 179–206.

——— (2016): "Estimating Multivariate Volatility Models Equation by Equation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 613–635.

González, M., J. Nave, and G. Rubio (2012): "The Cross Section of Expected Returns with MIDAS Betas," *Journal of Financial and Quantitative Analysis*, 47, 115–135.

Hafner, C. M., and A. Preminger (2009): "On Asymptotic Theory for Multivariate GARCH Models," *Journal of Multivariate Analysis*, 100, 2044–2054.

Hansen, P. R., A. Lunde, and V. Voev (2014): "Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility," *Journal of Applied Econometrics*, 29, 774–799.

Kan, R., and C. Zhang (1999): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203–235.

Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149–173.

Lintner, J. (1965): "Security Prices, Risk, and Maximal Gains from Diversification," *The Journal of Finance*, 20, 587–615.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pedersen, R. S., and A. Rahbek (2014): "Multivariate Variance Targeting in the BEKK–GARCH Model," *The Econometrics Journal*, 17, 24–55.

Sharpe, W. (1964): "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.