Generalized Conditional Autoregressive Betas

Stefano Grassi[1] and Francesco Violante[2]

March 3, 2022

Abstract

This paper introduces a new multivariate model, dubbed Generalized Conditional Autoregressive Beta (GCAB) GARCH, for jointly modeling the time-varying slope coefficients in a multiple regression conditionally heteroskedastic system. The model, which implies a structure tailored to the linear asset pricing framework, allows the coexistence of constant and time-varying betas, simplifies testing (or imposing) cross-sectional restrictions and, introduces new mechanisms of propagation of shocks, namely beta spillovers, in a coherent, explicit and parsimonious way. We derive conditions for stationarity and uniform invertibility and, to mitigate the problem of parameter proliferation in large dimensions, we provide estimators for beta and covariance tracking. We propose a variety of parallel and sequential maximum likelihood estimators and, we investigate their finite sample properties of by means of extensive Monte Carlo experiments. Finally, the GCAB is used in the Fama-French three factors asset pricing framework using real data.

Keywords: Cholesky decomposition; Multivariate GARCH, Asset Pricing.

JEL Classification: C12, C22, G12, G13

---

[1]University of Rome 'Tor Vergata', Via Columbia 2, Rome, Italy. Phone: +390672595914 - E-mail:stefano.grassi@uniroma2.it

[2]CREST, ILB, GENES-ENSAE, Institut Polytechnique de Paris, 5, avenue Henry Le Chatelier, TSA 96642, 91764 Palaiseau cedex, France. Phone: +33(0)170266750 - E-Mail:francesco.violante@ensae.fr

Stefano Grassi gratefully acknowledges financial support from the University of Rome 'Tor Vergata' by the 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 model rejection. Due to its intuitive appeal and ease of implementation, the 2-Pass Cross-Sectional Regression approach, introduced by Fama and MacBeth (1973), still represents a workhorse in the asset pricing literature. Time variation in the beta coefficients is attained by least square estimation of a linear factor asset pricing model on a rolling sample, resulting in dynamic properties of the conditional betas heavily determined by the estimation window size. Furthermore, the rolling OLS makes it difficult to distinguish between sampling variability and actual time variation in the beta coefficients.

To overcome these issues González et al. (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. Furthermore, the model assumes orthogonality of the risk factors, which is often rejected in practice.

Hansen et al. (2014) also construct a model based on mixed frequencies which exploit additional information coming from non-parametric realized measures of variances and correlations. However, because they model 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) proposed to model dynamic conditional betas using dynamic conditional correlations (DCB-DCC). In the DCB-DCC model, the betas are recovered as a (non-linear) 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. Also, this modeling approach makes the coexistence of dynamic and constant betas challenging because constancy of the betas, or some of them, requires constraints on the structure of the conditional covariance matrix itself and not simply on the parameters governing its dynamics. Indeed, the constancy constraint is imposed a priori and, then statistically validated ex-post, rather than via testable parameter restrictions on the conditional covariance dynamics.

In a recent paper, Darolles et al. (2018) propose a model for time-varying betas based on the Cholesky decomposition of the conditional covariance matrix of the factors-assets system. In their model, dubbed Cholesky-GARCH (CHAR), they explicitly specify the dynamics of the conditional betas by means of a parametric model. This simplifies the coexistence of constant and time-varying betas which can be recovered via simple parameter restrictions. The authors show that the CHAR proves superior to the DCB-DCC in terms of forecasting and beta hedging. However, in asset pricing applications, the CHAR model shows limitations stemming from the plain triangular Cholesky decomposition and more so when the cross-sectional dimension of the number of assets is large. Indeed, the CHAR is subject to dependence on the causal ordering of the coordinates implied by the Cholesky factorization. This means imposing a structure not just between factors and assets (which is natural), but also within the set of factors and the set of assets (which instead is arbitrary). As a consequence, the model is not invariant to permutation of the elements within each set. Also, because of the structure imposed within the set of assets, a linear asset pricing interpretation of the conditional mean of the assets is straightforward only when the system considered consists of possibly multiple risk factors but only a single asset. When multiple assets are considered, the triangular structure of the model introduces nuisance conditional betas, merely accounting for the correlation between assets, whose number increases quadratically with the number of assets. The presence of nuisance betas is not only unnatural in the asset pricing framework, but also inevitably rises the problem of parameter proliferation in large systems.[1] In asset pricing applications, Darolles et al. (2018) avoid the problem by marginalizing the distribution of the multi-asset system iteratively, isolating risk factors and a single asset at the time. Although valid, this approach limits the type of hypotheses on the asset pricing model that can be tested, particularly restrictions on the cross-section of assets.

Inspired by the recent literature, we introduce a new model tagged Generalized Conditional Autoregressive Beta (GCAB). The GCAB achieves orthogonalization between two sets of variables, in the asset pricing context the risk factors and the investment assets, via a Block-Cholesky de-

[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).

composition of the system covariance matrix. This provides several direct advantages.

First, the GCAB entails a causal ordering between the sets of factors and assets, i.e. it preserves the natural hierarchy existing between risk factors and assets, without imposing a structure within each set of variables. Consequently, the model is invariant to permutations of the coordinates in each set.

Second, in the asset pricing framework, the GCAB allows for a coherent and parsimonious multivariate evaluation of a multi-asset system. More precisely, the GCAB only requires modeling the conditional betas measuring the exposure of assets to risk factors, whose number grows linearly in the number of assets. Co-movements between variables, in each block, are instead modeled explicitly via time-varying covariances or correlations.

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. Because such innovations drive the conditional betas, this can potentially impact their dynamics in an undesirable way. The block orthogonalization of the GCAB, instead, only underlies a reduction of the magnitude of the orthogonalized 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 immediately interpretable way. Mirroring the dual nature of the block decomposition, we isolate two different sources of spillovers originating from factors or assets. A factor spillover occurs when shocks to the exposures of asset i to factor m causes a lagged impact on the exposure of asset i to factor $j \neq m$. Similarly, asset spillovers represent the transmission on the exposure of asset i to factor j of changes in the exposure of competing assets $n \neq i$, to the same factor j. Beta spillovers mimic a transmission mechanism similar, yet more complex, to the covariance/correlation spillovers familiar in the multivariate GARCH literature. Asset spillovers can act across any pair of assets because the block structure allows modelling their cross-section jointly and symmetrically in terms of cross sectional interactions. In this sense, our model bypasses the causal ordering of the coordinates imposed by the triangular structure of the CHAR, and thus its inherent sequentiality.

After introducing our generalized conditional beta specification and various specific sub-models, we study conditions for stationarity and uniform invertibility. Also, we provide solutions to the parameter proliferation problem in large dimensions in the form of beta tracking and block-orthogonalized innovations' covariance targeting. Concerning computational feasibility, we discuss alternative multi-step estimators computationally more convenient than the full quasi maximum estimator in large cross-sectional dimensions. We illustrate the finite sample properties of such estimators by Monte Carlo simulation.

We test our model in the context of the Fama and French (1993) three-factor framework. The empirical application aims at benchmarking a conditional beta specification driven only by idiosyncratic shocks against three alternative specifications accounting for beta spillovers. We perform a comprehensive historical analysis of the bivariate asset system composed of the Coal and Petroleum-Natural Gas value-weighted industry portfolios studied over a period spanning from January 1, 1927, to November 30, 2020. Besides 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 Petroleum-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 rest of the paper is organized as follows. Section 2 briefly introduces the notation and operators used throughout the paper. Section 3 describes the model while Section 4 discusses a generalized conditional beta specification, several sub-models obtained under parameter restrictions, the model's theoretical properties. Monte Carlo simulations results are reported in Section 5 and the empirical application in Section 6. Finally, Section 7 concludes.

2 Notation

We make use of the following matrix notation and operators. For any matrix $\mathbf{A}$ partitioned in blocks, we denote $\mathbf{A}_{ij}$ the ij-th block of $\mathbf{A}$ and, $a_{ij,[kl]}$ the kl element of the ij-th block of $\mathbf{A}$. The same notation, albeit using a single index, is used to denote a vector's partitions and its elements. 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 $\mathbf{0}_{(r)}$ denotes the null vector of size $r \times 1$. Furthermore, we denote the matrix square root operator, $\mathbf{A}^{1/2}$, any factorization of a positive-definite matrix $\mathbf{A}$ such that $\mathbf{A} = \mathbf{A}^{1/2}\left(\mathbf{A}^{1/2}\right)'$. For a square matrix $\mathbf{A}$, $\mathbf{a} = \text{vech}\left(\mathbf{A}_{(c)}\right)$ defines the $c(c+1)/2 \times 1$ vector that stacks the lower triangular portion, including the main diagonal, of $\mathbf{A}$. Similarly, define

$\mathbf{a} = \mathrm{vec}(\mathbf{A})$, the $rc \times 1$ vectorization of an $r \times c$ matrix $\mathbf{A}$ obtained by stacking its columns, and denote $\mathbf{A} = \mathrm{vec}^{-1}_{(r\times c)}(\mathbf{a})$ its inverse. Also, we define $\mathbf{a} = \mathrm{diag}\left(\mathbf{A}_{(c)}\right)$ the $c \times 1$ vector holding the main diagonal of $\mathbf{A}_{(c)}$.

For any two vectors $\mathbf{a}$ of size $r \times 1$ and $\mathbf{b}$ of size $c \times 1$, $\mathbf{a} \otimes \mathbf{b} = \mathrm{vec}\left(\mathbf{ab}'\right)$, 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 $\mathbf{A} \odot \mathbf{A}^{\odot -1} = \mathbf{e}_{(r)}\mathbf{e}^{'}_{(c)}$. Finally, $\mathcal{I}_{t-1}$ denotes the information set generated by past values of $\mathbf{x}_t$ and, $\mathrm{E}_{t-1}[\mathbf{x}_t] = \mathrm{E}[\mathbf{x}_t|\mathcal{I}_{t-1}]$, the conditional expectation operator. 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 $\mathrm{E}\left[\mathbf{x}_{1,t}\mathbf{x}^{'}_{2,t}\right] = \mathbf{0}$, i.e. $\mathbf{x}_{1,t}$ is statistically orthogonal to $\mathbf{x}_{2,t}$.

3 Model setup

Let $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}^{1/2}_t \boldsymbol{\eta}_t$, be a vector of heteroskedastic random variables, with conditional covariance matrix $\boldsymbol{\Sigma}_t = \mathrm{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}^{'}_t|\mathcal{I}_{t-1};\boldsymbol{\theta}\right]$, positive-definite, $\mathcal{I}_{t-1}$-measurable and depending on a vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ convex and, with $\boldsymbol{\eta}_t$ an independent and identically distributed random vector with zero mean and identity covariance matrix.

In linear asset pricing applications the elements of $\boldsymbol{\epsilon}_t$ are categorized in two groups: k market-wide risk factors and n investment assets returns. These applications aim at quantifying the dependence, also called the exposure, of each investment assets to the risk factors. Because such exposures are typically modeled in a linear regression framework, in the financial literature they are commonly referred to as the assets' "betas". Endowed with this categorization, we partition accordingly the elements of $\boldsymbol{\epsilon}_t$, $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\eta}_t$ into two blocks respectively of size k and n, that is

$$\boldsymbol{\epsilon}_t = \begin{bmatrix} \boldsymbol{\epsilon}_{1,t} \\ \boldsymbol{\epsilon}_{2,t} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_t = \begin{bmatrix} \boldsymbol{\Sigma}_{11,t} & \boldsymbol{\Sigma}^{'}_{12,t} \\ \boldsymbol{\Sigma}_{12,t} & \boldsymbol{\Sigma}_{22,t} \end{bmatrix}, \qquad \boldsymbol{\eta}_t = \begin{bmatrix} \boldsymbol{\eta}_{1,t} \\ \boldsymbol{\eta}_{2,t} \end{bmatrix}. \tag{1}$$

Because $\boldsymbol{\Sigma}_t$ is symmetric and positive definite, a conformable block Cholesky decomposition yields

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

where $\mathbf{L}_t$ is lower unit-triangular holding its identity matrix on the diagonal blocks, with inverse $\mathbf{L}^{-1}_t = \begin{bmatrix} \mathbf{I}_{(k)} & \mathbf{0}_{(k\times n)} \\ -\mathbf{B}_t & \mathbf{I}_{(n)}, \end{bmatrix}$, and where $\mathcal{S}_t$ admits $\mathcal{S}_t = \mathcal{S}^{1/2}_t\left(\mathcal{S}^{1/2}_t\right)'$, because its diagonal blocks,

$\mathcal{S}_{11,t} = \mathbf{\Sigma}_{11,t}$ and $\mathcal{S}_{22,t}$, the Schur complement of $\mathbf{\Sigma}_{22,t}$, are symmetric and positive-definite. Hence, it holds

$$\mathbf{L}_t^{-1}\boldsymbol{\epsilon}_t = \mathcal{S}_t^{1/2}\boldsymbol{\eta}_t, \tag{3}$$

or more explicitly $\boldsymbol{\epsilon}_{1,t} = \mathcal{S}_{11,t}^{1/2}\boldsymbol{\eta}_{1,t}$, and $\boldsymbol{\epsilon}_{2,t} = \mathbf{B}_t\boldsymbol{\epsilon}_{1,t} + \mathcal{S}_{22,t}^{1/2}\boldsymbol{\eta}_{2,t}$. Each row of the $n \times k$ matrix $\mathbf{B}_t$ collects the time-varying coefficients, i.e. the conditional betas, of the multivariate heteroskedastic regressions of $\epsilon_{2,[i],t}$ on $\boldsymbol{\epsilon}_{1,t}$, $i \in [1, n]$. This construction has several advantages: the model is invariant to permutations of the elements within each block while maintaining the causal ordering between blocks ($\boldsymbol{\epsilon}_{1,t}$ causes $\boldsymbol{\epsilon}_{2,t}$), the conditional betas can be modeled jointly (facilitating e.g. imposing cross-sectional constraints, modeling cross-feedback), the correlation structure of the first block and that of the beta neutralized second block are explicitly modeled.

The conditional covariance matrices $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$, as well as the conditional beta matrix $\mathbf{B}_t$ are assumed to be $\mathcal{I}_{t-1}$-measurable dynamic processes. In the reminder of the paper we assume $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$ following each a distinct stationary multivariate GARCH process, see Bauwens et al. (2006), Silvennoinen and Terasvirta (2009) and Boudt et al. (2019) for surveys. Finally, it is worth stressing that positive definiteness of $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$, together with (2) ensures positive definiteness of $\mathbf{\Sigma}_t$, $\forall t$, for any $\mathbf{B}_t \in \mathbb{R}^{n\times k}$. Hence, this setup does not require imposing any restriction on the sign and size of the elements of $\mathbf{B}_t$.[2] Dynamics of $\mathbf{B}_t$ and its properties are discussed in the next Section.

4 Generalized Conditional Autoregressive Beta dynamics

For the $\mathcal{I}_{t-1}$-measurable process $\mathbf{B}_t$ we propose the recurrence

$$\mathbf{B}_t = \mathbf{\Psi} + \left[\text{vec}^{-1}_{(k\times n)}\left(\sum_{p=1}^{P}\mathbf{\Omega}_p\left(\mathbf{v}_{2,t-p}\otimes\mathbf{v}_{1,t-p}\right) + \sum_{q=1}^{Q}\mathbf{\Gamma}_q\text{vec}\left(\mathbf{B}'_{t-q}\right)\right)\right]', \tag{4}$$

where $\mathbf{\Psi}$ (of dimension $n \times k$), $\mathbf{\Omega}_p$ and $\mathbf{\Gamma}_q$ (both of dimension $nk \times nk$) are matrices of static parameters. $\mathbf{B}_t$ is updated via the outer products of past block-orthogonal innovations, namely $\mathbf{v}_{1,t} = \boldsymbol{\epsilon}_{1,t}$ and $\mathbf{v}_{2,t} = \boldsymbol{\epsilon}_{2,t} - \mathbf{B}_t\boldsymbol{\epsilon}_{1,t}$, respectively. This choice relates to Koopman et al. (2013), in that the product of block-orthogonal innovations is proportional to the updating term emerging in their score-driven time-varying parameter regression framework.

[2] Parameter restriction ensuring positive definiteness of $\mathcal{S}_{11,t}$, and $\mathcal{S}_{22,t}$ depend on the specific model chosen and are extensively discussed in the mentioned references.

Remark 1 *An advantage of (4) is that the coexistence of constant and time-varying betas achievable via simple parameters restrictions. Denoting* $\boldsymbol{\omega}_{s,p}$ $(\boldsymbol{\gamma}_{s,p})$ *the s-th row of* $\boldsymbol{\Omega}_p$ $(\boldsymbol{\Gamma}_p)$*, the restriction* $\boldsymbol{\omega}_{s,p} = \boldsymbol{\gamma}_{s,p} = \mathbf{0}_{(nk)}$ *for some* $s \in [1, nk]$ *and* $\forall p$*, implies* $\beta_{s,t} = \beta_s$*, for the s-th element of* $\boldsymbol{\beta} = \mathrm{vech}\left(\mathbf{B}_{\mathrm{t}}^{'}\right)$*. Also, (4) nests the constant beta marginal regression* $\epsilon_{2,[i],t} = \sum_{j=1}^{k} \beta_{ij}\epsilon_{1,[j],t} + v_{2,[i],t}$ *for some* $i \in [1, n]$*. Denoting* $\boldsymbol{\Omega}_{ij,p}$ $(\boldsymbol{\Gamma}_{ij,p})$ *the* (i, j) $k \times k$ *block of* $\boldsymbol{\Omega}_p$ $(\boldsymbol{\Gamma}_p)$*, this is achieved when* $\boldsymbol{\Omega}_{ij,p} = \boldsymbol{\Gamma}_{ij,p} = \mathbf{0}_{(k\times k)}$ $\forall j \in [1, n]$ *and* $\forall p$*. Obviously, nullity of* $\boldsymbol{\Omega}_p$ *(thus* $\boldsymbol{\Gamma}_p$*)* $\forall p$ *implies constancy of all elements of* $\mathbf{B}_t$*. By construction, (3) reduces to the constant parameters heteroskedastic multiple linear regressions framework, i.e.* $\mathbf{B}_t = \boldsymbol{\Psi} = \mathrm{E}\left[\boldsymbol{\Sigma}_{12,\mathrm{t}}\right] \mathrm{E}\left[\boldsymbol{\Sigma}_{11,\mathrm{t}}\right]^{-1}$*.*

The inclusion of all cross-products of the elements of $\mathbf{v}_{1,t}$ and $\mathbf{v}_{2,t}$, as well as lagged betas, totals $nk(1 + 2nk(P + Q))$ static parameters. This makes (4) very general, but also computationally cumbersome in large dimensions. We next discuss restrictions which yield more parsimonious specifications suited to financial applications. These parameterisations fall in two categories: betas driven solely by idiosyncratic shocks and beta spillovers. Without loss of generality and to simplify the notation, we fix lag orders to $P = Q = 1$.

4.1 Betas driven solely by idiosyncratic shocks

The dynamics of $\beta_{ij,t}$ is driven solely by the corresponding (idiosyncratic) updating term $v_{2,[i],t-p}v_{1,[j],t-p}$ and own lag. This restriction renders the model feasible in very large cross-sectional dimensions (both k and n) and, it is comparable in essence e.g. to the scalar/diagonal restriction of the BEKK model of Engle and Kroner (1995) or the scalar dynamics of the DCC model of Engle (2002). In this first category, listed from most to least restricted, we find the following sub-models (in the reminder of the paper the label in parentheses will be used to refer to the particular sub-model):
Scalar (4.i): Our most parsimonious parameterisation, tailored to accommodate large cross-sectional dimensions. It imposes common dynamics of $\beta_{ij,t}$ for all i and j. It is obtained from (4) under the restrictions $\boldsymbol{\Omega} = \omega\mathbf{I}_{(nk)}$ and $\boldsymbol{\Gamma} = \gamma\mathbf{I}_{(nk)}$.
Semi-scalar (4.ii): It is a scalar model in the columns of $\mathbf{B}_t$. For each $j = 1, \ldots, k$, $\beta_{ij,t}$ shares the same dynamics across all $i \in [1, n]$. It is obtained under the restrictions $\boldsymbol{\Omega} = \mathbf{I}_{(n)} \otimes \overline{\boldsymbol{\Omega}}$ and $\boldsymbol{\Gamma} = \mathbf{I}_{(n)} \otimes \overline{\boldsymbol{\Gamma}}$ where $\overline{\boldsymbol{\Omega}}$ and $\overline{\boldsymbol{\Gamma}}$ are $k \times k$ diagonal matrices.
Heterogeneous (4.iii): Each element $\beta_{ij,t}$ is driven by its own set of parameters. It is obtained under diagonality of $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$. Under this restriction (4) can be equivalently expressed, using the Hadamard product, as:

$$\mathbf{B}_t = \boldsymbol{\Psi} + \left(\widetilde{\boldsymbol{\Omega}} \odot \mathbf{v}_{2,t-1}\mathbf{v}_{1,t-1}^{'}\right) + \left(\widetilde{\boldsymbol{\Gamma}} \odot \mathbf{B}_{t-1}\right), \qquad (4.\mathrm{iii})$$

where $\widetilde{\mathbf{\Omega}} = \left[\text{vec}^{-1}_{(k\times n)}\left(\text{diag}\left(\mathbf{\Omega}\right)\right)\right]^{'}$ and $\widetilde{\mathbf{\Gamma}} = \left[\text{vec}^{-1}_{(k\times n)}\left(\text{diag}\left(\mathbf{\Gamma}\right)\right)\right]^{'}$. This model strikes a good compromise between flexibility and feasibility in large dimension.[3]

4.2 Beta Spillovers

Spillovers on $\beta_{ij,t}$ are introduced via the products $v_{2,[r],t-p}v_{1,[s],t-p}$, $\{r, s\} \neq \{i, j\}$ and corresponding lagged betas. Mirroring the partition in (1), we isolate two sources of spillovers, dubbed *factor* and *asset*, in explicit reference to the aforementioned role of $\boldsymbol{\epsilon}_{1,t}$ and $\boldsymbol{\epsilon}_{2,t}$ in asset pricing applications. A factor spillover occurs when shocks to the exposures of asset $i \in [1, n]$ to factor $m \in [1, k]$ causes a lagged impact on the exposure of asset i to factor $j \neq m$, i.e. $\beta_{ij,t}$. Similarly, asset spillovers represent the transmission on $\beta_{ij,t}$ of changes in the exposure of competing assets $m \in [1, n], m \neq i$ to factor j. Beta spillover mimic a transmission mechanism similar, yet more complex, to covariance/correlation spillovers, familiar in the multivariate GARCH literature.

Factor spillovers (4.iv): Spillovers on $\beta_{ij,t}$ act via the products $\left\{v_{2,[i],t-1}v_{1,[s],t-1}\right\}_{s=1,\dots,k,s\neq j}$ and lagged betas $\{\beta_{is,t-1}\}_{s=1,\dots,k,s\neq j}$. To obtain this model, $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ in (4) are block diagonal with n blocks of size $k \times k$ on the main diagonal, i.e. $\mathbf{\Omega}_{ij} = \mathbf{\Gamma}_{ij} = \mathbf{0}_{(k\times k)}$ $i, j \in [1, n]$, $i \neq j$.

Asset spillovers (4.v): Spillovers on $\beta_{ij,t}$ enter via $\left\{v_{2,[s],t-1}v_{1,[j],t-1}\right\}_{s=1,\dots n,s\neq j}$ and lagged betas $\{\beta_{sj,t-1}\}_{s=1,\dots n,s\neq j}$. This model is obtained when $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ are $n \times n$ block matrices with $k \times k$ diagonal blocks. Since all blocks are diagonal, we can define a $nk \times nk$ permutation matrix $\mathbf{P}$ such that $\mathbf{\Omega}^{*} = \mathbf{P}\mathbf{\Omega}\mathbf{P}^{'}$ is block diagonal with $n \times n$ blocks on the main diagonal. By reordering accordingly the terms in (4), this sub-model can be written alternatively as

$$\mathbf{B}_t = \mathbf{\Psi} + \text{vec}^{-1}_{(n\times k)}\Big(\mathbf{\Omega}^{*}\left(\mathbf{v}_{1,t-1} \otimes \mathbf{v}_{2,t-1}\right) + \mathbf{\Gamma}^{*}\text{vec}\left(\mathbf{B}_{t-1}\right)\Big), \tag{4.v}$$

where $\mathbf{\Omega}^{*}$ and $\mathbf{\Gamma}^{*}$ are such that $\mathbf{\Omega}^{*}_{ij} = \mathbf{\Gamma}^{*}_{ij} = \mathbf{0}_{(n\times n)}$ $i, j \in [1, k]$, $i \neq j$.

Factor & asset spillovers (4.vi): This empirically relevant parameterisation mixes both types of spillovers described above. $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ in (4) are $n \times n$ 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 matrices if $i \neq j$, $i, j \in [1, n]$.

4.3 Stationarity, targeting and uniform invertibility

In this Section we derive the statistical properties of the model (3)-(4). Where possible, we also provide explicit results for the sub-models in Section 4.1 and 4.2. All proofs are in the Appendix.

[3] Alternatively (4.i) can be obtained from (4.iii) when $\widetilde{\mathbf{\Omega}} = \omega\mathbf{e}_{(n)}\mathbf{e}^{'}_{(k)}$ and $\widetilde{\mathbf{\Gamma}} = \gamma\mathbf{e}_{(n)}\mathbf{e}^{'}_{(k)}$. Similarly, (4.ii) can be obtained from (4.iii) under the restrictions $\widetilde{\mathbf{\Omega}} = \mathbf{e}_{(n)}\boldsymbol{\omega}^{'}$ and $\widetilde{\mathbf{\Gamma}} = \mathbf{e}_{(n)}\boldsymbol{\gamma}^{'}$, where $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$ are $k \times 1$ vectors.

Theorem 1 (Stationarity) *Under the following conditions:*

C.1) $\mathbf{z}_t = (\mathbf{z}_{1,t}, \mathbf{z}_{2,t})$ *with* $\mathbf{z}_{i,t} = (\mathbf{v}_{i,t}, \mathrm{vec}(\mathcal{S}_{\mathrm{ii,t}}))'$ *for* $i = 1, 2$, *is stationary and ergodic;*

C.2) $|\mathbf{\Gamma}_Q(z)| = \left|\mathbf{I}_{(nk\times nk)} - \sum_{q=1}^{Q} \mathbf{\Gamma}_q z^q\right| \neq 0$ *for all* $|z| \leq 1$,

then: (a) $\mathbf{B}_t$, *is stationary and ergodic, (b)* $\mathbf{B} \equiv \mathrm{E}[\mathbf{B}_\mathrm{t}] = \left[\mathrm{vec}^{-1}_{(\mathrm{k}\times\mathrm{n})}\left(\mathbf{\Gamma}_\mathrm{Q}(1)^{-1}\,\mathrm{vec}\left(\mathbf{\Psi}'\right)\right)\right]'$, *(c) a stationary ergodic solution* $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{1,t}, \boldsymbol{\epsilon}_{2,t})'$ *with* $\boldsymbol{\epsilon}_{1,t} = \mathbf{v}_{1,t}$ *and* $\boldsymbol{\epsilon}_{2,t} = \left(\mathbf{B}_t,\ \mathbf{I}_{(n)}\right)(\mathbf{v}_{1,t}, \mathbf{v}_{2,t})'$ *exists.*

For the six sub-models in Section 4.1 and 4.2, condition C.2 can be written more explicitly. To this end, denote $\lambda_j^{max}(\mathbf{A})^+$ the largest absolute eigenvalue of $\mathbf{A}$.

Corollary 1 *Condition C.2 simplifies to: (a)* $|\gamma| < 1$ *for (4.i), (b)* $\max\limits_{s\in[1,k]} |\overline{\gamma}_{ss}| < 1$ *(largest absolute diagonal element of* $\overline{\mathbf{\Gamma}}$*) for (4.ii), (c)* $\max\limits_{s\in[1,kn]} |\gamma_{ss}| < 1$ *(largest absolute diagonal element of* $\mathbf{\Gamma}$*) for (4.iii), (d)* $\max\limits_{i\in[1,n]} \lambda_i^{max}(\mathbf{\Gamma}_{ii})^+ < 1$ *where* $\mathbf{\Gamma}_{ii}$ *is the i-th* $k \times k$ *diagonal block of* $\mathbf{\Gamma}$ *for (4.iv), (e)* $\max\limits_{j\in[1,k]} \lambda_j^{max}(\mathbf{\Gamma}^*_{jj})^+ < 1$, *where* $\mathbf{\Gamma}^*_{jj}$ *is the j-th* $n \times n$ *diagonal block of* $\mathbf{\Gamma}^*$ *for (4.v), (f)* $\lambda^{max}(\mathbf{\Gamma})^+ < 1$ *for (4.vi).*

Even under the most parsimonious specifications, (4) is subject to the curse of dimensionality because of the $n \times k$ intercept $\mathbf{\Psi}$. However, provided a sample estimator for $\mathrm{E}[\mathbf{B}_t]$ is available, Theorem 1 *(b)* delivers a simple way to reduce the parameter space by means of beta targeting. To this end, let us consider the unconditional covariance between $\boldsymbol{\epsilon}_{1,t}$ and $\boldsymbol{\epsilon}_{2,t}$. Using (3) and recalling $\mathbf{v}_{1,t} = \boldsymbol{\epsilon}_{1,t}$ and $\mathbf{v}_{2,t} = \boldsymbol{\epsilon}_{2,t} - \mathbf{B}_t\boldsymbol{\epsilon}_{1,t}$, then $\mathrm{E}\left[\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{1,t}'\right] \equiv \mathbf{\Sigma}_{21} = \mathrm{E}\left[\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}'\right] + \mathrm{E}[\mathbf{v}_{2,t}\mathbf{v}_{1,t}']$. By the law of iterated expectations and the block orthogonality, $\mathbf{\Sigma}_{21} = \mathrm{E}\left[\mathbf{B}_t\mathcal{S}_{11,t}\right]$.

Theorem 2 (OLS targeting) *For the stationary model* (3)-(4)*, condition*

C.3) $\mathbf{v}_t = (\mathbf{v}_{1,t}, \mathbf{v}_{2,t})' \,|\, \mathcal{I}_{t-1} \sim \mathrm{N}\left(\mathbf{0}_{(k+n)}, \mathcal{S}_t\right)$ *is multivariate normally distributed and* $\partial\mathcal{S}_{ii,t}/\partial\mathbf{v}_{j,t-s} = 0,\ \forall s > 0,\ i, j \in [1, 2],\ i \neq j$

is sufficient to ensure $\mathrm{E}\left[\mathbf{B}_t\mathcal{S}_{11,t}\right] = \mathbf{B}\mathbf{\Sigma}_{11}$ *and* $\mathbf{\Psi} = \left[\mathrm{vec}^{-1}_{(k\times n)}\left(\mathbf{\Gamma}_Q(1)\,\mathrm{vec}\left(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\right)\right)\right]'$.

Condition *C.3* ensures that $\mathbf{B}_t$ and $\mathcal{S}_{11,t}$ are unconditionally uncorrelated. Theorem 2 shows that (3)-(4) is internally consistent in that the unconditional level of $\mathbf{B}_t$ equals the constant slope regression coefficients, i.e. $\mathrm{E}[\mathbf{B}_\mathrm{t}] \equiv \mathbf{B} = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$, constructed via the unconditional version of (2). This favorable property provides a simple estimator for beta targeting, namely the ordinary least square.

We now elaborate on the possibility of covariance tracking for the second block system as a further layer of reduction of the parameter space. From (2) we obtain $\mathrm{E}\left[\boldsymbol{\epsilon}_{1,t}\boldsymbol{\epsilon}_{1,t}^{'}\right] \equiv \boldsymbol{\Sigma}_{11} = \mathrm{E}[\mathcal{S}_{11,t}]$ and, under Theorem 2, $\mathrm{E}\left[\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{1,t}^{'}\right] \equiv \boldsymbol{\Sigma}_{21} = \mathbf{B}\boldsymbol{\Sigma}_{11}$. For the remaining block of $\mathrm{E}\left[\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t}^{'}\right]$ we have:

$$\mathrm{E}\left[\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{2,t}^{'}\right] \equiv \boldsymbol{\Sigma}_{22} = \mathrm{E}\left[\mathbf{B}_t\boldsymbol{\Sigma}_{11,t}\mathbf{B}_t^{'}\right] + \mathrm{E}\left[\mathcal{S}_{22,t}\right].$$

Second block covariance tracking requires availability of a sample estimator for $\mathrm{E}[\mathcal{S}_{22,t}]$. A natural candidate would be the Schur complement of $\boldsymbol{\Sigma}_{22}$, i.e. $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. However this estimator is not in general unbiased for $\mathrm{E}[\mathcal{S}_{22,t}]$, except for the limit case of constant beta, because it ignores the time variation in $\mathbf{B}_t$.

The unconditional covariance matrix $\mathrm{E}\left[\mathcal{S}_{22,t}\right]$ expressed in terms of the model's parameters and sample moments, is in general a highly complicated function which depends on 6-th order moments of the joint distribution of $\mathbf{v}_{1,t}$ and the unobserved $\mathbf{v}_{2,t}$. For sake of completeness, though limited by analytical tractability, we derive a semi-non-parametric estimator for $\mathcal{S}_{22}$ under the baseline beta specification (4.iii).

Theorem 3 (Second block covariance targeting) *For the stationary model* (3)-(4.iii)*, under condition C.3,* $\mathrm{E}[\mathcal{S}_{22,t}]$ *is given by:*

$$\mathrm{E}[\mathcal{S}_{22,t}] = \mathcal{A}^{\odot -1} \odot \left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right), \tag{5}$$

where $\mathcal{A} = \left(\mathbf{e}_{(n)}\mathbf{e}_{(n)}^{'} + \sum_{s=0}^{\infty}\mathcal{A}_s\right)$ *is finite and symmetric,* $\{\mathcal{A}_s\}_{s=0}^{\infty}$ *is an element-wise exponentially decaying sequence of matrices with typical element* $\alpha_{ij,s} = \mathbf{a}_{i,s}\mathbf{R}_s\mathbf{a}_{j,s}^{'}$, $i,j = 1,\ldots,n$, $s = 0,\ldots,\infty$ *and where* $\mathbf{a}_{i,s}$ *is the* i*-th row of the matrix of parameters* $\mathbf{A}_s = \widetilde{\boldsymbol{\Omega}} \odot \widetilde{\boldsymbol{\Gamma}}^{\odot s}$ *and,* $\mathbf{R}_s = \mathrm{E}\left[\left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{1,t-s-1}^{'}\right) \odot \left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\right]$ *with* $\lim_{s\to\infty}\mathbf{R}_s = \boldsymbol{\Sigma}_{11}^{\odot 2}$.

Theorem 3 shows that, in presence of time variation in the betas, the unconditional variance of $\mathbf{v}_{2,t}$ is a scaled version of the Schur complement of $\boldsymbol{\Sigma}_{22}$. Also, as mentioned above, in the limiting case $\widetilde{\boldsymbol{\Omega}} = \mathbf{0}$, i.e. $\mathbf{B}_t = \mathbf{B}$, $\forall t$, then $\mathcal{A} = \mathbf{e}_{(n)}\mathbf{e}_{(n)}^{'}$ and $\mathrm{E}[\mathcal{S}_{22,t}] = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ which conforms with the constant parameter regression case. The blocks of the unconditional covariance matrix $\boldsymbol{\Sigma} \equiv \mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\prime})$ in (5) can be estimated non-parametrically by the corresponding sample estimators. Furthermore, $\mathbf{R}_s$, which relates to the $(s+1)-$th order element-wise autocovariance function of the outer product of first block innovations, can be estimated either via sample autocovariance or, where analytically feasible, expressed in terms of the parameters of $\mathcal{S}_{11,t}$. To illustrate this, in the following Corollary, we derive the explicit solution for the scaling factor $\mathcal{A}$ in a system with $k = 1$ and arbitrary $n \geq 1$ and, where the (univariate) first block process is a GARCH(1,1).

Corollary 2 *For the stationary model* (3)-(4.iii) *with dimensions* $k = 1$ *and* $n \geq 1$, *i.e.* $\boldsymbol{\epsilon}_t = (v_{1,t}, \mathbf{B}_t v_{1,t} + \mathbf{v}_{2,t})' | \mathcal{I}_{t-1} \sim \mathrm{N}\left(\mathbf{0}_{(1+n)}, \mathcal{S}_t\right)$, $\mathbf{B}_t = \mathbf{B} \odot \left(\mathbf{e}_{(n)} - \widetilde{\boldsymbol{\Gamma}}\right) + \widetilde{\boldsymbol{\Omega}} \odot \left(\mathbf{v}_{2,t-1} v_{1,t-1}'\right) + \widetilde{\boldsymbol{\Gamma}} \odot \mathbf{B}_{t-1}$ *and* $\mathcal{S}_{11,t} = c + \tau_1 v_{1,t-1}^2 + \delta_1 \mathcal{S}_{11,t-1}$, *it holds:*

$$\mathbf{R}_s \equiv \mathrm{E}\left[v_{1,t}^2 v_{1,t-s-1}^2\right] = (\tau_1 + \delta_1)^{s+1}\left(\mu_4 - \mu_2^2\right)\kappa + \mu_2^2, \qquad s \in [0, \infty), \tag{6}$$

where $\mu_2 \equiv \mathrm{E}\left[v_{1,t}^2\right] = \frac{c}{1-\tau_1-\delta_1}$, $\mu_4 \equiv \mathrm{E}\left[v_{1,t}^4\right] = \frac{3\mu_2^2(1-(\tau_1+\delta_1)^2)}{1-\delta_1^2-3\tau_1^2-2\tau_1\delta_1}$ *and* $\kappa = \frac{\tau_1(1-\delta_1(\tau_1+\delta_1))}{(1-2\tau_1\delta_1-\delta_1^2)(\tau_1+\delta_1)}$. *The explicit solution for the (ij)-th element of* $\mathcal{A}$ *is*

$$\alpha_{ij} = 1 + \omega_i \omega_j \left(\frac{(\tau_1 + \delta_1)\left(\mu_4 - \mu_2^2\right)\kappa}{1 - \gamma_i \gamma_j (\tau_1 + \delta_1)} + \frac{\mu_2^2}{1 - \gamma_i \gamma_j}\right), \tag{7}$$

where ω_i *and* γ_i, $i \in [1, n]$ *are the elements of the* $n \times 1$ *matrices of parameters* $\widetilde{\boldsymbol{\Omega}}$ *and* $\widetilde{\boldsymbol{\Gamma}}$.

Remark 2 *Theorem 3 and Corollary 2 directly extend to the sub-models (4.i) and (4.ii).*

Lastly, we provide conditions for uniform invertibility of (4), i.e. asymptotic irrelevance of initial states, which is a desirable property for both stability of the estimation and prediction. Let define $\mathbf{B}_t(\boldsymbol{\theta}; \mathbf{B}_0)$ $\forall t > 0$, for any $\boldsymbol{\theta}$ and for an arbitrary random initial state $\mathbf{B}_0$. Then (4) is uniformly invertible if $\mathbf{B}_t(\boldsymbol{\theta}; \mathbf{B}_0)$ is consistently approximated by the $\mathcal{I}_{t-1}$-measurable function $\mathbf{B}_t(\boldsymbol{\theta})$.

Theorem 4 (Uniform invertibility) *A sufficient condition for uniform invertibility of* (4) *is:*

$$\varsigma_\Gamma + \varsigma_\Omega \sum_{i=1}^{k} \mathrm{E}\left[\epsilon_{1,[i],1}^2\right] < 1, \tag{8}$$

where $\varsigma_\Gamma = \varsigma_{max}(\Gamma)$ *and* $\varsigma_\Omega = \varsigma_{max}(\Omega)$ *are the largest singular value of* $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Omega}$.

Corollary 3 *For the sub-models in Section 4.1 and 4.2, equation* (8) *further simplifies to: (a)* $\varsigma_\Gamma = |\gamma|$, $\varsigma_\Omega = |\omega|$ *for(4.i), (b)* $\varsigma_\Gamma = \max_{s\in[1,k]} |\overline{\gamma}_{ss}|$, $\varsigma_\Omega = \max_{s\in[1,k]} |\overline{\omega}_{ss}|$ *(largest absolute diagonal elements of* $\overline{\boldsymbol{\Gamma}}$ *and* $\overline{\boldsymbol{\Omega}}$*) for (4.ii), (c)* $\varsigma_\Gamma = \max_{s\in[1,nk]} |\gamma_{ss}|$, $\varsigma_\Omega = \max_{s\in[1,nk]} |\omega_{ss}|$ *(largest absolute diagonal elements of* $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Omega}$ *for (4.iii), (d)* $\varsigma_\Gamma = \max_{i\in[1,n]} \varsigma_{max}(\boldsymbol{\Gamma}_{ii})$ *and* $\varsigma_\Omega = \max_{i\in[1,n]} \varsigma_{max}(\boldsymbol{\Omega}_{ii})$, *(*$\boldsymbol{\Gamma}_{ii}$ *and* $\boldsymbol{\Omega}_{ii}$ *are the corresponding* $k \times k$ *diagonal blocks of* $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Omega}$*) for (4.iv), (e)* $\varsigma_\Gamma = \max_{j\in[1,k]} \varsigma_{max}\left(\boldsymbol{\Gamma}_{jj}^{*'}\right)$ *and* $\varsigma_\Omega = \max_{j\in[1,k]} \varsigma_{max}\left(\boldsymbol{\Omega}_{jj}^{*'}\right)$ *(*$\boldsymbol{\Gamma}_{jj}^*$ *and* $\boldsymbol{\Omega}_{jj}^*$ *are the corresponding* $n \times n$ *diagonal blocks of the diagonalizations of* $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Omega}$*) for (4.v). For (4.vi) condition* (8) *cannot be further simplified.*

Remark 3 *Although more explicit than the general conditions for stochastic recurrences with non-i.i.d. stationary coefficients given in Brandt (1986), the sufficient condition in* (8) *is likely to be overly conservative in practice. This is because it is an upper bound obtained chaining transformations and norm inequalities.*

4.4 Estimation

Consistency and asymptotic normality of the full quasi maximum likelihood estimator (QMLE) of (3)-(4) directly extends from Darolles et al. (2018).[4] In this Section, we discuss alternative multi-step QML estimators (parallel and sequential) computationally more efficient than the full QMLE in large cross-sectional dimensions. Finite sample performances of these estimators are studied in Section 5.

Denote $\phi\left(\boldsymbol{\epsilon}_t; \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right)$ the probability density function of a multivariate normally distributed random vector $\boldsymbol{\epsilon}_t$ with covariance matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ depending on a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact parameter space containing the population value of $\boldsymbol{\theta}$.

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$, a partition of the parameter vector according to the two sub-systems in (3). Furthermore, let us partition each of the vectors $\boldsymbol{\theta}_s = \left(\boldsymbol{\theta}_s^-, \boldsymbol{\theta}_s^{\mathrm{C}}\right)$, $s = 1, 2$ where the $\boldsymbol{\theta}_s^-$ is the subset of parameters of the marginal distributions of $\mathbf{v}_{s,t}$, and $\boldsymbol{\theta}_s^{\mathrm{C}}$ is the parameter vector of the Gaussian copula $\boldsymbol{\mathcal{C}}\left(\boldsymbol{\xi}_{s,t}; \mathrm{C}_{s,t}\left(\boldsymbol{\theta}_s^{\mathrm{C}}\right)\right)$, with $\boldsymbol{\xi}_{s,t} = \left(\mathrm{I} \odot \mathcal{S}_{ss,t}\right)^{-1/2} \mathbf{v}_{s,t}$, and $\mathrm{C}_{s,t}\left(\boldsymbol{\theta}_s^{\mathrm{C}}\right)$ the conditional correlation implied by $\mathcal{S}_{ss,t}$. Then $\phi\left(\boldsymbol{\epsilon}_t; \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right)$ can be factorized as follows:

$$
\begin{aligned}
\phi\left(\boldsymbol{\epsilon}_t; \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right) &= \phi\left(\mathbf{v}_{1,t}; \mathcal{S}_{11,t}(\boldsymbol{\theta}_1)\right) \phi\left(\mathbf{v}_{2,t}; \mathcal{S}_{22,t}(\boldsymbol{\theta}_2), \mathbf{B}_t(\boldsymbol{\theta}_2) | \boldsymbol{\epsilon}_{1,t}, \boldsymbol{\theta}_1\right) \quad (9) \\
&= \left[\prod_{i=1}^{k} \phi\left(v_{1,[i],t}; \mathcal{S}_{11,[ii],t}(\boldsymbol{\theta}_1^-)\right)\right] \boldsymbol{\mathcal{C}}\left(\boldsymbol{\xi}_{1,t}; \mathrm{C}_{1,t}\left(\boldsymbol{\theta}_1^{\mathrm{C}}\right)\right) \times \\
&\quad \left[\prod_{j=1}^{n} \phi\left(v_{2,[j],t}; \mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-), \mathbf{b}_{j,t}(\boldsymbol{\theta}_2^-) | \mathbf{v}_{1,t}, \boldsymbol{\theta}_1^-\right)\right] \boldsymbol{\mathcal{C}}\left(\boldsymbol{\xi}_{2,t}; \mathrm{C}_{2,t}\left(\boldsymbol{\theta}_2^{\mathrm{C}}\right)\right) \quad (10)
\end{aligned}
$$

where $\mathbf{v}_{1,t} = \boldsymbol{\epsilon}_{1,t}$, $\mathbf{v}_{2,t} = \boldsymbol{\epsilon}_{2,t} - \mathbf{B}_t \boldsymbol{\epsilon}_{1,t}$ and $\mathbf{b}_{j,t}$ is the j-th row of $\mathbf{B}_t$, $j = 1, \ldots, n$.

The first two approaches that we propose are based on factorization (9). While the last three are based on (10). Although they progressively reduce the complexity of the optimization problem, to be feasible they restrict the dynamics of $\mathcal{S}_{11,t}$, $\mathcal{S}_{22,t}$ and $\mathbf{B}_t$, in terms of cross-sectional interactions and constraints.

Method 1 (Full quasi-maximum likelihood estimator, $\mathcal{M}_1$) . *This method is based on di-*

[4] To see this, notice that in a bivariate system, i.e. $n = k = 1$, the two models perfectly overlap. This is because trivially standard and block Cholesky decomposition yield the same structure. The difference between the two approaches emerges from the fact that, starting from the bivariate system, to increase the cross-sectional dimension Darolles et al. (2018) populate the system before the orthogonalization. Contrary, we first operate the orthogonalization of a bivariate system and then we populate each block. Although, as $n > 1 \cup k > 1$, the structure imposed on $\boldsymbol{\epsilon}_t$ by two models diverges, (3)-(4) remains in nature a bivariate system regardless of the dimension of the two blocks.

rect maximization of equation (9):

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

where the factorization of the the joint density in $\mathcal{M}_1$ stems directly from (2).

Method 2 (Two-step block-by-block estimator, $\mathcal{M}_2$) . *This method is based on the factorization in equation (9):*

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

$\mathcal{M}_2$ is feasible whenever $\partial \mathcal{S}_{11,t}/\partial \mathbf{v}_{2,t-s} = \mathbf{0}$. It is identical to $\mathcal{M}_1$ if (additionally) $\partial \mathcal{S}_{22,t}/\partial \mathbf{v}_{1,t-s} = \mathbf{0} \cap \partial \mathbf{B}_t/\partial \mathbf{v}_{1,t-s} = \mathbf{0}$. Otherwise, $\mathcal{M}_2$ can be implemented sequentially, albeit entailing a loss of efficiency wrt $\mathcal{M}_1$.

Method 3 (Joint estimation of marginal distributions, $\mathcal{M}_3$) *This method is based on the joint maximization of the marginals in equation (10):*

$$\hat{\boldsymbol{\theta}}^- = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{t=1}^{T} \left[\prod_{i=1}^{k} \phi\left(v_{1,[i],t}; \mathcal{S}_{11,[ii],t}\left(\boldsymbol{\theta}_1^-\right)\right) \times \prod_{j=1}^{n} \phi\left(v_{2,[j],t}; \mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-), \mathbf{b}_{j,t}\left(\boldsymbol{\theta}_2^-\right) | \mathbf{v}_{1,t}, \boldsymbol{\theta}_1^-\right) \right]. \qquad (\mathcal{M}_3)$$

Method 4 (Estimation block-by-block via marginal distributions, $\mathcal{M}_4$) . *This method is based on the block factorization of the marginals in equation (10):*

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1^- &= \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{t=1}^{T} \prod_{i=1}^{k} \phi\left(v_{1,[i],t}; \mathcal{S}_{11,[ii],t}(\boldsymbol{\theta}_1^-)\right), \\ \hat{\boldsymbol{\theta}}_2^- &= \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{t=1}^{T} \prod_{j=1}^{n} \phi\left(v_{2,[j],t}; \mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_2^-), \mathbf{b}_{j,t}(\boldsymbol{\theta}_2^-) | \mathbf{v}_{1,t}, \boldsymbol{\theta}_1^-\right). \end{aligned} \qquad (\mathcal{M}_4)$$

This method requires $\partial \mathcal{S}_{11,t}/\partial \mathbf{v}_{2,t-s} = \mathbf{0}$. It is identical to $\mathcal{M}_3$ when $\mathcal{S}_{22,t}$ and $\mathbf{B}_t$ do not depend on $\boldsymbol{\theta}_1$, otherwise is a sequential estimator.

Method 5 (Estimation equation-by-equation, $\mathcal{M}_5$) *This method is based on element-wise factorization of the distribution of $(\mathbf{v}_{1,t}, \mathbf{v}_{2,t})'$:*

$$\hat{\boldsymbol{\theta}}^- = \begin{cases} \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{t=1}^{T} \phi\left(v_{1,i,t}; \mathcal{S}_{11,[ii],t}(\boldsymbol{\theta}_{1,i}^-)\right), \; i = 1, \ldots, k, \\ \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{t=1}^{T} \phi\left(v_{2,j,t}; \mathcal{S}_{22,[jj],t}(\boldsymbol{\theta}_{2,j}^-), \mathbf{b}_{j,t}(\boldsymbol{\theta}_j^-) | \mathbf{v}_{1,t}, \boldsymbol{\theta}_1^-\right), \; j = 1, \ldots, n, \end{cases} \qquad (\mathcal{M}_5)$$

where $\boldsymbol{\theta}_{s,i}^{-}$ is the subset of $\boldsymbol{\theta}_{s}^{-}$ driving the i-th diagonal element of $\mathcal{S}_{ss,t}$, $s=1,2$. To be feasible, this approach requires $\mathcal{S}_{11,t}$, $\mathcal{S}_{22,t}$ and $\mathbf{B}_t$ to entail minimal or no cross-sectional interactions and parameter restrictions.

Remark 4 *$\mathcal{M}_5$ relies on the highest degree of likelihood factorization. It is the most computationally efficient method in large dimensions because, by breaking the estimation in nk univariate problems, it avoids the inversion of large matrices. This method is feasible when $\mathbf{B}_t$ follows (4.i), (4.ii), (4.iii) or (4.iv) and in absence of spillovers in the dynamics of $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$. For the sub-models (4.v) and (4.vi), $\mathcal{M}_5$ remains feasible provided that asset spillovers are restricted to act solely via the updating term.*

The last three methods, $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_5$, are appealing when the correlations implied by $\mathcal{S}_{11,t}$ and $\mathcal{S}_{22,t}$ are not directly of interest (nuisance), see Francq and Zakoian (2016) for a related approach. Otherwise, these methods can be used as a preliminary step to the estimation of the conditional covariances/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 then be obtained by filtering (e.g scalar or diagonal BEKK) or they can be estimated in an additional step using $\boldsymbol{\mathcal{C}}\left(\boldsymbol{\xi}_{s,t}; \mathrm{C}_{s,t}\left(\boldsymbol{\theta}_{s}^{\mathrm{C}}\right)\right)$ (e.g. conditional correlation models, of Bollerslev, 1990, Engle, 2002 and Aielli, 2013). Such multi-step approach simplifies the second block covariance targeting because knowledge of the diagonal elements of $\mathcal{S}_{22}$ allows targeting the off-diagonal elements by means of sample estimators using prior 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 $i\neq j$.

4.5 Remarks on devolatilized innovations

In the specifications discussed in Section 4, $\mathbf{B}_t$ is updated via 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 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 reduce the impact of abnormally large shocks on $\mathbf{B}_t$, which are common in empirical applications, we propose to let the conditional betas depend on the product of devolatilized innovations, defined as $\boldsymbol{\xi}_{s,t} = \left(\mathbf{I}\odot\mathcal{S}_{ss,t}\right)^{-1/2}\mathbf{v}_{s,t}$, $s=1,2$. In this case, stationarity conditions and beta targeting, discussed in Section 4.3, remain valid. However, the use of $\left(\boldsymbol{\xi}_{2,t}\otimes\boldsymbol{\xi}_{1,t}\right)$ in (4) results in non explicit invertibility conditions because these products are related to past observations through highly nonlinear recursions, and less intuitive 2nd step covariance targeting.

Theorem 5 *For model (3)-(4) with $\mathbf{B}_t$ updated by $(\boldsymbol{\xi}_{2,t}\otimes\boldsymbol{\xi}_{1,t})$ the following properties hold: (a) $|\boldsymbol{\Gamma}_Q(z)|\neq 0$ for all $|z|\leq 1$, then $\mathbf{B}_t$ is stationary; (b) if $(\boldsymbol{\xi}_{1,t},\boldsymbol{\xi}_{2,t})'|\mathcal{I}_{t-1}$ is multivariate normally* *distributed and* $\partial \mathcal{S}_{ii,t}/\partial \boldsymbol{\xi}_{j,t-s} = \mathbf{0}$ $\forall s > 0$, $i,j \in [1,2]$ $i \neq j$, *then* $\mathrm{E}(\mathbf{B}_t) \equiv \mathbf{B} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$*; (c) for the sub-model* (4.iii), $\mathrm{E}(\mathcal{S}_{22,t}) = \boldsymbol{\Sigma}_{22} - \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}^{'} - \mathcal{R}_{\boldsymbol{\xi}_2} \odot \sum_{s=0}^{\infty} \mathcal{A}_s$*, with* $\mathcal{A}_s$ *such that its* (ij)*-th element* $\alpha_{ij,s} = \mathbf{a}_{i,s}\mathrm{E}\left[(\boldsymbol{\xi}_{1,t-s-1}\boldsymbol{\xi}_{1,t-s-1}^{'}) \odot (\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'})\right]\mathbf{a}_{j,s}^{'}$, $i,j \in [1,n]$*, with* $\mathbf{a}_{i,s}$ *defined in Theorem 3, and where* $\mathcal{R}_{\boldsymbol{\xi}_2} = \mathrm{E}\left[\boldsymbol{\xi}_{2,t}\boldsymbol{\xi}_{2,t}^{'}\right]$.

When $\mathbf{B}_t$ is updated using devolatilized orthogonal innovations, 2nd block covariance targeting could reveal impractical in most cases. This is because $\mathcal{R}_{\boldsymbol{\xi}_2}$ ultimately depends on the unobserved $\mathbf{v}_{2,t}$. In fact, by the law of iterated expectations $\mathcal{R}_{\boldsymbol{\xi}_2} = \mathrm{E}\left[(\mathbf{I} \odot \mathcal{S}_{22,\mathrm{t}})^{-1/2}\mathcal{S}_{22,\mathrm{t}}\,(\mathbf{I} \odot \mathcal{S}_{22,\mathrm{t}})^{-1/2}\right]$, which makes the 2nd block covariance targeting the (non trivial) solution of $\mathrm{E}\left[\mathcal{S}_{22,\mathrm{t}}\right] + \mathrm{E}[(\mathbf{I} \odot \mathcal{S}_{22,\mathrm{t}})^{-1/2}\mathcal{S}_{22,\mathrm{t}}(\mathbf{I} \odot \mathcal{S}_{22,\mathrm{t}})^{-1/2}] \odot \sum_{\mathrm{s}=0}^{\infty}\mathcal{A}_{\mathrm{s}} = \boldsymbol{\Sigma}_{22} - \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}^{'}$.

Remark 5 *Because* $\mathcal{R}_{\boldsymbol{\xi}_2}$ *is unit diagonal by construction, 2nd block covariance targeting becomes straightforward under marginalization, i.e. using any of the factorization* $\mathcal{M}_3$, $\mathcal{M}_4$ *and* $\mathcal{M}_5$*. In this case Theorem 5 (c) simplifies to* $\mathrm{E}\left[\mathrm{diag}(\mathcal{S}_{22,t})\right] = \mathrm{diag}\left(\boldsymbol{\Sigma}_{22} - \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}^{'} - \sum_{s=0}^{\infty}\mathcal{A}_s\right)$*, where all quantities on the right hand side can be estimated using observable data.*

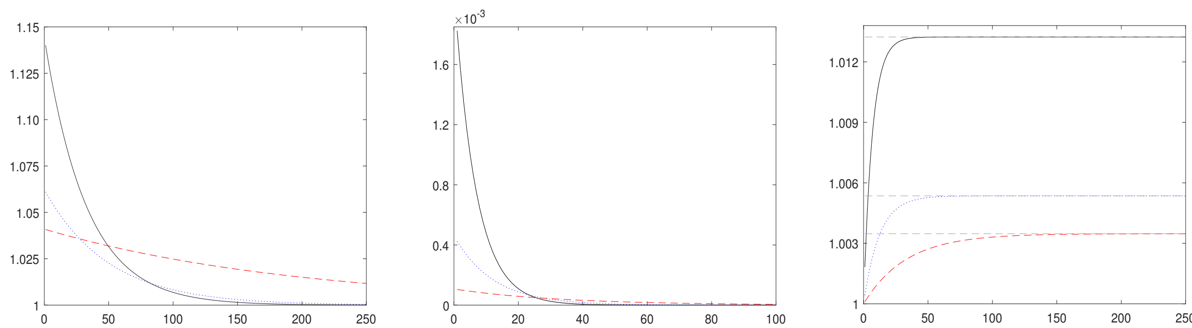
5 Monte Carlo study

In this Section, we perform three numerical exercises. The first illustrates Theorem 3 with a numerical example based Corollary 2. We assess the impact of truncation in the 2nd block covariance targeting estimator in a controlled setting where the closed-form solution of the infinite sum in (5) is known. The second quantifies the efficiency loss entailed by beta and 2nd block covariance targeting compared to the full QML estimator. The aim of the last simulation exercise is twofold: *i)* compare the finite sample properties of the multi-step estimation methods discussed in Section 4.4 to the full QML estimator and, *ii)* investigate the behavior of the QML estimator when the beta dynamics are driven by devolatilized orthogonal shocks, defined in Section 4.5, coupled with beta targeting.

5.1 Truncation in the 2nd block covariance targeting estimator

We consider the model (3)-(4.iii) with $k = 1$, $n = 2$ and with $\boldsymbol{\eta}_t \sim$ i.i.d $\mathrm{N}(\mathbf{0}, \mathbf{I})$. The conditional covariances $\mathcal{S}_{ii,t}$, $i = 1,2$, follow a GARCH(1,1) and a scalar BEKK(1,1), centered respectively in $\mathcal{S}_{11} = 1$ and $\mathrm{vech}(\mathcal{S}_{22}) = (1, 0.5, 1)^{'}$. The (2×1) process $\mathbf{B}_t$ is centered in $\mathbf{B} = (0.5, 0.5)^{'}$. For simplicity, the pairs innovation/smoothing parameters in $\mathcal{S}_{ii,t}$, namely (τ_i, δ_i), $i = 1,2$, and in $\mathbf{B}_t$, i.e. (ω_j, γ_j), $j \in [1,n]$, are restricted to be the same in all equations. We consider three sets of parameters, reflecting increasing persistence and smoothness: $(0.04, 0.93)$, $(0.02, 0.96)$ and $(0.01, 0.985)$.

Figure 1 shows $\mathbf{R}_s$, $\mathcal{A}_s$ and the truncated sum $1 + \sum_{s=0}^{S} \mathcal{A}_s$, defined in Theorem 3.[5] The higher the persistence in $\mathcal{S}_{11,t}$, the slower $\mathbf{R}_s$ converges to $\Sigma_{11}^{\odot 2}$ (panel (a)). This is because, according to (6), the decay factor of $\mathbf{R}_s = \mathrm{E}[v_{1,t}^2 v_{1,t-s-1}^2]$, namely $1 - (\tau_1 + \delta_1)$, shrinks with the persistence of $\Sigma_{11,t}$. The sequence $\mathcal{A}_s$ decays towards zero faster than $\mathbf{R}_s$ because of the compounded effect of $\omega_j^2 \gamma_j^{2s}$ (panel (b)). The truncated sum approaches the asymptotic limit $\mathcal{A}$, which for this process is explicitly given in (7), reasonably fast for all parameterisations (panel (c)). Furthermore, as the degree of smoothness of B_t increases (i.e. $(\omega_j/\gamma_j) \to 0$) we observe that, although the convergence of the truncated sums gets slower, the scaling factor, $\mathcal{A}$ in (5), approaches the unit lower bound.[6] These results suggest that truncating $\mathcal{A}$ in (5) to computationally feasible levels entails no substantial bias.



(a) $\mathbf{R}_s$ *$s \in [1, 250]$* ***(b)*** $\mathcal{A}_s$ *$s \in [1, 100]$* ***(c)*** $1 + \sum_{s=0}^{S} \mathcal{A}_s$, *$S \in [1, 250]$*

Figure 1: *R_s, $\mathcal{A}_s$, $1 + \sum_{s=0}^{S} \mathcal{A}_s$ (truncated sum) and its theoretical limit $\mathcal{A}$, for the sets of innovation/smoothing parameters: $(0.04, 0.93)$ (solid black), $(0.02, 0.96)$ (dotted blue) and $(0.01, 0.985)$ (dashed red).*

5.2 Relative efficiency under beta and 2nd block covariance targeting

The data generating process is the same as in Section 5.1, but with cross-sectional dimension increased to five ($k = 3$ and $n = 2$). $\mathcal{S}_{ii,t}$, $i = 1, 2$, follow scalar BEKK dynamics with $\mathcal{S}_{11,[jj]} = 1$ and $\mathcal{S}_{11,[jl]} = 0.1$, $j \neq l$, $j, l \in [1, k]$ for the first block, $\mathcal{S}_{22,[jj]} = 1$ and $\mathcal{S}_{22,[jl]} = 0.5$, $j \neq l$, $j, l \in [1, n]$ for the second block, respectively. Both covariance processes have innovation parameter

[5]The restriction of common parameters implies that the 2×2 symmetric matrix $\mathcal{A}_s$ (and thus its truncated sums) has a single unique value.

[6]Although to a lower extent, the reduction in $\mathcal{A}$ is furthered the smoother $\mathcal{S}_{11,t}$, i.e. $(\tau_i/\delta_i) \to 0$.

$\tau_i = 0.04$ and smoothing $\delta_i = 0.93$, $i = 1, 2$. The 2×3 matrices of parameters $\mathbf{B}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$, are set such that $\beta_{ij} = 0.5$, $\omega_{ij} = 0.04$, and $\gamma_{ij} = 0.93$, $i \in [1, n]$ and $j \in [1, k]$. Although, all betas share common dynamics, all parameters in (4.iii) are estimated individually without imposing equality restrictions. $\mathbf{B}$ is targeted using the least square estimator. For 2nd block covariance targeting, the truncation in (5) is set to 100, while $\mathbf{R}_s = \mathrm{E}\left[\left(\mathbf{v}_{1,t-s-1}\mathbf{v}'_{1,t-s-1}\right) \odot \left(\mathbf{v}_{1,t}\mathbf{v}'_{1,t}\right)\right]$ and $\boldsymbol{\Sigma}_{22} = \mathrm{E}[\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}'_{2,t}]$ are estimated by their sample counterpart. The four competing specifications (unrestricted, $\mathbf{B}$-targeting, $\mathcal{S}_{22}$-targeting and both) are estimated using $\mathcal{M}_2$. Because for this data generating process the estimation problem is separable without efficiency loss, only second block parameters are estimated.[7] We consider three sample sizes, $T = \{1500, 3000, 6000\}$, and 5000 replication. Table 1 reports bias and root mean squared errors (RMSE) averaged across the elements of $\mathbf{B}, \boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ (denoted β_{ij}, ω_{ij} and γ_{ij}) and across the diagonal elements of $\mathcal{S}_{22}$ (denoted $\mathcal{S}_{22,[ii]}$). Values reported are multiplied by 100.

Table 1: *Finite sample properties of the full QML estimator vs. beta and 2nd block covariance targeting. All reported values are multiplied by 100. The table reports BIAS and RMSE (both multiplied by 100) for the unrestricted intercepts (U/R),* $\mathbf{B}$*-targeting (*$\mathbf{B}$*),* $\mathcal{S}_{22}$*-targeting (*$\mathcal{S}_{22}$*), both (*$\mathbf{B}$*,*$\mathcal{S}_{22}$*) and, for sample size* $T = \{1500, 3000, 6000\}$*. The last row (# Par) reports the number of 2nd-block parameters estimated by QML.*

	T=1500				T=3000				T=6000			
	U/R	$\mathbf{B}$	$\mathcal{S}_{22}$	$\mathbf{B}$,$\mathcal{S}_{22}$	U/R	$\mathbf{B}$	$\mathcal{S}_{22}$	$\mathbf{B}$,$\mathcal{S}_{22}$	U/R	$\mathbf{B}$	$\mathcal{S}_{22}$	$\mathbf{B}$,$\mathcal{S}_{22}$
						BIAS						
β_{ij}	0.04	0.04	0.02	0.04	0.02	0.02	0.01	0.02	0.02	0.03	0.02	0.03
ω_{ij}	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01
γ_{ij}	-0.99	-0.95	-0.99	-0.96	-0.70	-0.68	-0.70	-0.68	-0.40	-0.38	-0.39	-0.38
$\mathcal{S}_{22,[ii]}$	-0.36	-0.36	-0.43	-0.50	-0.19	-0.18	-0.22	-0.26	-0.11	-0.11	-0.13	-0.15
$\mathcal{S}_{22,[12]}$	0.20	0.18	0.25	0.18	0.05	0.05	0.09	0.05	0.03	0.03	0.07	0.05
τ_2	0.03	0.00	0.02	-0.01	0.03	0.02	0.02	0.01	0.01	0.00	0.00	-0.00
δ_2	-0.55	-0.52	-0.52	-0.50	-0.30	-0.29	-0.28	-0.27	-0.13	-0.13	-0.13	-0.12
						RMSE						
β_{ij}	3.84	4.28	3.89	4.28	2.77	3.12	2.82	3.12	1.95	2.22	1.99	2.22
ω_{ij}	1.26	1.26	1.26	1.25	0.90	0.90	0.90	0.90	0.63	0.63	0.63	0.63
γ_{ij}	3.90	3.88	3.90	3.89	2.98	2.97	2.97	2.97	2.02	2.01	2.02	2.01
$\mathcal{S}_{22,[ii]}$	3.79	3.79	3.93	3.95	2.72	2.72	2.84	2.84	1.92	1.92	2.01	2.02
$\mathcal{S}_{22,[12]}$	5.15	5.16	5.47	5.62	3.74	3.74	3.97	4.09	2.66	2.66	2.85	2.93
τ_2	0.75	0.74	0.74	0.74	0.52	0.52	0.52	0.52	0.37	0.37	0.37	0.37
δ_2	1.71	1.69	1.70	1.69	1.13	1.12	1.12	1.12	0.75	0.75	0.75	0.75
# Par	23	17	20	14	23	17	20	14	23	17	20	14

[7] Because the second block likelihood does not depend on first block parameters, $\mathcal{M}_2$ is equivalent to $\mathcal{M}_1$.

The finite sample bias is negligible even for the shortest sample size (T =1500). The truncation level in (5) reveals sufficient, in that we do not find evidence of discernible asymptotic bias. The RMSE decays at the appropriate $T^{1/2}$ rate. The efficiency loss of the targeted parameters is invariant to the sample sizes and it amounts to about 12% for the betas, 4.5% for the 2nd block unconditional variances and 8% for the covariance. As expected, targeting does not affect the properties of the parameters estimated by QML.

5.3 Finite sample properties of full-QMLE vs. multi-step methods, when betas are driven by devolatilized shocks and beta tracking

In this Monte Carlo exercise, the data generating process is virtually unchanged from Section 5.2. The only, yet substantial, difference stands in the substitution of the innovation term in (4.iii) with the outer product $(\boldsymbol{\xi}_{2,t}\boldsymbol{\xi}^{'}_{1,t})$. The model parameters are estimated using $\mathcal{M}_1$, $\mathcal{M}_4$ and $\mathcal{M}_5$ together with beta targeting.[8] Because $\mathcal{S}_{ii,t}$, $i = 1, 2$, both follow scalar BEKK dynamics, the conditional variance of the marginal processes in $\mathcal{M}_4$ and $\mathcal{M}_5$ are univariate GARCH, either under the constraint of common dynamics within each block ($\mathcal{M}_4$), or unconstrained ($\mathcal{M}_5$). Table 2 reports average bias and RMSE (values reported are multiplied by 100). Note that, the use of $\boldsymbol{\xi}_{1,t}$ mandates the estimation of the first block parameters, which are not reported to save space. However, it is worth mentioning that $\mathcal{S}_{11,t}$ was targeted to the first block sample covariance, while (τ_1, δ_1) are estimated by QML. The number of first block parameters estimated by QML is reported in parenthesis in the last row of Table 2. For all methods biases are very small for all parameters and vanishing with the sample size. Unlike in Section 5.2, in this case any estimation method based on the likelihood factorization entails sequentiality and thus always a loss of efficiency. The loss of efficiency of $\mathcal{M}_3$ and $\mathcal{M}_4$ compared to the full QML ($\mathcal{M}_1$) is more pronounced for the parameters governing the dynamics of $\mathcal{S}_{22,t}$, less so for those governing the conditional betas.

[8] The full set of simulation results including $\mathcal{M}_2$ and $\mathcal{M}_3$ in not included to save space but it is available upon request.

Table 2: *Finite sample properties of $\mathcal{M}_1$, $\mathcal{M}_4$ and $\mathcal{M}_5$ with $\mathbf{B}_t$ driven by devolatilized shocks and beta targeting. The table reports BIAS and RMSE (both multiplied by 100) for $\mathcal{M}_1$, $\mathcal{M}_4$ and $\mathcal{M}_5$ and for sample size $T = \{1500, 3000, 6000\}$. The last row (# Par) reports the number of 2nd-block parameters estimated by QML. The number in parenthesis represents the number of first block parameters.*

	T=1500			T=3000			T=6000		
	$\mathcal{M}_1$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_1$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_1$	$\mathcal{M}_4$	$\mathcal{M}_5$
				BIAS					
β_{ij}	-0.02	-0.06	-0.06	0.01	0.01	0.00	-0.01	-0.01	-0.01
ω_{ij}	0.03	0.03	0.04	0.02	0.04	0.04	0.01	0.02	0.02
δ_i	-0.95	-0.88	-0.88	-0.68	-0.72	-0.72	-0.38	-0.43	-0.43
$\mathcal{S}_{22,[ii]}$	-0.44	-0.42	-0.41	-0.21	-0.21	-0.19	-0.11	-0.11	-0.10
$\mathcal{S}_{22,[12]}$	-0.05			-0.05			-0.03		
τ_2	0.01	0.04	0.11	0.01	0.01	0.03	0.01	0.01	0.02
δ_2	-0.55	-0.76	-1.40	-0.27	-0.32	-0.50	-0.13	-0.16	-0.24
				RMSE					
β_{ij}	4.41	4.42	4.42	3.14	3.14	3.14	2.22	2.22	2.22
ω_{ij}	1.27	1.39	1.39	0.91	1.02	1.01	0.63	0.72	0.72
δ_i	3.92	4.03	4.03	2.98	3.23	3.23	2.03	2.29	2.30
$\mathcal{S}_{22,[ii]}$	3.86	4.24	4.44	2.77	3.03	3.10	1.95	2.12	2.18
$\mathcal{S}_{22,[12]}$	5.26			3.72			2.66		
τ_2	0.75	0.97	1.38	0.52	0.67	0.91	0.37	0.46	0.63
δ_2	1.75	2.56	4.95	1.10	1.45	2.16	0.75	0.95	1.36
# Par	23 (2)	22 (2)	24 (6)	23 (2)	22 (2)	24 (6)	23 (2)	22 (2)	24 (6)

6 Empirical Application

The dataset used in the empirical application are from the well known and widely used Kenneth French's data library [9]. Our study spans the period from January 1, 1927, to October 30 2020, totaling 24682 daily observations. To our knowledge this is the first attempt at performing such a comprehensive historical analysis using this data.

We benchmark the conditional beta specification, driven only by heterogeneous idiosyncratic shocks, against the three specifications accounting for beta spillovers in the context of the Fama and French (1992) 3-factor model. The three risk factors are: market ($f_{mkt,t}$), approximated by the excess return on the portfolio formed by all US firms listed on the NYSE, AMEX, and NASDAQ, size (small-minus-big, $f_{smb,t}$) constructed as long-short portfolio of firms sorted by

[9]The dataset is freely available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The composition of the industry two portfolios, number 18 and 19 of the list of 30 industry portfolios, based on four-digit SIC code is available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/Siccodes30.zip.

size and, value (high-minus-low, $f_{hml,t}$) also constructed as self-financing portfolio of firms sorted by book-to-market ratio. Using the notation introduced in Section 3, $\boldsymbol{\epsilon}_{1,t} = (\mathbf{f}_t - \boldsymbol{\mu}_f)$, where $\mathbf{f}_t = (f_{mkt,t}, f_{smb,t}, f_{hml,t})'$ and $\boldsymbol{\mu}_f \equiv \mathrm{E}[\mathbf{f}_t]$.

For the set of investment assets, we consider the excess returns on two US industry portfolios: Coal (C) and Petroleum-Natural Gas (P), hence $\boldsymbol{\epsilon}_{2,t} = (\mathbf{r}_t - \boldsymbol{\mu}_r)$, with $\mathbf{r}_t = (r_{C,t}, r_{P,t})'$.

6.1 Beta Spillovers for Coal and Petroleum-Natural Gas industry portfolios

The two-blocks partition of (3) naturally fits the linear asset pricing model of Fama and French (1992), with the addition of allowing time variation in the factors exposure. The model in (3) with $k = 3$ and $n = 2$, becomes

$$\begin{bmatrix} \mathbf{I}_{(3)} & \mathbf{0}_{(3\times 2)} \\ -\mathbf{B}_t & \mathbf{I}_{(2)} \end{bmatrix} \begin{bmatrix} (\mathbf{f}_t - \boldsymbol{\mu}_f) \\ (\mathbf{r}_t - \boldsymbol{\mu}_r) \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{f,t} \\ \mathbf{v}_{r,t} \end{bmatrix}, \tag{11}$$

where $\mathbf{v}_{f,t} = \mathcal{S}_{f,t}^{1/2} \boldsymbol{\eta}_{f,t}$ and $\mathbf{v}_{\mathbf{r},t} = \mathcal{S}_{r,t}^{1/2} \boldsymbol{\eta}_{r,t}$. In this context, it is often of interest testing the nullity of the pricing model's intercept, namely the "unconditional" Jensen's alpha. The model with time-varying slopes in (11) implicitly entails conditionally time-varying intercepts $\boldsymbol{\alpha}_{r,t} = \boldsymbol{\mu}_r - \mathbf{B}_t \boldsymbol{\mu}_f$. However, because $\boldsymbol{\alpha}_r \equiv \mathrm{E}(\boldsymbol{\alpha}_{r,t}) = \boldsymbol{\mu}_r - \mathrm{E}(\mathbf{B}_t)\boldsymbol{\mu}_f = \boldsymbol{\mu}_r - \mathbf{B}\boldsymbol{\mu}_f$ (coinciding with the intercept of the constant parameter regression), the unconditional alpha can be explicitly added to (11), whose lower block becomes $\mathbf{r}_t = \boldsymbol{\alpha}_r + \mathbf{B}_t \mathbf{f}_t - (\mathbf{B}_t - \mathbf{B})\boldsymbol{\mu}_f + \mathbf{v}_{r,t}$.

To assess existence, type and extent of beta spillovers, we estimate (4), updated by the product of devolatilized orthogonal innovations, under the four parameter restrictions: (4.iii) (benchmark), (4.iv), (4.v) and (4.vi). To contain the number of parameters, we let the spillovers enter only via the updating term, while we restrict the contribution of the smoothing term to be idiosyncratic, with $\boldsymbol{\Gamma}$ diagonal in all specifications. For the same reason, we also target, the unconditional level of $\mathbf{B}_t$ to the sample OLS estimate in all specifications. Using Theorem 5, this corresponds to reparameterising the intercept in (4) as $\boldsymbol{\Psi} = \mathbf{B} \odot \left[\mathrm{vec}_{(k\times n)}^{-1} \left(\mathrm{diag}(\mathbf{I}_{(nk)} - \boldsymbol{\Gamma}) \right) \right]'$.

The model is completed by assuming (distinct) scalar BEKK dynamics for $\mathcal{S}_{f,t}$ and $\mathcal{S}_{r,t}$:

$$\mathcal{S}_{i,t} = \mathcal{S}_i \left(1 - \tau_i - \delta_i\right) + \tau_i \left(\mathbf{v}_{i,t-1} \mathbf{v}_{i,t-1}' \right) + \delta_i \mathcal{S}_{i,t-1} \qquad i = f, r.$$

The factors' block unconditional covariance matrix, $\mathcal{S}_f$, is targeted to the sample covariance. Positive definiteness of $\mathcal{S}_r$ is ensured via triangular decomposition, i.e. $\mathcal{S}_r = \boldsymbol{C}_r \boldsymbol{C}_r'$, where $\boldsymbol{C}_r$ is a lower triangular matrix of parameters. The model's free parameters, namely $\tau_f, \delta_f, \boldsymbol{\alpha}_r, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{C}_r, \tau_r, \delta_r$, are estimated by Gaussian QML using $\mathcal{M}_2$.

Table 3 reports the second step parameter estimates [10]. The unconditional exposure of the two portfolios to the market factor is $\beta^C_{mkt} = 1.17$ and $\beta^P_{mkt} = 0.93$. Despite being more correlated to the market than the Coal portfolio ($\text{Corr}(r_{P,t}, f_{mkt,t}) = 0.78$ against $\text{Corr}(r_{C,t}, f_{mkt,t}) = 0.59$), the Petroleum portfolio exhibits a much less erratic behavior ($\text{Var}(r_{P,t}) = 1.76$ against $\text{Var}(r_{C,t}) = 4.64$). The unconditional exposures to the size factor are $\beta^C_{smb} = 0.45$, while β^P_{smb} is negative but close very to zero, implying near unconditional orthogonality. Last, the Coal portfolio shows an exposure to the value factor more than double that of the Petroleum one ($\beta^C_{hml} = 0.51$ and $\beta^P_{hml} = 0.23$).

The three factor pricing model predicts that the risk factors are sufficient to price assets. Formally, the intercept $\boldsymbol{\alpha}_r = (\alpha^C, \alpha^P)'$ is expected to vanish. The null hypothesis is not rejected only for α^P. The rejection of the nullity of the intercept for the Coal portfolio, suggests possible misspecification of the pricing model.

We find evidence of time variation in all six conditional betas, which are 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 figures are in line with the findings of Darolles et al. (2018) for a comparable modeling approach, sampling frequency and assets class.

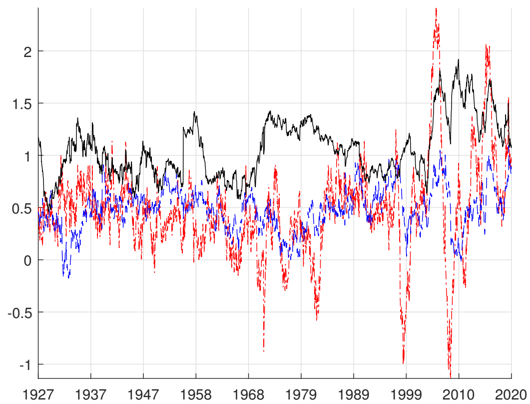
The data also shows evidence of both factor and asset spillovers. Shocks on $f_{smb,t}$ impact the exposure of both portfolios to the market factor, albeit in opposite directions. Evidence specific to each portfolio is found for the size and the value factor exposures. For the Coal portfolio, we find statistically significant asset spillovers on β^C_{hml}. For the Petroleum-Natural gas portfolio, the spillovers in the exposures to $f_{smb,t}$ and $f_{hml,t}$ involve both asset and factor spillovers. For the latter, in both cases, the source of the spillover is the market factor.

Finally, estimates of the parameters associated with the idiosyncratic shock (the diagonal elements of $\boldsymbol{\Omega}$) are close across specifications. This suggests that the inclusion of spillover effects, although not independent nor orthogonal to the idiosyncratic shock, provide novel and relevant information. This is confirmed by the likelihood ratio test which finds (4.iv) statistically superior at standard confidence levels. Figure 2 plots the filtered time series of the three risk exposures for the two

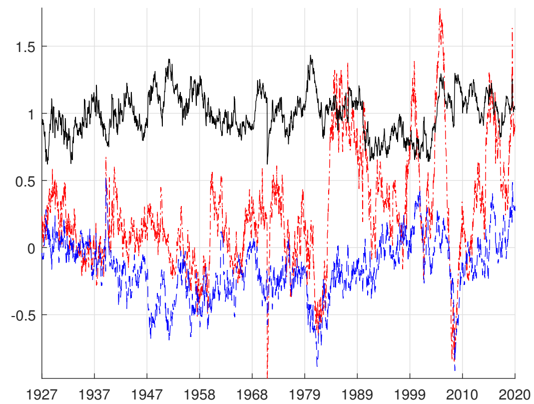
[10] Parameter estimates for the first step, common to all specifications, are $\tau_f = 0.0641$ (0.0031) and $\delta_f = 0.9327$ (0.0033), standard errors in parenthesis.

Table 3: *Estimation results of the GCAB model for the Coal and Petroleum-Natural Gas portfolios under the sub-models (4.iii), (4.iv), (4.v) and (4.iv). The table reports source of spillover (SO), parameter estimates (Est), standard error (SE) and, p-value (p-val). Inference on the elements of the* $\mathbf{B}$*, common across specifications, is based on OLS. Inference on the elements of* $\mathcal{S}_r = \boldsymbol{C}_r\boldsymbol{C}_r^{'}$ *is obtained using the delta method. Log-likelihood* ($\log \boldsymbol{\mathcal{L}}$) *reported in the last row.*

		Idiosyncratic			Assets spillovers			Factors spillovers			Factors & Assets spill.		
Par	SO	Est	SE	p-val	Est	SE	p-val	Est	SE	p-val	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}	petr				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}	petr				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}	petr				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	



(a) *Coal*



(b) *Petroleum-Natural Gas*

Figure 2: *Time-varying exposures to the risk factors:* $\beta^n_{mkt,t}$ *(solid black),* $\beta^n_{smb,t}$ *(dashed blue),* $\beta^n_{hml,t}$ *(dotted red),* $n = C, P$.

portfolios. The exposure to the size factor of the Coal portfolio, β^C_{smb}, is only sporadically negative indicating that the Coal portfolio consistently trades like a small stock. For the Petroleum portfolio, β^P_{smb} is systematically negative from the mid-40s to the mid-90s indicating that the sector in this 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 $f_{hml,t}$ shows similar peaks and troughs for both portfolios but with β^C_{hml} behaving more erratically. 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 portfolio shows a trough, trading as a growth stock, the Petroleum portfolio shows a mirroring peak, trading in the same period as a value stock.

Although, as stated in Section 4.5, there are no explicit invertibility conditions for (4) updated by the product of devolatilized orthogonal innovation, the asymptotic irrelevance of initial values can be verified empirically using the method described in Franq and Zakoian (2019), see also Winterberger (2013) and Blasques et al. (2018) for a similar approach. This consist in filtering sample paths, initialized at arbitrary starting state, using estimated parameters and model's residuals. Figure 3, plots such sample paths for $\beta^C_{mkt,t}$ (the most persistent in our set). The

Figure 3: *$\beta^C_{mkt,t}$ starting at the unconditional level* 1.1706 *(black solid) against alternative filtered sample paths initialized at arbitrary starting states $\beta^C_{mkt,0} \in [-0.5, 2.5]$.*

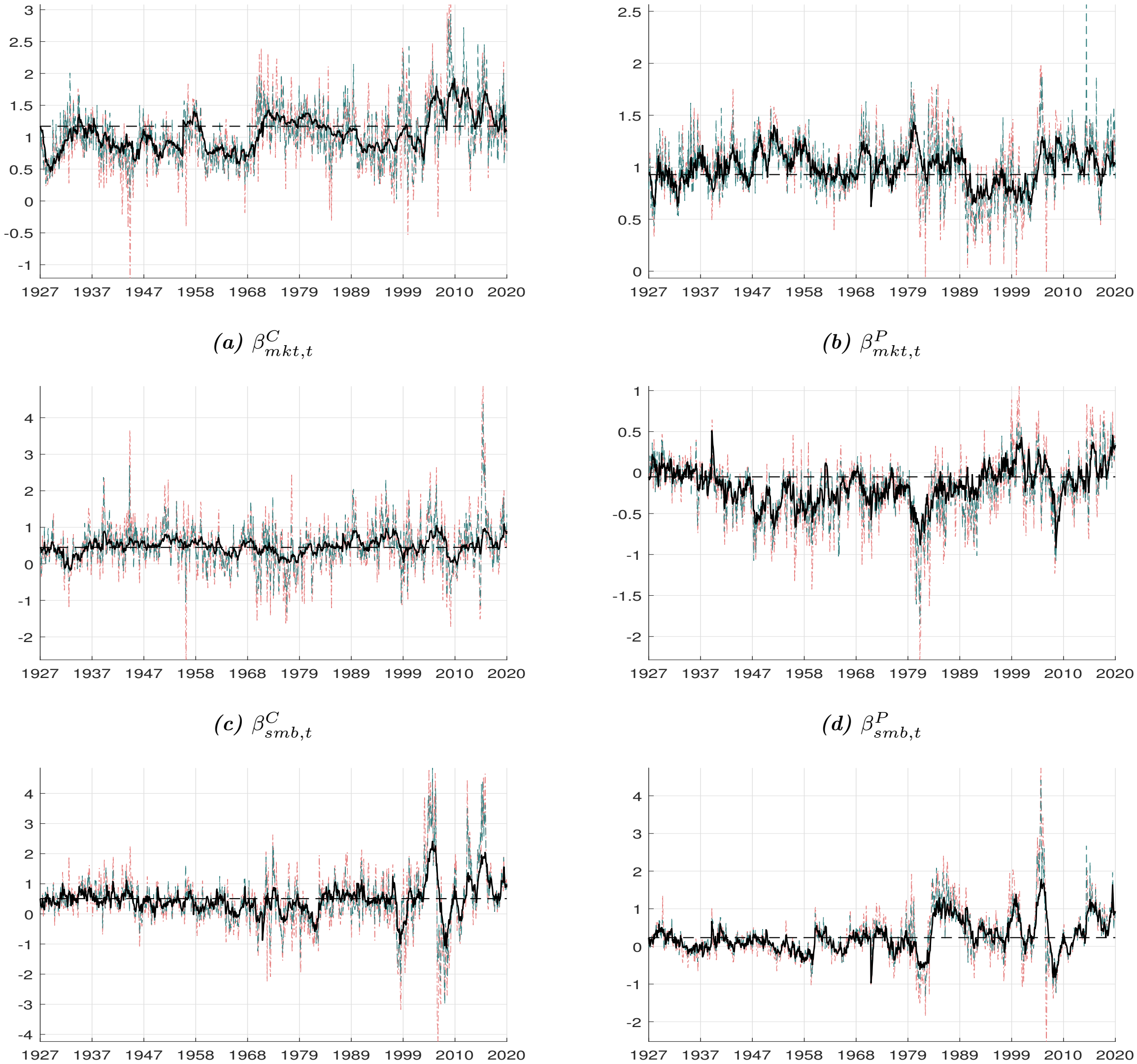
benchmark path is initialized at the unconditional beta (black solid line), while the alternatives start at $\beta^C_{mkt,0} \in [-0.5, 2.5]$. Although, because of the high persistence of the process, the impact of the initial state used to compute recursively $\mathbf{B}_t$ dissipates slowly, Figure 3 shows that $\beta^C_{mkt,0}$ has no effect asymptotically. Qualitatively similar results, not reported to save space but available upon request, are observed for the remaining elements of $\mathbf{B}_t$.

6.1.1 Comparison with the DCB-DCC (and Rolling OLS)

Figure 4 plots the six conditional betas, $\beta^i_{k,t}$, $i = C, P$, $k = \{mkt, smb, hml\}$, estimated using (4.vi), as well as those obtained using two competing approaches: direct estimation by DCB-DCC of Engle (2016) (green) and rolling OLS (red) estimated on a window of 100 observations. The latter is often used in empirical asset pricing to account for time variation in the betas.[11] The horizontal lines represent the constant parameters regression's betas estimated by full-sample OLS.

In general, the three methods track well each other. Setting aside the rolling OLS, where the degree of smoothness directly depends on the window size, the DCB-DCC yields comparatively very erratic conditional betas, characterized by short-lived bursts of unrealistic size. For example,

[11] The rolling OLS has the advantage of an intuitive interpretation and easy implementation. However, it makes difficult to distinguish between actual time variation (signal) and estimator's sampling variability (noise).



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

(c) $\beta^{C}_{smb,t}$ *(d)* $\beta^{P}_{smb,t}$

(e) $\beta^{C}_{hml,t}$ *(f)* $\beta^{P}_{hml,t}$

Figure 4: *Dynamic conditional betas: GCAB (solid black), DCB-DCC (dashed green) and 100-obs. rolling OLS (dotted red).*

$\beta^{C}_{mkt,t}$ peaks above 3.4 in January 2009, or $\beta^{C}_{smb,t}$ climbs rapidly from -1.8 to 5.4 between July and December 2015. A similar behavior is noticeable for all the other betas and through the sample. For ease of comparison, Table 4 reports sample moments of the filtered betas plotted in Figure 4. Compared to our model, the two competing approaches show variability that is twice to almost 10-fold. Our model struggles only in presence of abrupt shifts caused by episodes of sharp reversion, i.e. clusters of large unidirectional shocks, where the persistent dynamics generate a

Table 4: Summary statistics for $\mathbf{B}_t$ estimated using the GCAB, the DCB-DCC and the 100-obs. rolling OLS.

Beta	Model	Mean	Variance	Skew.	Kurt.	Min.	Max.
$\beta^C_{mkt,t}$	GCAB	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
$\beta^C_{smb,t}$	GCAB	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
$\beta^C_{hml,t}$	GCAB	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
$\beta^P_{mkt,t}$	GCAB	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
$\beta^P_{smb,t}$	GCAB	-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
$\beta^P_{hml,t}$	GCAB	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

rather slow transition while the DCB-DCC and the rolling OLS exhibit a faster adjustment. In our sample, we observe one of such instances in $\beta_{hml,t}$ during the third quarter of 2006.

In a beta hedging strategy, more volatile betas imply more re-balancing and thus much larger transaction costs, as well as potentially a more volatile beta hedged portfolio. Although limited

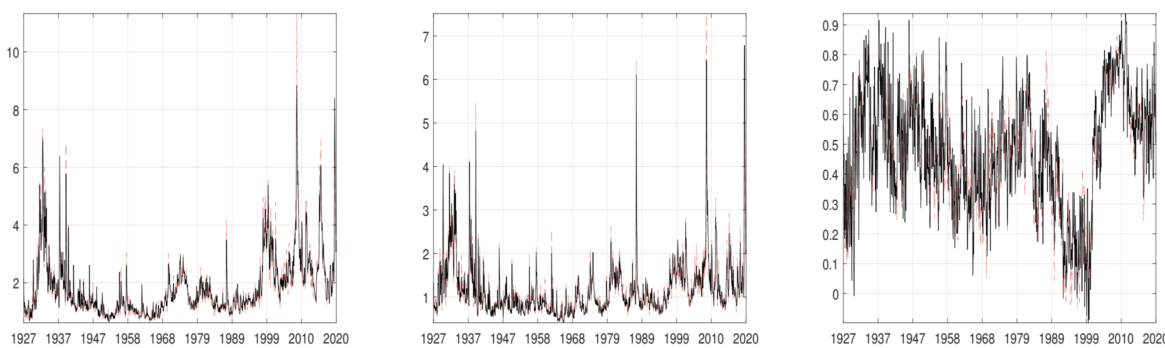
Table 5: In-sample performance of the beta hedged portfolios for the GCAB, the DCB-DCC and the 100-obs. rolling OLS

	Mean		Var		TO	
Model	Coal	Petr.	Coal	Petr.	Coal	Petr.
GCAB	-0.0127	-0.0026	2.6451	0.5484	0.0281	0.0274
DCB-DCC	-0.0127	-0.0024	2.6796	0.5648	0.1666	0.0927
roll-OLS	-0.0099	-0.0037	2.6458	0.5496	0.1056	0.0502

to an in-sample evaluation, Table 5 reports the performance of the beta hedged portfolios. The most striking figure is observed for the turnover (TO), $T^{-1}\sum_{k=1}^{3}|\beta_{ik,t}-\beta_{ik,t-1}|$, $i=C,P$, k indexing the three risk factors, which for our model is up to six times smaller than the two competing alternatives.

6.1.2 Indirect estimation of Σ_t

Endowed with $\mathbf{B}_t, \mathcal{S}_{11,t}, \mathcal{S}_{22,t}$ we can use (2) to indirectly infer the $\mathcal{I}_{t-1}$-conditional covariance matrix of the cross-section of assets $\Sigma_{\mathbf{r},t}$. The last question to be addressed is then whether the differences in the filtered betas observed between our model and the DCB-DCC (indirect estimation of $\mathbf{B}_t$ - direct estimation of $\Sigma_{\mathbf{r},t}$) translate into differences in $\Sigma_{\mathbf{r},t}$. A graphical comparison of the conditional standard deviations and the conditional correlation between our two portfolios, reported in Figures 5, shows no substantial difference between the two models which track each other extremely closely. We remark, though, that our model delivers less noisy sample paths.



(a) Standard deviation Coal *(b) Standard deviation Petroleum* *(c) Correlation Coal-Petroleum*

Figure 5: *Volatility and correlation of Coal and Petroleum portfolios estimated indirectly using DCB (solid black) and directly using DCC (dashed red).*

More precisely, we observe less extreme responses of the conditional variance to isolated large shocks, which on the contrary is the characteristic reaction of the direct variance modeling by means of GARCH-type dynamics. Summary statistics in Table 6 support these conclusions. This suggests an inherent robustness of our model to isolated extreme shocks, that is an attractive feature for risk management or portfolio allocation applications, as well as any other situation where $\mathbf{\Sigma}_{\mathbf{r},t}$ is of interest.

7 Conclusions

The paper introduced a new model, the Generalized Conditional Autoregressive Beta (GCAB), to estimate time-series regressions with time-varying coefficients and conditional heteroskedasticity. The GCAB achieves orthogonalization between two sets of variables, in the asset pricing

Table 6: Summary statistics for $\boldsymbol{\Sigma}_{r,t}$ estimated indirectly using the GCAB vs. direct estimation using the DCB-DCC model

Variable	Model	Mean	Variance	Skew.	Kurt.	Min.	Max.
$\boldsymbol{\Sigma}_{r,[11],t}$ (C)	GCAB	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
$\boldsymbol{\Sigma}_{r,[22],t}$ (P)	GCAB	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
$\boldsymbol{\Sigma}_{r,[12],t}$	GCAB	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
$\rho_{12,t}$	GCAB	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

context the risk factors and the investment assets, via a Block-Cholesky decomposition of the factors-assets system's covariance matrix. This provides several direct advantages over competing models available in the literature: the model is invariant to permutation of the element in each block, it avoids parameter proliferation, it introduces beta spillovers, it allows for different layers of likelihood factorization which eases computational feasibility in large dimension.

We derive conditions for stationarity and invertibility, as well as beta tracking and orthogonalized innovations covariance targeting. We also provide efficient computational strategies for likelihood maximization. The finite sample properties of such estimators are studied by means of an extensive Monte Carlo simulation.

The empirical application aims at comparing a baseline conditional beta specification driven only by idiosyncratic shocks against three alternative specifications accounting for different types of beta spillovers, in the context of the Fama and French (1993) three-factor framework. We consider a bivariate asset system composed of the Coal and Petroleum-Natural Gas 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. Finally, we benchmark our model against the DCB-GARCH model of Engle (2016)] and the rolling OLS, widely used in the empirical asset pricing literature. We find that our model delivers much more realistic conditional beta dynamics, compared to those unrealistically volatile delivered by the two competing models.

Appendix A. Proofs

Proof of Theorem 1. Define $\boldsymbol{\beta}_t = \text{vec}\left(\mathbf{B}_t^{'}\right)$ and $\mathbf{w}_t = \text{vec}\left(\boldsymbol{\Psi}^{'}\right) + \sum_{p=1}^{P} \boldsymbol{\Omega}_p \left(\mathbf{v}_{2,t-p} \otimes \mathbf{v}_{1,t-p}\right)$ with $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}_p$ defined according to equation (4). Conditions satisfying *C.1* are specific to the MGARCH parameterization of $\mathcal{S}_{ii,t}$, $i = 1, 2$, and examples can be found in Pedersen and Rahbek (2014), Boussama et al. (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 Prop. 3.36 in White (1984), $(\mathbf{v}_{2,t} \otimes \mathbf{v}_{1,t})$ is stationary and ergodic and, by the ergodic theorem, so is $\mathbf{w}_t$. It follows that, result (a), under condition *C.2* a stationary and ergodic solution to (4) is $\boldsymbol{\beta}_t = \boldsymbol{\Gamma}_Q(\mathrm{L})^{-1}\mathbf{w}_t = \sum_{i=0}^{\infty} \mathcal{G}_i \mathbf{w}_{t-i}$, $\mathcal{G}_0 = \mathbf{I}_{(nk\times nk)}$ and $\sum_{i=0}^{\infty} \|\mathcal{G}_i\| < \infty$, where L denotes the lag operator.

The result in (b) stems directly form orthogonality of $\mathbf{v}_{1,t}$ and $\mathbf{v}_{2,t}$:

$$\begin{aligned} \mathrm{E}[\boldsymbol{\beta}_t] &= \boldsymbol{\Gamma}_Q(1)^{-1}\text{vec}(\boldsymbol{\Psi}^{'}) + \sum_{i=0}^{\infty} \mathcal{G}_i \left(\sum_{p=1}^{P} \boldsymbol{\Omega}_p \mathrm{E}\left[\mathbf{v}_{2,t-p-i} \otimes \mathbf{v}_{1,t-p-i}\right] \right) \\ &= \boldsymbol{\Gamma}_Q(1)^{-1}\text{vec}(\boldsymbol{\Psi}^{'}), \end{aligned} \tag{12}$$

because $\mathrm{E}\left[\mathbf{v}_{2,t-p-i} \otimes \mathbf{v}_{1,t-p-i}\right] = \mathbf{0}_{(nk)}$ $\forall t$, then $\mathrm{E}[\mathbf{B}_{\mathrm{t}}] = \left[\text{vec}^{-1}_{(\mathrm{k}\times\mathrm{n})}\left(\boldsymbol{\Gamma}_{\mathrm{Q}}(1)^{-1}\ \text{vec}\left(\boldsymbol{\Psi}^{'}\right)\right)\right]^{'}$.
Ergodicity and stationarity of $\boldsymbol{\epsilon}_t$, result (c), follows. ■

Proof of Theorem 2. For ease of exposition and analytical tractability we first derive the result for baseline parameterisation (4.iii) and $P = Q = 1$. Under Theorem 1, (4.iii) admits the representation:

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

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

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

Expectations of the second term on the right hand side depend on linear combinations of $k^2 n$ 4th-order moments $\mathrm{E}[v_{1,[m],t}v_{1,[i],t}v_{1,[i],t-s}v_{2,[j],t-s}]$, $i, m = 1, \ldots, k$, $j = 1, \ldots, n$ $s \in \mathbb{N}$. Because condition *C.3* entails contemporaneous (conditional) and inter-temporal (unconditional) independence between the two blocks, all such moments are null. This result directly extends to the general specification in (4), because the second term on the right hand side of:

$$\mathrm{E}[\mathbf{B}_t \mathbf{v}_{1,t} \mathbf{v}_{1,t}^{'}] \quad = \quad \mathbf{B}\Sigma_{11} + \mathrm{E}\left[\mathrm{vec}_{(k\times n)}^{-1} \left(\sum_{i=0}^{\infty} \mathcal{G}_i \sum_{p=1}^{P} \boldsymbol{\Omega}_p \left(\mathbf{v}_{2,t-p-i} \otimes \mathbf{v}_{1,t-p-i} \right) \right)^{'} \left(\mathbf{v}_{1,t} \mathbf{v}_{1,t}^{'} \right) \right],$$

depends on linear combinations of 4th-order moments of the same form as in the previous case. Hence, $\mathrm{E}\left[\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{1,t}^{'}\right] = \mathrm{E}\left[\mathbf{B}_t \mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right] = \mathbf{B}\Sigma_{11}$ which implies $\mathbf{B} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$. By rearranging (*b*) in Theorem 1, we obtain $\boldsymbol{\Psi} = \left[\mathrm{vec}_{(k\times n)}^{-1}\left(\boldsymbol{\Gamma}_Q(1)\,\mathrm{vec}\left(\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)\right)\right]^{'}$. ■

Proof of Theorem 3. Recalling $\mathbf{B}_t = \mathbf{B} + \widetilde{\boldsymbol{\Omega}} \odot \sum_{i=0}^{\infty} \widetilde{\boldsymbol{\Gamma}}^{\odot i} \odot \left(\mathbf{v}_{2,t-i-1}\mathbf{v}_{1,t-i-1}^{'}\right)$, we can write the first term on the rhs of $\boldsymbol{\epsilon}_{2,t}\boldsymbol{\epsilon}_{2,t}^{'} = \mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\mathbf{B}_t^{'} + \mathcal{S}_{22,t}$ as:

$$\begin{aligned}\mathbf{B}_t \mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\mathbf{B}_t^{'} = \mathbf{B}\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\mathbf{B}^{'} &+ \left[\sum_{s=0}^{\infty} \mathbf{A}_s \odot \left(\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}^{'}\right)\right]\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\mathbf{B}^{'} + \\ &+ \mathbf{B}\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\left[\sum_{s=0}^{\infty}\left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{2,t-s-1}^{'}\right)\odot \mathbf{A}_s^{'}\right] + \\ &+ \left[\sum_{s=0}^{\infty} \mathbf{A}_s \odot \left(\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}^{'}\right)\right]\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\left[\sum_{s=0}^{\infty}\left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{2,t-s-1}^{'}\right)\odot \mathbf{A}_s^{'}\right],\end{aligned} \tag{13}$$

where $\mathbf{A}_s = \widetilde{\boldsymbol{\Omega}} \odot \widetilde{\boldsymbol{\Gamma}}^{\odot s}$, $s = 1,\ldots,\infty$, with typical element $\left(\widetilde{\omega}_{ij}\widetilde{\gamma}_{ij}^{s}\right)$, $i = 1,\ldots,n$ and $j = 1,\ldots,k$. Expectations of the second and third terms on the rhs of (13) are null under *C.3*. Expectations of the last term can be rewritten as:

$$\sum_{s=0}^{\infty}\sum_{r=0}^{\infty} \mathrm{E}\left[\left(\mathbf{A}_s \odot \left(\mathbf{v}_{2,t-s-1}\mathbf{v}_{1,t-s-1}^{'}\right)\right)\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\left(\left(\mathbf{v}_{1,t-r-1}\mathbf{v}_{2,t-r-1}^{'}\right)\odot \mathbf{A}_r^{'}\right)\right], \tag{14}$$

and depends on linear combinations of k^2n^2 non-zero 6th-order moments of the form:

$$\mathrm{E}\left[v_{1,[m],t}v_{1,[i],t}v_{1,[m],t-s}v_{1,[i],t-s}v_{2,[j],t-s}v_{2,[l],t-s}\right],\ i,m = 1,\ldots,k;\ j,l = 1,\ldots,n;\ s \in \mathbb{N}.$$

Expectations of the terms such that $r \neq s$ in (14) are null under *C.3*. Thus:

$$\begin{aligned}\mathrm{E}\left[\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\mathbf{B}_t^{'}\right] &= \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}^{'} + \sum_{s=0}^{\infty}\mathcal{A}_s \odot \mathrm{E}\left[\mathcal{S}_{22,t}\right] \\ &= \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} + \sum_{s=0}^{\infty}\mathcal{A}_s \odot \mathrm{E}\left[\mathcal{S}_{22,t}\right],\end{aligned}$$

where $\mathbf{B} = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$ under Theorem 2, $\mathcal{A}_s$ is a $n \times n$ matrix with typical element $\alpha_{ij,s} = \mathbf{a}_{i,s}\mathrm{E}\left[\left(\mathbf{v}_{1,t-s-1}\mathbf{v}_{1,t-s-1}^{'}\right) \odot \left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\right]\mathbf{a}_{j,s}^{'}$, $i,j = 1,\ldots,n$, and $\mathbf{a}_{i,s}$ is the i-th row of the matrix of parameters $\mathbf{A}_s$. Hence,

$$\mathbf{\Sigma}_{22} = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} + \sum_{s=0}^{\infty}\mathcal{A}_s \odot \mathrm{E}\left[\mathcal{S}_{22,t}\right] + \mathrm{E}\left[\mathcal{S}_{22,t}\right].$$

The term $\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$, being the Schur complement of $\mathbf{\Sigma}_{22}$, is finite, symmetric and positive definite and so is $\mathrm{E}[\mathcal{S}_{22,t}]$. Therefore, under Theorem 1, $\left(\mathbf{e}_{(n)}\mathbf{e}_{(n)}^{'} + \sum_{s=0}^{\infty}\mathcal{A}_s\right)$ is finite and symmetric and, $\mathrm{E}[\mathcal{S}_{22,t}] = \left(\mathbf{e}_{(n)}\mathbf{e}_{(n)}^{'} + \sum_{s=0}^{\infty}\mathcal{A}_s\right)^{\odot -1} \odot \left(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\right)$. Finally $\lim_{s\to\infty}\mathbf{R}_s = \mathbf{\Sigma}_{11}^{\odot 2}$ stems directly from stationarity of $\boldsymbol{\epsilon}_{1,t}$. ■

Proof of Corollary 2. Because $v_{1,t}$ is a GARCH(1,1) process, $v_{1,t}^2$ admits an ARMA(1,1) representation with moments:

$$\mu_4 \equiv \mathrm{E}\left[v_{1,t}^4\right] = \sigma^2\frac{1-\delta_1\left(2\tau_1+\delta_1\right)}{1-\left(\tau_1+\delta_1\right)^2} + \mu_2^2, \tag{15}$$

$$\mathbf{R}_s \equiv \mathrm{E}\left[v_{1,t}^2 v_{1,t-s}^2\right] = \left(\tau_1+\delta_1\right)^s\left(\mu_4-\mu_2^2\right) + \mu_2^2 + \left(\tau_1+\delta_1\right)^{s-1}\delta_1\sigma^2, \ \ s \geq 1, \tag{16}$$

where $\mu_2 \equiv \mathrm{E}[v_{i,t}^2] = \frac{c}{1-\tau_1-\delta_1}$ and the nuisance parameter $\sigma^2 \equiv \mathrm{E}\left[(v_{1,t}^2 - \mathcal{S}_{11,t})^2\right]$. Using (15) we can write $\sigma^2 = (\mu_4-\mu_2^2)\frac{1-(\tau_1+\delta_1)^2}{1-\delta_1(2\tau_1+\delta_1)}$, which substituted in (16) gives:

$$\mathbf{R}_s = \left(\tau_1+\delta_1\right)^s\left(\mu_4-\mu_2^2\right)\kappa + \mu_2^2, \ \ s \geq 1,$$

where $\kappa = \frac{\tau_1(1-\delta_1(\tau_1+\delta_1))}{(1-2\tau_1\delta_1-\delta_1^2)(\tau_1+\delta_1)}$. The typical ij-th element of $\mathcal{A}$ thus writes:

$$\alpha_{ij} = 1 + \sum_{s=0}^{\infty}\omega_i\omega_j(\gamma_i\gamma_j)^s\left((\tau_1+\delta_1)^{s+1}\left(\mu_4-\mu_2^2\right)\kappa + \mu_2\right).$$

From Corollary 1, stationarity of $\boldsymbol{\epsilon}_t$ entails $\max_{i\in[1,n]}(|\gamma_i|) < 1$ and $(\tau_1+\delta_1) < 1$. Then $\sum_{s=0}^{\infty}(\gamma_i\gamma_j)^s = (1-\gamma_i\gamma_j)^{-1}$ and $\sum_{s=0}^{\infty}\left(\gamma_i\gamma_j(\tau_1+\delta_1)\right)^s = (1-\gamma_i\gamma_j(\tau_1+\delta_1))^{-1}$, thus

$$\alpha_{ij} = 1 + \omega_i\omega_j\left(\frac{(\tau_1+\delta_1)(\mu_4-\mu_2^2)\kappa}{1-\gamma_i\gamma_j(\tau_1+\delta_1)} + \frac{\mu_2^2}{1-\gamma_i\gamma_j}\right). \tag{17}$$

For a conditionally gaussian GARCH(1,1) process, He and Terasvirta (1999) derive μ_4 directly in terms of the GARCH parameters (c,τ_1,δ_1) as:

$$\mu_4 = \frac{3\mu_2^2(1+\tau_1+\delta_1)}{1-\delta_1^2-3\tau_1^2-2\tau_1\delta_1}. \tag{18}$$

Substituting (18) in (17), together with $\mu_2 = \frac{c}{1-\tau_1-\delta_1}$ completes the proof.

■

Proof of Theorem 4. The process $\mathbf{B}_t(\boldsymbol{\theta};\mathbf{B}_0)$ can be expressed in vector form in terms of the observables, $(\boldsymbol{\epsilon}_{1,t},\ \boldsymbol{\epsilon}_{2,t})^{'}$, as the random coefficients recurrence

$$\boldsymbol{\beta}_t(\boldsymbol{\theta};\mathbf{B}_0) = \mathbf{w}_{t-1} + \mathbf{S}_{t-1}\boldsymbol{\beta}_{t-1}(\boldsymbol{\theta};\mathbf{B}_0), \tag{19}$$

with $\boldsymbol{\beta}_t(\boldsymbol{\theta};\mathbf{B}_0) = \text{vec}\left(\mathbf{B}_t(\boldsymbol{\theta};\mathbf{B}_0)^{'}\right)$, $\mathbf{w}_t = \text{vec}(\boldsymbol{\Psi}^{'}) + \boldsymbol{\Omega}(\boldsymbol{\epsilon}_{2,t}\otimes\boldsymbol{\epsilon}_{1,t})$, $\mathbf{S}_t = \boldsymbol{\Gamma} - \boldsymbol{\Omega}\left(\mathbf{I}_{(n)}\otimes\left(\boldsymbol{\epsilon}_{1,t}\boldsymbol{\epsilon}_{1,t}^{'}\right)\right)$. From equation (19), by recursive substitution, we obtain

$$\boldsymbol{\beta}_t(\boldsymbol{\theta},\mathbf{B}_0) = \mathbf{w}_{t-1} + \sum_{j=1}^{t-1}\left(\prod_{i=t-j}^{t-1}\mathbf{S}_i\right)\mathbf{w}_{t-j-1} + \left(\prod_{i=0}^{t-1}\mathbf{S}_i\right)\boldsymbol{\beta}_0, \tag{20}$$

where $\boldsymbol{\beta}_0 = \text{vec}(\mathbf{B}_0^{'})$. By Theorem 1, the sequence $(\mathbf{w}_t,\mathbf{S}_t)$ is ergodic and stationary. Recalling the conditions in Brandt (1986) for convergence of stochastic recurrences with non-i.i.d stationary coefficients, if $i)$ $\max(0,\text{E}[\log||\mathbf{w}_0||] < \infty)$ and $ii)$ $\text{E}\left[\log||\mathbf{S}_0||\right] < 0$, then $\boldsymbol{\beta}_t(\boldsymbol{\theta}) = \mathbf{w}_{t-1} + \sum_{j=1}^{\infty}\left(\prod_{i=t-j}^{t-1}\mathbf{S}_i\right)\mathbf{w}_{t-j-1}$ converges absolutely almost surely, and for any arbitrary initial random state $\mathbf{B}_0$, $\text{Prob}\left(\lim_{t\to\infty}||\boldsymbol{\beta}_t(\boldsymbol{\theta};\mathbf{B}_0) - \boldsymbol{\beta}_t(\boldsymbol{\theta})|| = 0\right) = 1$.
Using Jensen's inequality, condition $i)$ holds if $\text{E}\left[\left\|\text{vec}(\boldsymbol{\epsilon}_{2,0}\boldsymbol{\epsilon}_{1,0}^{'})\right\|^{s_0}\right] < \infty$ for some $s_0 > 0$, since $\left\|\text{vec}(\boldsymbol{\epsilon}_{2,0}\boldsymbol{\epsilon}_{1,0}^{'})\right\| = \left\|\boldsymbol{\epsilon}_{2,0}\otimes\boldsymbol{\epsilon}_{1,0}^{'}\right\|$. By similar arguments, $ii)$ holds if $\text{E}\left[\|\mathbf{S}_0\|\right] < 1$ for some sub-additive and sub-multiplicative norm. Using the spectral norm, i.e. $\varsigma_A \equiv \|\mathbf{A}\|_2 = \varsigma_{max}(\mathbf{A})$ is the largest singular value of $\mathbf{A}$, and noting that $\left\|\mathbf{I}_{(n)}\otimes\left(\boldsymbol{\epsilon}_{1,0}\boldsymbol{\epsilon}_{1,0}^{'}\right)\right\|_2 = \left(\boldsymbol{\epsilon}_{1,0}^{'}\boldsymbol{\epsilon}_{1,0}\right)$, uniform invertibility holds if $\varsigma_\Gamma + \varsigma_\Omega\sum_{i=1}^{k}\text{E}\left(\epsilon_{1,[i],0}^2\right) < 1$. ■

Proof of Theorem 5. The results in *(a)* and *(b)* stem directly from Theorem 1 and Theorem 2. For *(c)*, by the same arguments used in the proof of Theorem 3 we have

$$\begin{aligned}
\text{E}\left[\mathbf{B}_t\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\mathbf{B}_t^{'}\right] &= \mathbf{B}\text{E}[\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}]\mathbf{B}^{'} + \\
&\quad + \sum_{s=0}^{\infty}\text{E}\left[\left(\mathbf{A}_s\odot\left(\boldsymbol{\xi}_{2,t-s-1}\boldsymbol{\xi}_{1,t-s-1}^{'}\right)\right)\left(\mathbf{v}_{1,t}\mathbf{v}_{1,t}^{'}\right)\left(\left(\boldsymbol{\xi}_{1,t-r-1}\boldsymbol{\xi}_{2,t-r-1}^{'}\right)\odot\mathbf{A}_r^{'}\right)\right] \\
&= \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}^{'} + \text{E}\left[\boldsymbol{\xi}_{2,t}\boldsymbol{\xi}_{2,t}^{'}\right]\odot\sum_{s=0}^{\infty}\mathcal{A}_s,
\end{aligned}$$

where the ij-th element of $\mathcal{A}_s$ is $\alpha_{ij,s} = \mathbf{a}_{i,s}\mathrm{E}\left[(\boldsymbol{\xi}_{1,t-s-1}\boldsymbol{\xi}^{'}_{1,t-s-1}) \odot (\mathbf{v}_{1,t}\mathbf{v}^{'}_{1,t})\right]\mathbf{a}^{'}_{j,s}$, $i, j = 1, \dots, n$, and $\mathbf{a}_{i,s}$ is the i-th row of $\mathbf{A}_s = \widetilde{\boldsymbol{\Omega}} \odot \widetilde{\boldsymbol{\Gamma}}^{\odot s}$. Hence,

$$\mathrm{E}\left[\mathcal{S}_{22,t}\right] = \left(\boldsymbol{\Sigma}_{22} - \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}^{'}\right) - \mathrm{E}\left[\boldsymbol{\xi}_{2,t}\boldsymbol{\xi}^{'}_{2,t}\right] \odot \sum_{s=0}^{\infty} \mathcal{A}_s,$$

where, from *(b)*, $\mathbf{B} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$. ■

References

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

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

Blasques, F., Gorgi, P., Koopman, S. J., and Winterberger, O. (2018). Feasible Invertibility conditions and Maximum Likelihood Estimation for Observation Driven Models. *Electronic Journal od Statistics*, 12:1019–1052.

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.

Boudt, K., Galanos, A., Payseur, S., and Zivot, E. (2019). Multivariate GARCH Models for Large-Scale Applications: A Survey. In Vinod, H. D. and Rao, C. R., editors, *Conceptual Econometrics Using R*, volume 41 of *Handbook of Statistics*, pages 193–242. Elsevier.

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

Brandt, A. (1986). The Stochastic Equation $Y_{n+1} = A_n Y_n + B_n$ with Stationary Coefficients. *Advances in Applied Probability*, 18:211–220.

Darolles, S., Francq, C., and Laurent, S. (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.

Engle, R. F. (2016). Dynamic Conditional Beta. *Journal of Financial Econometrics*, 14:643–667.

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

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

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

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

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

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

Francq, C. and Zakoian, J. M. (2016). Estimating Multivariate Volatility Models Equation by Equation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78:613–635.

Franq, C. and Zakoian, J.-M. (2019). *GARCH Models*. Wiley.

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

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

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

He, C. and Terasvirta, T. (1999). Properties of Moments of a Family of GARCH Processes. *Journal of Econometrics*, 92:173–192.

Koopman, S. J., Creal, D. D., and Lucas, A. (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, 28:777–795.

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

Silvennoinen, A. and Terasvirta, T. (2009). Multivariate GARCH Models for Large-Scale Applications: A Survey. Handbook of Financial Time Series, pages 201–229. Springer.

White, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press.

Winterberger, O. (2013). Continuous Invertibility and Stable QML Estimation of the EGARCH(1,1) Model. *Scandinavian Journal of Statistics*, 40:846–867.