



**HAL**  
open science

## Autoregressive conditional betas

F. Blasques, Christian Francq, Sébastien Laurent

► **To cite this version:**

F. Blasques, Christian Francq, Sébastien Laurent. Autoregressive conditional betas. *Journal of Econometrics*, 2024, 238 (2), pp.105630. 10.1016/j.jeconom.2023.105630 . hal-04676069

**HAL Id: hal-04676069**

**<https://hal.science/hal-04676069v1>**

Submitted on 23 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Autoregressive conditional betas

F. Blasques\*, Christian Francq<sup>†</sup> and Sébastien Laurent<sup>‡</sup>

September 14, 2023

## Abstract

This paper introduces an autoregressive conditional beta (ACB) model that allows regressions with dynamic betas (or slope coefficients) and residuals with GARCH conditional volatility. The model fits in the (quasi) score-driven approach recently proposed in the literature, and it is semi-parametric in the sense that the distributions of the innovations are not necessarily specified. The time-varying betas are allowed to depend on past shocks and exogenous variables. We establish the existence of a stationary solution for the ACB model, the invertibility of the score-driven filter for the time-varying betas, and the asymptotic properties of one-step and multistep QMLEs for the new ACB model. The finite sample properties of these estimators are studied by means of an extensive Monte Carlo study. Finally, we also propose a strategy to test for the constancy of the conditional betas. In a financial application, we find evidence for time-varying conditional betas and highlight the empirical relevance of the ACB model in a portfolio and risk management empirical exercise.

**Keywords:** Score driven model, Time-varying parameters, GARCH model, Betas.  
**JEL:** C01, C22, C58.

## 1 INTRODUCTION

The introduction of time-dependent variance models has revolutionized time series applications, especially in finance, which is an area where conditional means can sometimes be assumed to be constant (and even zero), but variances cannot. Much more recently, it has been realized that the betas of regression models, such as those used in finance for hedging or portfolio selection, are varying over time. In this paper, we propose and study time-varying beta models which, like GARCH models, have updating and persistence mechanisms.

---

\*VU University Amsterdam, Tinbergen Institute, The Netherlands. E-mail: f.blasques@vu.nl. Francisco Blasques acknowledges the financial support of the Dutch Science Foundation (NWO) under grant Vidi.195.099.

<sup>†</sup>CREST, Institut Polytechnique de Paris and University of Lille E-mail: christian.francq@univ-lille3.fr

<sup>‡</sup>Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, Aix-Marseille Graduate School of Management – IAE, France. E-mail: sebastien.laurent@univ-amu.fr.

Failing to account for time-varying betas yields poor predictions and biased selection rules. An active literature has developed different models and methods to account for time-varying betas - we review this literature in detail below. However, existing approaches suffer from two main limitations: either they lack updating and persistence mechanisms or their implementation is computationally challenging. Our new model addresses these limitations. We study in details their statistical properties and show that these models can be estimated using a multistep QML estimator, making their estimation straightforward.

Consider a time series  $\{y_t\}_{t \in \mathbb{Z}}$  (e.g., financial returns) that depends linearly on a set of  $p$  exogenous variables  $x_t = (x_{1,t}, \dots, x_{p,t})'$ . Linear regression models play an important role in economics and finance. However, the assumption of constancy for the parameters can be inappropriate in a number of applications. When the slope (betas) and variance parameters of the model are time-varying, the model can be specified as follows:

$$y_t = \beta_{1,t}x_{1,t} + \dots + \beta_{p,t}x_{p,t} + v_t, \quad v_t = g_t\eta_t, \quad (1)$$

where  $g_t$  is the conditional volatility of  $y_t$  and  $\eta_t$  is an *i.i.d.* random variable with mean 0 and unit variance.

If the conditional betas are time-varying but incorrectly assumed to be constant, this can lead to inaccurate forecasts and poor decision-making. In such a case, it is therefore desirable to render the betas as time-varying parameters. To do so, several solutions have been proposed in the literature.

The simplest method to obtain time-varying estimates of the betas is to estimate the model on rolling windows, as in Fama and MacBeth (1973), in the context of the CAPM model in finance. Practitioners appreciate the fact that the estimated coefficients obtained with this method are very smooth by construction. This method gives the illusion of producing dynamic conditional betas. However, under the stationarity assumption, the ordinary least squares (OLS) estimator targets the marginal (constant) betas and the time-varying estimates are the result not only of possible variation in the conditional information, but also from statistical uncertainty due to the limited dimension of the window.

Interaction variables can also be used to introduce dynamics into betas, for instance, as in Gagliardini et al. (2016). This method has the advantage of simplicity, but it also requires the availability and selection of suitable variables to drive the entire dynamics of the betas. The main drawback of this approach is that conditional betas can only inherit memory from the selected covariates, since they do not depend on their own past values. The ACB model proposed in this paper is an attempt to overcome this problem.

Occasionally, data are available at a higher frequency than that needed for betas to be estimated (e.g., the intraday returns to compute daily betas). In such settings, Barndorff-Nielsen and Shephard (2004) introduce the concept of realized betas that builds upon realized volatility and covariance measures. Realized betas are used in Patton and Verardo (2012) and Boudt et al. (2017), among other works while Ait-Sahalia et al. (2020) study their theoretical properties in the context of a multiple linear regression model. The main drawback of this approach is that data are not always available to

compute realized betas. Furthermore, realized betas are known to be very noisy when computed during short windows of time (e.g., daily realized betas computed on 5-minute data). For instance Ait-Sahalia et al. (2020) average daily realized betas at a monthly frequency to obtain smoother trajectories of the realized betas. Note also that unlike realized betas, conditional betas obviously depend on the conditioning information that is available. In this paper, the information set contains the past values of the dependent variable and of covariates.

State space models are also commonly used to obtain time-varying betas, where the betas are either assumed to be mean reverting but most often follow random walks. As described, for instance, in Hamilton (1994) and Durbin and Koopman (2001), the estimation of state space models is accomplished using Gaussian ML and requires the application of the Kalman filter. There are, however, technical and numerical challenges with the implementation of state space models, especially when dealing with nonlinear parameter updates and/or non-Gaussian innovations. By suggesting a QMLE method and an observation-driven score-updating procedure, we achieve a substantially simpler and easier model implementation. Note also that the state space models involve two latent noises and produce stochastic time-varying betas which are not measurable with past observations, and are thus difficult to interpret and compare with the conditional betas.

Building upon the work of Bollerslev et al. (1988), Engle (2016) introduce the *dynamic conditional beta* (DCB) model. This model allows for the retrieval of the time-varying betas from an estimate of the full conditional covariance matrix  $\Sigma_t$  using a multivariate GARCH model estimated on both the endogenous variable  $y_t$  and the  $p$  explanatory variables  $x_t$  in the model. This approach is very intuitive because it relies on a conditional version of the OLS estimator, where the unconditional covariance matrix of  $x$  and the unconditional covariance matrix of  $y$  and  $x$  are replaced by their conditional counterparts extracted from  $\Sigma_t$ . However, this approach also presents some major drawbacks. First, testing and imposing the constancy of the conditional betas is sometimes impractical. Second, while it is possible to introduce exogenous variables into the model, it is impossible to identify precisely which variables influence the evolution of the different betas. Third, the estimated conditional betas are often found to be highly volatile.

Darolles et al. (2018) propose a new multivariate GARCH model, called *Cholesky-GARCH* (CHAR), where the conditional covariance matrix  $\Sigma_t$  of a  $(p + 1)$ -dimensional system is obtained from a modelization of its Cholesky decomposition. The elements of this decomposition are obtained from a system of  $p + 1$  linear models with time-varying betas, whose values depend on a constant, their lags and past shocks. It was found that the CHAR model outperforms a model with constant betas and the DCB model of Engle (2016). Grassi and Violante (2022) propose a block-CHAR model with time-varying betas between a set of regressands and a set of regressors. The drawback of the CHAR approach is, however, that it requires estimating the full multivariate system (as in the DCB model) even when there is only one equation of interest. For instance, the conditional betas of the last equation (which is similar to (1)) require the estimation of  $p$  models (one for each exogenous variable). Furthermore, Darolles et al. (2018) and Grassi and Violante (2022) do not allow exogenous variables to enter into the conditional beta

equations.

In this paper, we propose a new and flexible model, which we call the *autoregressive conditional beta* (ACB) model. In the ACB model, the conditional betas are obtained using a score-driven (SD) or quasi score-driven (QSD) updating equation featuring an additional set of  $q$  predetermined exogenous variables. The SD approach, developed by Harvey and Chakravarty (2008) and Creal et al. (2012), relies on a scaled score to drive the time variation in conditional moments. Blasques et al. (2023) propose the class of QSD models that allows for the design of updating equations which are not linked to the density of the error term. This extension is important to explain key empirical regularities observed in the data, and achieves improved empirical performance in certain applications. Unlike the CHAR, the ACB model does not require the estimation of a system of equations with time-varying betas but only the equation of interest plus  $p$  univariate GARCH models when the explanatory variables display volatility clustering. The  $p$  univariate GARCH models play an important role in obtaining smoother betas, as they regulate the step size in the updating equation for the betas. In particular, in periods of high volatility in the regressors, the betas are updated more conservatively than in periods where the regressors are less volatile. This ultimately leads to a model that performs well empirically and filtered betas that are more stable and less noisy than those obtained with a DCB model, for instance. Although the ACB model features a QSD update constructed from a Gaussian score function, the model is studied and estimated without assuming any particular distributions for the innovation processes, as allowed by the QSD methodology. This approach is therefore semi-parametric and can, for instance, accommodate fat-tailed innovations.

The remainder of the paper is structured as follows. Section 2 presents the model and its properties (i.e., the stationarity and invertibility properties). The estimation of this model is studied in Section 3. Our main theoretical results are stated in Sections 2 and 3, but their proofs are relegated to the online supplement. Section 4 studies in greater detail the finite sample properties of the multistep QMLE method of the ACB model via a Monte Carlo simulation. The empirical application is presented in Section 5. Section 6 concludes the paper. The main assumptions are gathered in the appendix. Finally, the online supplement also contains additional Monte Carlo simulation results and a second application.

## 2 ACB MODEL AND ITS PROPERTIES

### 2.1 ACB updating equations

In time series, a model often refers to a data generating process (DGP) or a filter used for estimation or prediction. Where ambiguity is possible, we will use the terminology of ACB process to refer to a time series generated by the ACB model. The ACB model is a regression model with (Q)SD time-varying betas, as in (1), where the updating equation of the conditional betas is obtained from a Gaussian score function. The model also allows for heteroskedastic errors and stochastic regressors displaying GARCH dynamics. Specifically, we assume that each of the  $p$  regressors  $x_{i,t}$  ( $i = 1, \dots, p$ ) has conditional

mean  $\mu_{i,t}$  and conditional variance  $g_{i,t}^2$ ; i.e.,  $x_{i,t} = \mu_{i,t} + g_{i,t}\varepsilon_{i,t}$ , where  $\varepsilon_{i,t}$  is an *i.i.d.* random variable with mean 0 and unit variance. The ACB model is based on the score-driven approach, which has been successfully applied in many fields. More than 280 papers are referenced on the website [gasmodel.com](http://gasmodel.com); these models have applications in financial econometrics (credit risk modeling, stock volatility and correlation modeling, modeling time-varying dependence structures, CDS spread modeling, systemic risk, and high-frequency data), macroeconomics, and public health, among others.

Assume that  $y_t$  follows a conditional density  $p(y_t|f_t, \Omega_t, \theta) \forall t \in \mathbb{Z}$ , where  $f_t$  is a time-varying parameter of interest and  $\Omega_t$  is the information set available at time  $t$ . A QSD model for  $f_t$  features an autoregressive updating equation of the following form:

$$f_{t+1} = \varpi + \xi S(f_t) s_t(f_t) + c f_t, \quad (2)$$

where  $\varpi, \xi$  and  $c$  are unknown parameters to be estimated,  $S(f_t)$  is a scaling function, and  $s_t(f_t)$  is a partial derivative of a loss function of interest with respect to  $f_t$ . In the special case where the loss function corresponds the conditional log-likelihood of  $y_t$  given its past, we obtain a SD model, where  $s_t(f_t) = \frac{\partial \log p(y_t|f_t, \Omega_t, \theta)}{\partial f_t}$  is the score of the predictive likelihood with respect to  $f_t$ .

The SD literature does not prescribe a specific scaling function  $S(f_t)$ . While Blasques et al. (2015) note that the scaling function should be bounded away from zero  $S(f) > 0 \forall f$  for their optimality results to hold, those results are only applicable for univariate filters and do not extend to the current multivariate setting. In practice, the SD model literature is rich in scaling matrices. Examples include unit/identity scaling (Blasques et al., 2015); inverse information scaling (Creal et al., 2010); square root inverse information (Creal et al., 2011); Moore-Penrose pseudo-inverse (Bazzi et al., 2017). Similarly, the QSD model approach allows also for flexibility in the choice of the scaling matrix.

For the ACB model, we specify an updating equation for the time-varying betas which corresponds to a SD update when the regression error  $\eta_t$  is *i.i.d.*  $\mathcal{N}(0, 1)$ . Further, the scaling of the score corresponds to a ‘diagonalized’ inverse information, i.e., one which holds for a regression with independent regressors. Alternatively, the ACB can be seen as a QSD model where the update is derived using a Gaussian log-likelihood as a loss function, regardless of the distribution of the regression error  $\eta_t$ .

In the context of the linear regression model in (1), assuming that  $\eta_t$  *i.i.d.*  $\mathcal{N}(0, 1)$ , the log-likelihood contribution at time  $t$  is given by

$$l_t = -\frac{1}{2} \left\{ \frac{v_t^2}{g_t^2} + \log(g_t^2) \right\}, \quad v_t = y_t - \sum_{i=1}^p \beta_{i,t} x_{i,t}. \quad (3)$$

This implies that  $\frac{\partial l_t}{\partial \beta_{i,t}} = \frac{v_t x_{i,t}}{g_t^2}$  while the inverse of the information matrix for orthogonal regressors is

$$S(\beta_{i,t}) = - \left( E_{t-1} \frac{\partial^2 l_t}{\partial^2 \beta_{i,t}} \right)^{-1} = \frac{g_t^2}{\mu_{i,t}^2 + g_{i,t}^2}.$$

Other choices for the scaling function are possible, like the identity or the inverse-square-root Fisher, but we found these specifications to be inferior than the one based on the

inverse Fisher (i.e., much lower log-likelihoods) when applied to the data used in the empirical application.

The ACB model uses an update for the betas of the form  $S(\beta_{i,t}) \frac{\partial l_t}{\partial \beta_{i,t}} = \frac{v_t x_{i,t}}{\mu_{i,t}^2 + g_{i,t}^2}$ . We note that, in line with the QSD modeling approach, the adoption of this scaling matrix does not imply any statement on the independence of the regressors, but rather, we take a simple and convenient scaling function which delivers an update which works well empirically.

An extended QSD updating equation for the  $i$ -th ( $i = 1, \dots, p$ ) time-varying beta in (1), featuring  $q$  additional exogenous variables  $z_{1,t}, \dots, z_{q,t}$ , is thus given by

$$\beta_{i,t+1} = \varpi_i + \xi_i \frac{v_t x_{i,t}}{\mu_{i,t}^2 + g_{i,t}^2} + c_i \beta_{i,t} + \gamma_{1,i} z_{1,t} + \dots + \gamma_{q,i} z_{q,t}. \quad (4)$$

This updating equation is intuitive, as the term  $v_t x_{i,t}$  implies that the QSD update attempts to obtain values for the beta parameter for which the residual term  $v_t$  becomes orthogonal to the regressor  $x_{i,t}$  so that  $x_{i,t}$  is not only unconditionally uncorrelated with  $v_t$  but also conditionally uncorrelated.<sup>1</sup> If, for instance,  $\beta_{i,t}$  is undervalued, then

$$v_t x_{i,t} = y_t x_{i,t} - \sum_{j \neq i} \beta_{j,t} x_{i,t} x_{j,t} - \beta_{i,t} x_{i,t}^2 > 0.$$

Therefore, if  $\beta_{i,t}$  is too small and  $\xi_i > 0$ , then  $\beta_{i,t+1}$  tends to increase.<sup>2</sup> The model thus allows dynamic monitoring of the orthogonality condition between the error term of the regression and the regressors. Additionally, by scaling  $v_t x_{i,t}$  by the time-varying quantity  $(\mu_{i,t}^2 + g_{i,t}^2)^{-1}$ , the score-driven update ensures that the updating step size for the beta parameter is smaller when the regressor  $x_{i,t}$  is more volatile (i.e., when it features high conditional volatility  $g_{i,t}^2$ ). This makes the filtered beta less noisy and, in some sense, more robust to clusters of volatility in the regressors.

Naturally, when a regressor  $x_{i,t}$  is *i.i.d.*, then a time series model does not need to be fitted to filter the conditional mean and volatility of  $x_{i,t}$ , since we have that  $\mu_{i,t+1} = \mu_i$  and  $g_{i,t+1}^2 = g_i^2$ . For such regressors, the scaling term  $\mu_i^2 + g_i^2$  can also be removed, as the parameter  $\xi_i$  can be estimated accordingly. For this reason, it may be important to test for the presence of serial correlation and GARCH effects in the regressors when specifying the ACB model. In the case of an *i.i.d.*  $i$ -th regressor, the updating equation for the corresponding time-varying beta is thus given by

$$\beta_{i,t+1} = \varpi_i + \xi_i v_t x_{i,t} + c_i \beta_{i,t} + \gamma_{1,i} z_{1,t} + \dots + \gamma_{q,i} z_{q,t},$$

where in this case  $\xi_i$  absorbs the effect of  $(\mu_i^2 + g_i^2)^{-1}$ . Given that we allow for multiple regressors, the ACB model takes the form of a multivariate (Q)SD updating equation and is closely related to a SD update (albeit with added regressors  $z_{1,t}, \dots, z_{q,t}$ ). A

<sup>1</sup>Lange et al. (2022) propose an observation-driven modelling framework that permits time variation in the model's parameters using a proximal-parameter (ProPar) update. Interestingly, they applied this framework to the modelling of a single conditional beta and also found that  $v_t x_{i,t}$  is a key variable driving the dynamic of the conditional beta.

<sup>2</sup>Even if  $\xi_i > 0$  is natural, we will not impose any sign constraint on the coefficients.

multivariate version of the optimality results in Blasques, Koopman, Lucas (2015, 2018) is missing in the (Q)SD literature. As such, existing results are not directly applicable to the current model. We further note that one faces substantial difficulties in empirically verifying the sufficient conditions required for optimality in Blasques, Koopman, Lucas (2015, 2018). As a result, the score update cannot be theoretically justified by existing optimality results and the proposed dynamic model should be judged empirically on its own merits. As we shall see, the current ACB model joins a wide range of multivariate SD filters which have nonetheless been proposed in the literature and used successfully in applications.

We further note that the ACB model also allows for the addition of a time-varying intercept in the regression equation. In particular, this happens when a regressor  $x_{i,t}$  is constant and more particularly when  $x_{1,t} = 1 \forall t$ . In such a setting, the time-varying beta  $\beta_{1,t}$  plays the role of a time-varying intercept. In this case, the resulting updating equation is given by

$$\beta_{1,t+1} = \varpi_1 + \xi_1 v_t + c_1 \beta_{1,t} + \gamma_{1,1} z_{1,t} + \dots + \gamma_{q,1} z_{q,t}.$$

Note that the updating mechanism of the previous equation is very simple because  $\mu_1^2 = 1$  and  $g_1^2 = 0$ . Example 1 below shows that ARMA(1,1) belongs to the family of ACB models.

**EXAMPLE 1 (ARMA(1,1)).** *Assume that  $p = 1$ ,  $q = 0$  and  $x_{1t} = 1 \forall t$ . Model (1) with conditional betas that follow (4) can be rewritten as  $y_t = \varpi_1 + (\xi_1 - c_1)v_{t-1} + c_1 y_{t-1} + v_t$  and therefore is an ARMA(1,1) model with GARCH effects.*

## 2.2 Stationarity and invertibility of the ACB model

In this section, we establish the stationarity of data generated by the ACB model and the invertibility of the filters for both the time-varying conditional betas and the conditional volatilities.

It is important to highlight that the proposed model specification, including the Gaussian score, and the scaling by a ‘diagonalized’ inverse information matrix, was carefully selected so that one can show that the filtering theory holds. Specifically, the results that follow are achievable in part because, as we shall see, the specification is such that the filtering equations are related to each other according to a specific order, meaning that they can be analyzed sequentially. In particular, we note that (i) the conditional volatilities of the regressors do not depend on other filtered parameters, (ii) the conditional betas depend on the filtered conditional volatilities of their regressor only, and (iii) the conditional volatility of the observation equation’s error depends on all the other filters.

Let  $\mathbf{z}_t = (z_{1,t}, \dots, z_{q,t})^\top$  be the vector of covariates and  $\mathcal{F}_t$  the sigma field generated by  $\{\eta_u, \epsilon_u^\top, \mathbf{z}_u^\top, u \leq t\}$ . For simplicity, we assume a constant conditional mean and a



GARCH(1,1) volatility for the regressors, and consider the DGP:

$$y_t = \beta_{1,t}x_{1,t} + \cdots + \beta_{p,t}x_{p,t} + v_t, \quad v_t = g_t\eta_t, \quad (5)$$

$$\beta_{i,t+1} = \varpi_{0i} + \xi_{0i} \frac{v_t x_{i,t}}{\mu_{0i}^2 + g_{i,t}^2} + c_{0i}\beta_{i,t} + \gamma_{01,i}z_{1,t} + \cdots + \gamma_{0q,i}z_{q,t}, \quad (6)$$

$$g_{t+1}^2 = \omega_0 + \alpha_0 v_t^2 + \beta_0 g_t^2, \quad (7)$$

$$x_{i,t} = \mu_{0i} + g_{i,t}\varepsilon_{i,t}, \quad (8)$$

$$g_{i,t+1}^2 = \omega_{0i} + \alpha_{0i} (x_{i,t} - \mu_{0i})^2 + \beta_{0i} g_{i,t}^2 \quad (9)$$

with  $\varpi_{0i}, \xi_{0i}, c_{0i}, \gamma_{0j,i}, \mu_{0i} \in \mathbb{R}$  for  $i \in \{1, \dots, p\}$  and  $j \in \{1, \dots, q\}$ ,  $\omega_0 > 0$ ,  $\alpha_0 > 0$ ,  $\beta_0 \geq 0$ ,  $\omega_{0i} > 0$ ,  $\alpha_{0i} > 0$  and  $\beta_{0i} \geq 0$  for  $i \in \{1, \dots, p\}$ . The following proposition provides conditions for the existence of a stationary solution to the time series process defined by (5)-(9).

**PROPOSITION 1** (Stationarity of the ACB process). *Under **A1-A3** in the appendix, there exists a unique stationary, ergodic and  $\mathcal{F}_t$ -measurable sequence  $(y_t, \beta_{1,t+1}, \dots, \beta_{p,t+1}, g_{t+1}^2)$  satisfying (5)-(7). Moreover, the components of the previous vector admit small-order moments.*

The previous proposition illustrates an advantage of the ACB approach compared to the DCB of Engle (2016): the stationarity conditions are obvious. Indeed, under standard constraints on the parameters of the univariate volatilities of the regressors and of the errors term of the regression, the vector  $(v_t, x_{1,t}, \dots, x_{p,t})$  is stationary and ergodic. When  $|c_{0i}| < 1$ ,  $\beta_{i,t}$  inherits these properties, regardless of the values of the other parameters. This contrasts with the DCC process (often used in the DCB approach) whose stationarity—and more specifically ergodicity—are notoriously difficult to study.

In the rest of the article, we assume that the DGP of  $\{y_t\}_{t \in \mathbb{Z}}$  and  $\{x_{i,t}\}_{t \in \mathbb{Z}}$  for  $i = 1, \dots, p$  is given by (5)-(9). Obviously, the theoretical results of this article no longer hold if the DGP is not the one assumed. However, we have performed numerical experiments showing that the results are not radically changed if, for example, the volatility of the regressors is not a standard GARCH(1,1), but follows a more complicated specification. Let  $\varphi_0$  be the vector of the true value of all the parameters of the DGP and  $\Theta_\varphi$  be the parameter space, which extends the positivity constraints of  $\varphi_0$  to all its elements. Under the assumptions of Proposition 1, let  $(y_t, \mathbf{x}_t^\top, \mathbf{z}_t^\top)$  be a stationary, ergodic and  $\mathcal{F}_t$ -measurable solution to (5)-(9). The sigma field generated by the past observations  $\{y_u, \mathbf{x}_u^\top, \mathbf{z}_u^\top, u \leq t\}$  is denoted by  $\mathcal{F}_t^*$ . For  $\varphi \in \Theta_\varphi$ , we now study the existence of the  $\mathcal{F}_t^*$ -measurable solutions

$$\mathbf{Y}_t(\varphi) := (g_{1,t+1}^2(\varphi), \dots, g_{p,t+1}^2(\varphi), g_{t+1}^2(\varphi), \beta_{1,t+1}(\varphi), \dots, \beta_{p,t+1}(\varphi), v_t(\varphi))$$

to the filter

$$v_t(\boldsymbol{\varphi}) = y_t - \beta_{1,t}(\boldsymbol{\varphi})x_{1,t} - \cdots - \beta_{p,t}(\boldsymbol{\varphi})x_{p,t}, \quad (10)$$

$$\beta_{i,t+1}(\boldsymbol{\varphi}) = \varpi_i + \xi_i \frac{v_t(\boldsymbol{\varphi})x_{i,t}}{\mu_i^2 + g_{i,t}^2(\boldsymbol{\varphi})} + c_i \beta_{i,t}(\boldsymbol{\varphi}) + \gamma_{1,i} z_{1,t} + \cdots + \gamma_{q,i} z_{q,t}, \quad (11)$$

$$g_{t+1}^2(\boldsymbol{\varphi}) = \omega + \alpha v_t^2(\boldsymbol{\varphi}) + \beta g_t^2(\boldsymbol{\varphi}), \quad (12)$$

$$g_{i,t+1}^2(\boldsymbol{\varphi}) = \omega_i + \alpha_i (x_{i,t} - \mu_i)^2 + \beta_i g_{i,t}^2(\boldsymbol{\varphi}). \quad (13)$$

Note that while Equations (5)-(9) seem deceptively similar to those in (10)-(13), they are crucially different in key aspects. First, Equations (5) and (7) make statements about how the data  $y_t$  and  $x_t$  are generated, and take  $\varepsilon_{i,t}$  and  $v_t = g_t \eta_t$  to be innovations which bring randomness to the DGP. These equations are thus different from the filtering updates in (10)-(13). Second, the filtering equations do not rely on the unobserved  $\varepsilon_{i,t}$  and  $v_t$ . Instead, they use observed data only, i.e., empirical counterparts  $(x_{i,t} - \mu_i)$  and  $v_t(\boldsymbol{\varphi})$  defined in (11), which are crucially different in nature. For example,  $v_t(\boldsymbol{\varphi})$  will generally not be an uncorrelated innovation for  $\boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$ . This may lead to different dynamics which are the focus of Proposition 2. Further, even at  $\boldsymbol{\varphi}_0$ , the initialized filter will be such that even  $\tilde{v}_t(\boldsymbol{\varphi}_0) \neq v_t$  a.s. at any  $t$ . This is the focus of Proposition 3.

Let  $\mathbf{g}_t^2(\boldsymbol{\varphi}) = (g_{1,t}^2(\boldsymbol{\varphi}), \dots, g_{p,t}^2(\boldsymbol{\varphi}))^\top$ . Note that if it exists, the time-varying beta parameter  $\boldsymbol{\beta}_t = (\beta_{1,t}(\boldsymbol{\varphi}), \dots, \beta_{p,t}(\boldsymbol{\varphi}))^\top$  satisfies a stochastic recurrence equation (SRE) of the form

$$\boldsymbol{\beta}_{t+1} = \psi_{\boldsymbol{\varphi}}(y_t, \mathbf{x}_t, \mathbf{z}_t, \mathbf{g}_t^2(\boldsymbol{\varphi}); \boldsymbol{\beta}_t) = \psi_t(\boldsymbol{\varphi}; \boldsymbol{\beta}_t) \quad (14)$$

for some measurable function  $\psi_{\boldsymbol{\varphi}} : \mathbb{R}^{p+q+1} \times (0, \infty)^p \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ . Applying the general theory of the SREs developed in Bougerol (1993) and Straumann and Mikosch (2006), we obtain the following result.

**PROPOSITION 2 (Stability of the ACB filter).** *Let  $\{y_t, \mathbf{x}_t^\top, \mathbf{z}_t^\top\}_{t \in \mathbb{Z}}$  be stationary and ergodic with a finite small-order moment, and suppose that **A4** holds. Then, for all  $\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}$ , there exists a unique, strictly stationary, ergodic and  $\mathcal{F}_t^*$ -measurable solution  $\{\mathbf{Y}_t(\boldsymbol{\varphi})\}_{t \in \mathbb{Z}}$  to (10)-(13).*

The results in Proposition 2 assume that the regressors are stationary and ergodic. A high-level assumption of this nature is broadly in line with standards of the regression and SD literature. In practice, one may also proceed by assuming a specific dynamic DGP for the regressors which leads to primitive testable conditions.

In practice, the vector  $\mathbf{Y}_t(\boldsymbol{\varphi})$  defined in Proposition 2 must be approximated by a measurable function of the available past observations  $\{y_i, \mathbf{x}_i^\top, \mathbf{z}_i^\top\}$  for  $i = 1, \dots, t$ . To achieve this, let us choose starting values  $\tilde{g}_{i,1}^2(\boldsymbol{\varphi}) = \tilde{g}^2 \geq 0$  and

$$\tilde{g}_{i,t+1}^2(\boldsymbol{\varphi}) = \omega_i + \alpha_i (x_{i,t} - \mu_i)^2 + \beta_i \tilde{g}_{i,t}^2(\boldsymbol{\varphi}) \text{ for } t \geq 1.$$

For all starting values  $\boldsymbol{\beta}^0 \in \mathbb{R}^p$  and  $\tilde{g}^2 \in [0, \infty)$ ,  $\tilde{\boldsymbol{\beta}}_t$  is defined by  $\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}^0$  and

$$\tilde{\boldsymbol{\beta}}_{t+1} = \tilde{\psi}_t(\boldsymbol{\varphi}; \tilde{\boldsymbol{\beta}}_t) \text{ for } t \geq 1,$$

where  $\tilde{\psi}_t(\cdot)$  is obtained by replacing the  $g_{i,t}^2(\varphi)$  values with the  $\tilde{g}_{i,t}^2(\varphi)$  values in  $\psi_t(\cdot)$ . Straumann and Mikosch (2006) called this the *perturbed SRE*. Note that the sequence of stochastic recurrence equations  $\{\tilde{\psi}_t(\cdot)\}_{t \in \mathbb{N}}$  in the perturbed SRE is not stationary and ergodic. Similarly,  $\tilde{g}_t^2(\varphi)$  is defined with an arbitrary positive initial value, and  $\tilde{v}_t(\varphi) = y_t - \tilde{\beta}_t^\top \mathbf{x}_t$ . Replacing the elements of  $\mathbf{Y}_t(\varphi)$  with their “tilde” version, we obtain  $\tilde{\mathbf{Y}}_t(\varphi)$ . Proposition 3 shows that  $\|\tilde{\mathbf{Y}}_t(\varphi) - \mathbf{Y}_t(\varphi)\|$  converges exponentially fast to zero uniformly over  $\Theta_\varphi$ .

**PROPOSITION 3** (Invertibility of the ACB model). *Let the assumptions of Proposition 2 hold, and suppose that  $E \sup_{\varphi \in \Theta_\varphi} \|\mathbf{Y}_t(\varphi)\|^s < \infty$  for some  $s > 0$ . Then, there exists  $\varrho \in (0, 1)$  such that, as  $t \rightarrow \infty$ , a.s.*

$$\varrho^{-t} \sup_{\varphi \in \Theta_\varphi} \|\tilde{\beta}_t - \beta_t\| \rightarrow 0, \quad (15)$$

and

$$\varrho^{-t} \sup_{\varphi \in \Theta_\varphi} \left\{ |\tilde{g}_t^2(\varphi) - g_t^2(\varphi)| + \sum_{i=1}^p |\tilde{g}_{i,t}^2(\varphi) - g_{i,t}^2(\varphi)| + |\tilde{v}_t(\varphi) - v_t(\varphi)| \right\} \rightarrow 0. \quad (16)$$

Proposition 3 ensures that the ACB model is uniformly invertible (i.e., that the filter of the time-varying parameters converges almost surely and exponentially fast to a unique limit solution regardless of the initialization of the filter) and that this result holds uniformly over the parameter space  $\Theta_\varphi$ . As shown in Section 3, this property is essential for finding a consistent estimator of  $\varphi_0$  and for using the model in practice.

### 2.3 Empirically checking the invertibility of the ACB

Note that  $\tilde{\mathbf{Y}}_t(\varphi) = \tilde{\mathbf{Y}}_t(\varphi, y_0)$  depends on the vector of initial values  $y_0 = (\tilde{g}^2, \beta^0)$ . Since the initial values are arbitrarily fixed, to use the ACB model in practice, the dependence of  $\tilde{\mathbf{Y}}_t(\varphi, y_0)$  on  $y_0$  must vanish for large  $t$ . Note that such invertibility issues generally hold for time series models based on nonlinear SRE. In particular, the invertibility of the EGARCH has been studied by Wintenberger (2013) and Blasques et al. (2018a) studied the invertibility of more general SD models. Propositions 2–3 show that the model is uniformly invertible when **A4** is satisfied. The condition  $E \log \sup_{\varphi \in \Theta_\varphi} \left\| \prod_{j=1}^k \Lambda_{t-j+1}(\varphi) \right\| < 0$ , where  $\Lambda_t(\cdot)$  is defined in **A4** and  $k \geq 1$ , is infeasible, but the expectation can be approximated by

$$\frac{1}{n} \sum_{t=k}^n \log \sup_{\varphi \in \Theta_\varphi} \left\| \prod_{j=1}^k \tilde{\Lambda}_{t-j+1}(\varphi) \right\|, \quad (17)$$

for  $k \leq n$ , where  $\tilde{\Lambda}_t(\varphi)$  is the  $p \times p$  matrix whose entry on the  $i$ -th row and  $j$ -th column is  $-\frac{\xi_i x_{i,t} x_{j,t}}{\mu_i^2 + \tilde{g}_{i,t}^2(\varphi)} + c_i 1_{\{i=j\}}$ . Note however that the evaluation of (17) is time-consuming,

since it requires numerically computing  $n$  suprema. For this reason, we focus on

$$\Delta_n(\boldsymbol{\varphi}, k) = \frac{1}{n} \sum_{t=k}^n \log \left\| \prod_{j=1}^k \tilde{\Lambda}_{t-j+1}(\boldsymbol{\varphi}) \right\|, \quad (18)$$

for some particular values of  $\boldsymbol{\varphi}$ . Obviously, if  $\lim_{n \rightarrow \infty} \Delta_n(\boldsymbol{\varphi}, k) > 0$  a.s. for any  $k \geq 1$  and any operator norm, then **A4** cannot hold true. For the numerical illustrations presented below, we used the spectral operator norm (since it tends to produce smaller values of  $\Delta_n(\boldsymbol{\varphi}, k)$  than with the operator norms induced by the  $\ell_1$  or  $\ell_\infty$  vector norms). If the empirical mean in (18) is found significantly negative (for instance by recalculating it over several sub-samples) for some  $k$ , the ACB model can be estimated (see Section 3 below) and used for forecasting when  $\Theta_\varphi$  is contained in some sufficiently small neighborhood of  $\boldsymbol{\varphi}$  (by continuity arguments). As **A4** is only a sufficient condition for invertibility, and also because it is always good to do robustness exercises, the previous analysis can be supplemented by the study of trajectories  $t \mapsto \mathbf{e}_i^\top \tilde{\mathbf{Y}}_t(\boldsymbol{\varphi}, y_0)$ , where  $\mathbf{e}_i$  is the  $i$ -th column of the identity matrix of size  $2(p+1)$ , as function of  $y_0$  (as done for instance in Figure 4.1 of Francq and Zakoian, 2019 and in Figure 4 of Darolles et al., 2018, see Figures 1, 2 and 4 below).

## 2.4 Forecasting conditional betas

In some empirical applications (e.g., that presented in Section 5), out-of-sample forecasts of the conditional betas are requested. In this section, we show how to obtain  $h$ -step-ahead forecasts of the betas, which are denoted as  $E_t(\beta_{i,t+h})$ , where  $E_t(\cdot)$  refers to the conditional expectation operator based on the information set at time  $t$ . When the  $i$ -th conditional beta follows (6), for  $h > 0$ ,  $E_t(\beta_{i,t+h})$  can be written as

$$E_t(\beta_{i,t+h}) = \varpi_{0i} + \xi_{0i} E_t \left[ \frac{v_{t+h-1} x_{i,t+h-1}}{\mu_i^2 + g_{i,t+h-1}^2} \right] + c_{0i} E_t(\beta_{i,t+h-1}) + \sum_{j=1}^q \gamma_{0j,i} E_t(z_{j,t+h-1}). \quad (19)$$

The vector  $\boldsymbol{\varphi}_0$  of the parameters is unknown, but in practice, it can be replaced by a consistent estimate. For  $h = 1$ , Equation (19) reduces to

$$E_t(\beta_{i,t+1}) = \varpi_{0i} + \xi_{0i} \frac{v_t x_{i,t}}{\mu_i^2 + g_{i,t}^2} + c_{0i} \beta_{i,t} + \sum_{j=1}^q \gamma_{0j,i} z_{j,t}, \quad (20)$$

while for  $h > 1$ , since  $E_t[v_{t+h-1} x_{i,t+h-1}] = 0$  under **A7**, we obtain

$$E_t(\beta_{i,t+h}) = \varpi_{0i} + c_{0i} E_t(\beta_{i,t+h-1}) + \sum_{j=1}^q \gamma_{0j,i} E_t(z_{j,t+h-1}). \quad (21)$$

Note that when  $q = 0$  and  $|c_{0i}| < 1$ ,  $E_t(\beta_{i,t+h}) \rightarrow \frac{\varpi_{0i}}{1-c_{0i}}$  as  $h \rightarrow \infty$ . Therefore, in the absence of covariates in the conditional betas, out-of-sample forecasts of the conditional betas converge exponentially to a constant as the horizon  $h$  tends to infinity.

### 3 FULL, ONE-STEP AND MULTISTEP QML ESTIMATORS

The ACB process contains multiple time-varying parameters, ranging from the unconditional mean and conditional volatility of the regressors to the time-varying conditional betas and conditional volatility of the error term. As such, the structure of the ACB model lends itself to an estimation by QMLE in one or more steps. Building on the stationarity and invertibility results obtained thus far, this section establishes asymptotic properties for both one-step and multistep QML estimators.

As noted in the previous section, the ACB model has been designed with a specific “sequential” dependence structure between the SREs for the filtering of time-varying parameters. This structure allowed us to establish stability and invertibility results which would otherwise be difficult to attain in a fully non-linear multivariate setting.

Naturally, the adopted model structure also leads us to consider both a multi-step QMLE and a one-step QMLE. This an element of novelty compared to the generality of the (Q)MLE results in the score literature.

#### 3.1 Multistep QMLE

Let us first consider the regressors’ conditional volatility parameter and the role it plays in the ACB model. Naturally, in the first step, for  $i \in \{1, \dots, p\}$ , the GARCH(1,1) parameters  $\boldsymbol{\theta}_0^{(i)} = (\mu_{0i}, \omega_{0i}, \alpha_{0i}, \beta_{0i})^\top$  can be estimated independently by standard QML,

$$\hat{\boldsymbol{\theta}}_n^{(i)} = \arg \min_{\boldsymbol{\tau} \in \Theta} \tilde{O}_n^{(i)}(\boldsymbol{\tau}), \quad \tilde{O}_n^{(i)}(\boldsymbol{\tau}) = \frac{1}{n} \sum_{t=2}^n \tilde{\ell}_{i,t}(\boldsymbol{\tau}), \quad (22)$$

where  $\boldsymbol{\tau} = (\mu, \omega, \alpha, \beta)^\top$  denotes a generic element of the parameter space  $\Theta$  and

$$\tilde{\ell}_{i,t}(\boldsymbol{\tau}) = \frac{(x_{i,t} - \mu)^2}{\tilde{g}_{i,t}^2(\boldsymbol{\tau})} + \log \tilde{g}_{i,t}^2(\boldsymbol{\tau}), \quad \tilde{g}_{i,t}^2(\boldsymbol{\tau}) = \omega + \alpha(x_{i,t-1} - \mu)^2 + \beta \tilde{g}_{i,t-1}^2(\boldsymbol{\tau})$$

with a given initial value  $\tilde{g}_{i,1}^2(\boldsymbol{\tau}) = \tilde{g}^2 \geq 0$ . In (22), the sum starts at  $t = 2$ , but it could start at  $t = r_0$  with any fixed  $r_0 \geq 1$  without changing the asymptotic behavior. The choice of  $r_0 > 0$  aims to reduce the influence of the starting values in the finite samples. Note that  $\tilde{g}_{i,t}^2(\boldsymbol{\tau})$  is denoted by  $\tilde{g}_{i,t}^2(\boldsymbol{\varphi})$  in Proposition 2, but we use this new notation to underline that  $\tilde{g}_{i,t}^2(\boldsymbol{\varphi})$  only depends on  $(\mu_i, \omega_i, \alpha_i, \beta_i)$  and not on the other components of  $\boldsymbol{\varphi}$ . Note also that the consistency and asymptotic normality of the first-step QMLE are well known (see for example Francq and Zakoian, 2004). The first-step estimation (the fact that the unknown parameters  $\boldsymbol{\theta}_0^{(i)}$  are replaced by their QMLE), however, influences the asymptotic distribution of the QMLE of the other parameters, obtained in a second stage.

In the second step, we estimate

$$\boldsymbol{\vartheta}_0^{(0)} = (\omega_0, \alpha_0, \beta_0)^\top \quad \text{and} \quad \boldsymbol{\vartheta}_0^{(i)} = (\varpi_{0i}, \xi_{0i}, c_{0i}, \gamma_{01,i}, \dots, \gamma_{0q,i})^\top,$$

for  $i \in \{1, \dots, p\}$ . Let  $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^{(1)\top}, \dots, \boldsymbol{\theta}_0^{(p)\top})^\top$ ,  $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_n^{(1)\top}, \dots, \widehat{\boldsymbol{\theta}}_n^{(p)\top})^\top$  and  $\boldsymbol{\vartheta}_0 = (\boldsymbol{\vartheta}_0^{(0)\top}, \boldsymbol{\vartheta}_0^{(1)\top}, \dots, \boldsymbol{\vartheta}_0^{(p)\top})^\top$ . Let  $\boldsymbol{\vartheta}$  be a generic element of some parameter space  $\Theta_\vartheta \subset (0, \infty)^2 \times [0, \infty) \times \mathbb{R}^{p(3+q)}$  and  $\boldsymbol{\varphi} = (\boldsymbol{\theta}^\top, \boldsymbol{\vartheta}^\top)^\top$ . We estimate  $\boldsymbol{\vartheta}_0$  by

$$\widehat{\boldsymbol{\vartheta}}_n = \arg \min_{\boldsymbol{\vartheta} \in \Theta_\vartheta} \widetilde{O}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\vartheta}), \quad \widetilde{O}_n(\boldsymbol{\varphi}) = \frac{1}{n} \sum_{t=2}^n \widetilde{\ell}_t(\boldsymbol{\varphi}), \quad (23)$$

where

$$\begin{aligned} \widetilde{\ell}_t(\boldsymbol{\varphi}) &= \frac{\widetilde{v}_t^2(\boldsymbol{\varphi})}{\widetilde{g}_t^2(\boldsymbol{\varphi})} + \log \widetilde{g}_t^2(\boldsymbol{\varphi}), & \widetilde{v}_t(\boldsymbol{\varphi}) &= y_t - \sum_{i=1}^p \widetilde{\beta}_{i,t}(\boldsymbol{\varphi}) x_{i,t}, \\ \widetilde{g}_t^2(\boldsymbol{\varphi}) &= \omega + \alpha \widetilde{v}_{t-1}^2(\boldsymbol{\varphi}) + \beta \widetilde{g}_{t-1}^2(\boldsymbol{\varphi}), \\ \widetilde{\beta}_{i,t}(\boldsymbol{\varphi}) &= \varpi_i + \xi_i \frac{\widetilde{v}_{t-1}(\boldsymbol{\varphi}) x_{i,t-1}}{\mu_i^2 + \widetilde{g}_{i,t-1}^2(\widehat{\boldsymbol{\theta}}_n)} + c_i \widetilde{\beta}_{i,t-1}(\boldsymbol{\varphi}) + \sum_{j=1}^q \gamma_{j,i} z_{j,t-1}; \end{aligned}$$

here, the obvious (abuse of) notation  $\widetilde{g}_{i,t-1}(\widehat{\boldsymbol{\theta}}_n)$  and additional initial values for  $\widetilde{g}_1^2(\boldsymbol{\varphi})$  and  $\widetilde{\boldsymbol{\beta}}_1 = (\widetilde{\beta}_{1,1}(\boldsymbol{\varphi}), \dots, \widetilde{\beta}_{p,1}(\boldsymbol{\varphi}))^\top$  are used. Note that by assuming that  $\widetilde{g}_1^2(\boldsymbol{\varphi}) \geq 0$  and  $\Theta_\varphi := \Theta \times \Theta_\vartheta$  is a compact set, there exists a constant  $\underline{\omega} > 0$  such that

$$\inf_{t \geq 1} \inf_{\boldsymbol{\varphi} \in \Theta_\varphi} g_t^2(\boldsymbol{\varphi}) \geq \underline{\omega} \quad \text{and} \quad \inf_{t \geq 1} \inf_{\boldsymbol{\varphi} \in \Theta_\varphi} \widetilde{g}_t^2(\boldsymbol{\varphi}) \geq \underline{\omega}. \quad (24)$$

Similarly,  $\widetilde{g}_{i,1}^2(\boldsymbol{\tau})$  is uniformly bounded away from zero. Therefore, the multistep QML estimator  $\widehat{\boldsymbol{\varphi}}_n = (\widehat{\boldsymbol{\theta}}_n^\top, \widehat{\boldsymbol{\vartheta}}_n^\top)^\top$  is well defined by (22) and (23).

### 3.2 One-step QMLE

An alternative to multistep QMLE is to estimate all the parameters in one step. The one-step QML estimator  $\widehat{\boldsymbol{\varphi}}_n^{1S}$  satisfies

$$\widehat{\boldsymbol{\varphi}}_n^{1S} = \arg \min_{\boldsymbol{\varphi} \in \Theta_\varphi} \widetilde{O}_n(\boldsymbol{\varphi}). \quad (25)$$

Note that, contrary to multistep QMLE, the one-step estimator does not exploit the likelihood of the  $x_{i,t}$  values to estimate certain components of  $\boldsymbol{\varphi}$ . With this estimator, all the parameters—even those involved in the proper dynamics of the  $x_{i,t}$  processes—are estimated from the likelihood of  $y_t$  conditional on  $\mathcal{F}_{t-1}^*$ . Hence, it is slightly surprising for a one-step estimator to be consistent. However, with the help of Lemma 1, the distribution of  $y_t$  conditional on  $\mathcal{F}_{t-1}^*$  is sufficient to identify all the components of  $\boldsymbol{\varphi}_0$ . For  $\boldsymbol{\tau} \in \Theta$ , with the change of notation just introduced, the solution of the filter (13) is  $g_{i,t}^2(\boldsymbol{\tau}) = \sum_{j \geq 0} \beta^j \{\omega + \alpha(x_{i,t-j-1} - \mu)^2 + \beta\}$ .

LEMMA 1. *Under **A1** and **A5** (a),  $g_{i,t}^2(\boldsymbol{\tau}) = a g_{i,t}^2 + b$  a.s., where  $a \neq 0$  implies  $a = 1$  and  $(\mu_i, \alpha_i, \beta_i) = (\mu_{0i}, \alpha_{0i}, \beta_{0i})$ .*

Since the one-step estimator does not integrate all the information provided by the likelihood of the  $x_{i,t}$  processes, it is intuitively very inefficient, as confirmed by the Monte Carlo experiments we performed, but not reported to save space. Clearly, we do not recommend using the one-step QMLE method in practice. The study of the asymptotic behavior of (25) is, however, simpler than that of the multistep estimator, and will constitute a first step to determine the asymptotic distribution of  $\widehat{\varphi}_n$ .

### 3.3 Full QMLE

Note that one-step estimator (25) is not a standard/full QML estimator, which requires a parameterization of the conditional correlation matrix  $\mathbf{R}_t(\varphi)$  of  $\mathbf{x}_t$  (by adding supplementary parameters in  $\varphi$ ) and is defined by

$$\widehat{\varphi}_n^F = \arg \min_{\varphi \in \Theta_\varphi} \sum_{t=2}^n \widetilde{\ell}_t(\varphi) + (\mathbf{x}_t - \boldsymbol{\mu})^\top \widetilde{\boldsymbol{\Sigma}}_t^{-1} (\mathbf{x}_t - \boldsymbol{\mu}) + \log \left| \widetilde{\boldsymbol{\Sigma}}_t^{-1} \right|, \quad (26)$$

where  $\widetilde{\boldsymbol{\Sigma}}_t = \widetilde{\mathbf{D}}_t \widetilde{\mathbf{R}}_t \widetilde{\mathbf{D}}_t$  with  $\widetilde{\mathbf{D}}_t^2 = \text{diag}(\widetilde{g}_{1,t}^2(\boldsymbol{\theta}), \dots, \widetilde{g}_{p,t}^2(\boldsymbol{\theta}))$  and other obvious notations. To derive this estimator, we write the conditional distribution of  $(y_t, \mathbf{x}_t)$  as the conditional distribution of  $y_t$  given  $\mathbf{x}_t$  multiplied by the distribution of  $\mathbf{x}_t$  given  $\mathcal{F}_{t-1}$ .

It is clear that, in terms of assumptions (the specification of  $\mathbf{R}_t$ ) and complexity (optimization on a parameter space of a larger dimension), the estimator  $\widehat{\varphi}_n^F$  is less attractive than  $\widehat{\varphi}_n^{1S}$  and is especially less attractive than the multistep estimator  $\widehat{\varphi}_n$  defined by (22) and (23). We therefore do not continue the study of (26).

### 3.4 Consistency of the one-step and multistep QML estimators

Let  $\ell_t(\varphi) = v_t^2(\varphi)/g_t^2(\varphi) + \log g_t^2(\varphi)$  and  $O_n(\varphi) = \frac{1}{n} \sum_{t=2}^n \ell_t(\varphi)$ . The following result establishes the strong consistency of both the multistep and one-step QML estimators. Specifically, we show that  $\widehat{\varphi}_n^{1S} \rightarrow \varphi_0$  a.s. and  $\widehat{\boldsymbol{\nu}}_n \rightarrow \boldsymbol{\nu}_0$  a.s., as  $n \rightarrow \infty$ . The latter result, when combined with the consistency of the first-step QML estimator for the GARCH parameters, i.e.,  $\widehat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0$  a.s. as  $n \rightarrow \infty$ , ensures the consistency of the multistep QML estimator.

**THEOREM 1** (Consistency of the two estimators). *Under Assumptions **A1**, **A4** (a) and **A5** (a),  $\widehat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0$  a.s. as  $n \rightarrow \infty$ . If in addition the assumptions of Propositions 1 and 2, **A6** (a), **A7**, **A8** and **A9** hold, and  $\xi_{0i} \neq 0$  for  $i = 1, \dots, p$ , then  $\widehat{\boldsymbol{\nu}}_n \rightarrow \boldsymbol{\nu}_0$  a.s. and  $\widehat{\varphi}_n^{1S} \rightarrow \varphi_0$  a.s.*

### 3.5 Asymptotic normality

We first need to study the derivatives of the filter in (10)-(13). It is well known that under **A4** (a), the volatilities  $g_{i,t}^2(\cdot)$  admit derivatives of any order uniformly in  $\Theta_\varphi$ . Moreover,  $E \sup_{\varphi \in \Theta_\varphi} \left\| \partial g_{i,t}^2(\varphi) / \partial \varphi \right\|^s < \infty$  for some  $s > 0$ , and there exists  $\varrho \in (0, 1)$  such that  $t \rightarrow \infty$  a.s.

$$\varrho^{-t} \sup_{\varphi \in \Theta_\varphi} \left\| \frac{\partial \widetilde{g}_{i,t}^2(\varphi)}{\partial \varphi} - \frac{\partial g_{i,t}^2(\varphi)}{\partial \varphi} \right\| \rightarrow 0. \quad (27)$$

To obtain similar results for the derivative of  $\tilde{\boldsymbol{\beta}}_{t+1}$  with respect to some component  $\varphi_k$  of  $\boldsymbol{\varphi}$  and with  $k$  varying from 1 to the total number of parameters  $d = pq + 7p + 3$ , let us consider the new SRE

$$\boldsymbol{\beta}_{t+1}^{(k)} = \mathbf{w}_t^{(k)} + \Lambda_t \boldsymbol{\beta}_t^{(k)}, \quad (28)$$

where  $\Lambda_t$  is defined in **A4** (b) and the  $i$ -th component of the vector  $\mathbf{w}_t^{(k)}$  is

$$\begin{aligned} w_{i,t}^{(k)} = & \frac{\partial}{\partial \varphi_k} \left\{ \varpi_i + \xi_i \frac{y_t x_{i,t}}{\mu_i^2 + g_{i,t}^2(\boldsymbol{\varphi})} + \gamma_{1,i} z_{1,t} + \cdots + \gamma_{q,i} z_{q,t} \right\} \\ & - \boldsymbol{\beta}_t^\top \mathbf{x}_t \frac{\partial}{\partial \varphi_k} \left\{ \frac{\xi_i}{\mu_i^2 + \beta_{i,t}(\boldsymbol{\varphi}) \frac{\partial}{\partial \varphi_k} c_i + g_{i,t}^2(\boldsymbol{\varphi})} \right\}. \end{aligned}$$

Note that  $\tilde{\boldsymbol{\beta}}_t^{(k)} := \partial \tilde{\boldsymbol{\beta}}_t / \partial \varphi_k$  satisfies the perturbed SRE  $\tilde{\boldsymbol{\beta}}_{t+1}^{(k)} = \tilde{\mathbf{w}}_t^{(k)} + \tilde{\Lambda}_t \tilde{\boldsymbol{\beta}}_t^{(k)}$ , where  $\tilde{\mathbf{w}}_t^{(k)}$  and  $\tilde{\Lambda}_t$  are obtained by replacing  $g_{i,t}^2(\boldsymbol{\varphi})$  with  $\tilde{g}_{i,t}^2(\boldsymbol{\varphi})$  in  $\mathbf{w}_t^{(k)}$  and  $\Lambda_t$ . Using the arguments of the proofs of Propositions 2 and 3, we obtain the following result.

**PROPOSITION 4** (Derivatives of the filter). *Let the assumptions of Proposition 3 hold. For all  $k \in \{1, \dots, d\}$  and all  $\boldsymbol{\varphi} \in \Theta_\varphi$ , there exists a unique stationary and ergodic process  $\{\mathbf{Y}_t^{(k)}(\boldsymbol{\varphi})\}_{t \in \mathbb{N}}$ , with*

$$\mathbf{Y}_t^{(k)}(\boldsymbol{\varphi}) := \left( g_{1,t+1}^{2,(k)}(\boldsymbol{\varphi}), \dots, g_{p,t+1}^{2,(k)}(\boldsymbol{\varphi}), g_{t+1}^{2,(k)}(\boldsymbol{\varphi}), \boldsymbol{\beta}_{t+1}^{(k)} \right) \in \mathcal{F}_t^*$$

such that the SREs (28),

$$g_{i,t+1}^{2,(k)}(\boldsymbol{\varphi}) = \frac{\partial \{\omega_i + \alpha_i (x_{i,t} - \mu_i)^2\}}{\partial \varphi_k} + g_{i,t}^2(\boldsymbol{\varphi}) \frac{\partial \beta_i}{\partial \varphi_k} + \beta_i g_{i,t}^{2,(k)}(\boldsymbol{\varphi})$$

and

$$g_{t+1}^{2,(k)}(\boldsymbol{\varphi}) = \frac{\partial \{\omega + \alpha v_t^2(\boldsymbol{\varphi})\}}{\partial \varphi_k} + g_t^2(\boldsymbol{\varphi}) \frac{\partial \beta}{\partial \varphi_k} + \beta g_t^{2,(k)}(\boldsymbol{\varphi}), \quad \frac{\partial v_t(\boldsymbol{\varphi})}{\partial \varphi_k} = -\mathbf{x}_t^\top \boldsymbol{\beta}_t^{(k)},$$

hold. Suppose that  $E \sup_{\boldsymbol{\varphi} \in \Theta_\varphi} \|\mathbf{Y}_t^{(k)}(\boldsymbol{\varphi})\|^s < \infty$  for some  $s > 0$ . Then, there exists  $\varrho \in (0, 1)$  such that, as  $t \rightarrow \infty$ , a.s.

$$\varrho^{-t} \sup_{\boldsymbol{\varphi} \in \Theta_\varphi} \left\| \tilde{\mathbf{Y}}_t^{(k)}(\boldsymbol{\varphi}) - \mathbf{Y}_t^{(k)}(\boldsymbol{\varphi}) \right\| \rightarrow 0, \quad (29)$$

where  $\tilde{\mathbf{Y}}_t^{(k)}(\boldsymbol{\varphi}) := \left( \tilde{g}_{1,t+1}^{2,(k)}(\boldsymbol{\varphi}), \dots, \tilde{g}_{p,t+1}^{2,(k)}(\boldsymbol{\varphi}), \tilde{g}_{t+1}^{2,(k)}(\boldsymbol{\varphi}), \tilde{\boldsymbol{\beta}}_{t+1}^{(k)} \right)$ ,  $\tilde{g}_{i,t}^{2,(k)}(\boldsymbol{\varphi}) = \partial \tilde{g}_{i,t}^2(\boldsymbol{\varphi}) / \partial \varphi_k$  and  $\tilde{g}_t^{2,(k)}(\boldsymbol{\varphi}) = \partial \tilde{g}_t^2(\boldsymbol{\varphi}) / \partial \varphi_k$ .

We finally turn to the asymptotic distribution of both the one-step and the multistep QML estimators. We first consider one-step QMLE. For notational simplicity, we define

$$\mathbf{I}_{1S} = E \frac{\partial \ell_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \frac{\partial \ell_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top}, \quad \mathbf{J}_{1S} = E \frac{\partial^2 \ell_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^\top}.$$

Theorem 2 establishes the asymptotic normality of the one-step QML estimator. The notation  $\xrightarrow{\mathcal{L}}$  stands for the convergence in law as  $n \rightarrow \infty$ .



THEOREM 2 (Asymptotic distribution of the one-step estimator). *Under the assumptions of Propositions 1 and 4 and A5-A10, if  $\xi_{0i} \neq 0$  for  $i = 1, \dots, p$ , we have*

$$\sqrt{n} \left( \widehat{\varphi}_n^{1S} - \varphi_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \mathbf{J}_{1S}^{-1} \mathbf{I}_{1S} \mathbf{J}_{1S}^{-1} \right).$$

Having stated the asymptotic normality of the one-step QML estimator, we now turn to the multistep QML estimator, which involves the estimation of all sets of GARCH parameters for the conditional volatility of both the error term and all the regressors (contained in the vector  $\boldsymbol{\theta}$ ) as well as the set of regression parameters used in the update of all time-varying betas (contained in  $\boldsymbol{\vartheta}$ ). Recall that the dimensions of  $\boldsymbol{\theta}$  and  $\boldsymbol{\vartheta}$  are  $d_1 = 4p$  and  $d_2 = 3(p+1) + pq$ , respectively. Let us define the score  $\mathbf{S}_t = \left( \frac{\partial \ell_{1,t}(\boldsymbol{\theta}_0^{(1)})}{\partial \boldsymbol{\tau}^\top}, \dots, \frac{\partial \ell_{p,t}(\boldsymbol{\theta}_0^{(p)})}{\partial \boldsymbol{\tau}^\top}, \frac{\partial \ell_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\vartheta}^\top} \right)^\top$  and the information matrices

$$\mathbf{I} = E \mathbf{S}_t \mathbf{S}_t^\top = \begin{pmatrix} \mathbf{I}_\theta & \mathbf{I}_{\theta\vartheta} \\ \mathbf{I}_{\theta\vartheta}^\top & \mathbf{I}_\vartheta \end{pmatrix}, \quad \mathbf{J}_{1S} = \begin{pmatrix} \mathbf{J}_\theta & \mathbf{J}_{\theta\vartheta} \\ \mathbf{J}_{\theta\vartheta}^\top & \mathbf{J}_\vartheta \end{pmatrix},$$

with obvious matrix partitions. Note that  $\mathbf{I}_\vartheta$  is also the bottom-right submatrix of size  $d_2 \times d_2$  of the matrix  $\mathbf{I}_{1S}$ . In addition, let  $\mathbf{J}_* = \text{diag} \left( \mathbf{J}^{(1)}, \dots, \mathbf{J}^{(p)} \right)$ .

Theorem 3 establishes the asymptotic normality of the multistep QML estimator.

THEOREM 3 (Asymptotic distribution of the multistep QML estimator). *Under the assumptions of Theorem 2, we have*

$$\sqrt{n} \left( \widehat{\varphi}_n - \varphi_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \boldsymbol{\Sigma} \right), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_\theta & \boldsymbol{\Sigma}_{\theta\vartheta} \\ \boldsymbol{\Sigma}_{\theta\vartheta} & \boldsymbol{\Sigma}_\vartheta \end{pmatrix},$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_\theta &= \mathbf{J}_*^{-1} \mathbf{I}_\theta \mathbf{J}_*^{-1}, \quad \boldsymbol{\Sigma}_{\theta\vartheta} = -\mathbf{J}_*^{-1} \mathbf{I}_\theta \mathbf{J}_*^{-1} \mathbf{J}_{\theta\vartheta} \mathbf{J}_\vartheta^{-1} + \mathbf{J}_*^{-1} \mathbf{I}_{\theta\vartheta} \mathbf{J}_\vartheta^{-1}, \\ \boldsymbol{\Sigma}_\vartheta &= \mathbf{J}_\vartheta^{-1} \mathbf{J}_{\theta\vartheta}^\top \mathbf{J}_*^{-1} \mathbf{I}_\theta \mathbf{J}_*^{-1} \mathbf{J}_{\theta\vartheta} \mathbf{J}_\vartheta^{-1} + \mathbf{J}_\vartheta^{-1} \mathbf{I}_\vartheta \mathbf{J}_\vartheta^{-1} - 2 \mathbf{J}_\vartheta^{-1} \mathbf{J}_{\theta\vartheta}^\top \mathbf{J}_*^{-1} \mathbf{I}_{\theta\vartheta} \mathbf{J}_\vartheta^{-1}. \end{aligned}$$

When  $\eta_t$  and  $\boldsymbol{\epsilon}_t$  are independent, we have  $\mathbf{I}_{\theta\vartheta} = \mathbf{0}_{d_1 \times d_2}$ .

### 3.6 Testing the assumption of constant conditional betas

Testing the assumption of the constancy of conditional betas is vital, as it provides an empirical justification for time-varying parameter models such as the ACB model. Naturally, a number of tests for the constancy of regression parameters are available in the literature. These include the class of CUSUM tests originally proposed by Brown et al. (1975), the Lagrange multiplier tests proposed by Nyblom (1989) and Hansen (1992), the tests for structural breaks proposed by Andrews (1993), and the parameter constancy tests proposed by Bai (1996) and Bai and Perron (1998). The available battery of tests is considerable.

Specifically, Nyblom (1989) proposes a test for the null hypothesis that all parameters are constant against the alternative that some of the parameters are time-varying.

Building upon the Nyblom test, Hansen (1992) allows for testing the constancy of single parameters. Let  $\hat{\beta}_i$  be an estimate of  $\beta_i$  obtained by maximizing the empirical mean of (3) under the constancy null hypothesis of the  $i$ -th beta; i.e.,  $H_0 : \beta_{i,t} = \beta_i$ . Hansen (1992) derives the limiting distribution of the statistic  $L_i = \frac{1}{nV_i} \sum_{t=1}^n F_{i,t}^2$  under the null hypothesis, where  $f_{i,t} = \frac{\partial \hat{l}_t}{\partial \beta_i}$ ,  $F_{i,j} = \sum_{j=1}^t f_{i,j}$  and  $V_i = \sum_{t=1}^n f_{i,t}^2$ , and suggests rejecting  $H_0$  at the nominal level of 5% when  $L_i > 0.470$ . This test is designed to have power under various alternatives but has the locally highest power when  $\beta_{i,t}$  follows a martingale process.

Additionally, it is important to test for the constancy of the time-varying betas that are specifically filtered by the ACB model. Such tests explore the structure of the ACB model and can give us further confidence that the betas are time-varying. There are, however, substantial challenges in deriving such tests because under the null hypothesis, some parameters might not be identifiable. In particular, in the absence of exogenous variables in the  $i$ -th conditional beta, Equation (4) is simplified as

$$\beta_{i,t+1} = \varpi_i + \xi_i \frac{v_t x_{i,t}}{\mu_i^2 + g_{i,t}^2} + c_i \beta_{i,t}, \quad (30)$$

and therefore,  $\beta_{i,t+1}$  converges to  $\varpi_i / (1 - c_i)$  when  $\xi_i = 0$ . Imposing the constancy of some conditional betas is therefore desirable, not only to avoid convergence problems due to this identifiability issue but also to reduce the number of parameters to estimate.

The problem of testing hypotheses when a model is not identified under the null has been well studied in the literature, see Amendola and Francq (2009) and the references therein. An obvious solution consists in fixing the parameter that is not identified under the null hypothesis to a plausible value.

In our case, this corresponds to fixing  $c_i$  to some value  $\bar{c}_i$ . The first strategy is to choose  $\bar{c}_i = 0$ . The  $i$ -th conditional beta therefore simplifies to  $\beta_{i,t+1} = \varpi_i + \xi_i \frac{v_t x_{i,t}}{\mu_i^2 + g_{i,t}^2}$ . In this setting, there is no substantial issue with testing the null hypothesis for which  $H_0 : \xi_i = 0$  since there is no unidentified parameter under the null. This solution is similar to that of testing for no conditional heteroskedasticity in the context of an ARCH model rather than a GARCH model. This test can be extended to account for  $q$  exogenous regressors  $z_{1t}, \dots, z_{qt}$  in the updating equation of  $\beta_{it}$ , i.e.,

$$\beta_{i,t+1} = \varpi_i + \xi_i \frac{v_t x_{i,t}}{\mu_i^2 + g_{i,t}^2} + \gamma_{1,i} z_{1,t} + \dots + \gamma_{q,i} z_{q,t}. \quad (31)$$

The null hypothesis of parameter constancy corresponds to  $H_0 : \xi_i = \gamma_{1,i} = \dots = \gamma_{q,i} = 0$ . Again, the absence of an autoregressive term ensures that there is no identification problem under the null hypothesis. The advantage of this test is that it is expected to have no size distortion, but its main drawback is that it is likely to have low power when  $\beta_{i,t}$  is very persistent.

The second option is to choose a value for  $\bar{c}_i$  that is plausible under the alternative hypothesis. Two values are considered in the simulations reported in the next section, i.e.,  $\bar{c}_i = 0.9$  and  $\bar{c}_i = 0.95$ . Naturally, the performance of this test depends on the choice of  $\bar{c}_i$ . A supremum test statistic over all possible values of  $c_i$  in a compact (partially)

solves the problem of the choice of  $\bar{c}_i$ , but leads to a non standard distribution under the null, and does not always increase the power.

In the next section, we investigate the properties of these tests across a wide range of empirically relevant scenarios.

## 4 SIMULATIONS

In this section, we present a Monte Carlo experiment that studies the finite-sample properties of the ACB model. The endogenous variable is specified as in (1).

Several cases are considered. In all simulations, we set  $p = 4$ . An intercept is included in the model by setting  $x_{1,t} = 1\forall t$ . We let  $x_{2,t}$ ,  $x_{3,t}$  and  $x_{4,t}$  be the first three explanatory variables used in the empirical application, i.e., the daily log-returns (in %) of the market, SMB and HML. A GARCH(1,1) model is estimated on these three series to obtain the quantities  $\mu_i$  and  $g_{i,t}$  involved in the dynamics of the conditional betas.

The conditional betas follow the specifications given in (4), while the conditional variance of the residuals  $v_t$  follows the GARCH(1,1) specification given in (7). For the parameters of the conditional beta equations, we choose  $\varpi_2 = 0.06$ ,  $\varpi_3 = 0.04$ ,  $\varpi_4 = 0.02$  and  $\xi_i = 0.05$ ,  $c_i = 0.94$  for  $i = 2, 3, 4$ . As such, in the absence of exogenous variables in the conditional betas,  $\beta_{i,t}$  meander around the values of 1, 0.5 and 0.25, respectively for  $i = 2, 3, 4$ . For the GARCH specification of the error term  $v_t$ , we set  $\omega = 0.005$ ,  $\alpha = 0.05$  and  $\beta = 0.94$  so that the unconditional variance of  $v_t$  equals 0.5.

We consider three specifications of the ACB model:

- Case 1: ACB model with an intercept and no exogenous variable in the betas. We obtain this model by setting  $q = 0$ ,  $\varpi_1 = 0.001$  and  $\xi_1 = c_1 = 0$  so that  $\beta_{1,t} = 0.001\forall t$ . The other three conditional betas are, however, time-varying.
- Case 2: ACB model with a time-varying intercept and no exogenous variable (i.e.,  $q = 0$ ) in the betas. This model extends Case 1 by setting  $\varpi_1 = 0.001$ ,  $\xi_1 = 0.05$ ,  $c_1 = 0.94$  so that  $\beta_{1,t}$  varies over time around its unconditional level of about 0.0166.
- Case 3: ACB model with a time-varying intercept and  $q = 1$  exogenous variable in the conditional betas of  $x_{2,t}$ ,  $x_{3,t}$  and  $x_{4,t}$ . This model extends Case 2 by adding a variable  $z_{1,t}$  corresponding to the value of the VIX of the previous day. We set  $\gamma_{1,2} = 0.02$  and  $\gamma_{1,3} = \gamma_{1,4} = 0$  so that an increase in the VIX translates into an increase in the market beta the day after but has no effect on the conditional betas of the other two explanatory variables. This result is compatible with the empirical result we obtained for the Durbl portfolio, one of the 12 industry portfolios considered in the second application reported in the online supplement.

Before presenting the results on the finite-sample properties of the ACB model, we illustrate the importance of the invertibility condition discussed in Sections 2.2-2.3. Figure 1 plots one path of the first 1,000 values of the three time-varying conditional betas  $\beta_{i,t}(\varphi_0)$  for  $i = 2, \dots, 4$  (solid black lines) generated with the model corresponding to Case 1. We also plot 11 paths of filtered time-varying conditional betas  $\tilde{\beta}_{i,t}(\varphi_0)$  at the

true parameter values but using different initial values  $\tilde{\beta}_{i,0}$  set to  $\tilde{\beta}_{i,0} = \beta_{i,0} + 0.1k$  for  $k = -5, 4, \dots, 4, 5$ , where  $\beta_{i,0}$  is the true value of the  $i$ -th beta at time  $t = 0$ . The bottom right graph plots  $\Delta_n(\varphi_0, k)$  (see Section 2.3) for  $k = 1, \dots, 20$ . This figure shows that the effect of the initialization vanishes rapidly and therefore that the filter is invertible. This is confirmed by the bottom right panel because  $\Delta_n(\varphi_0, k) < 0$  for all the chosen values of  $k$ .

Figure 2 plots the same information as in Figure 1 but for an ACB model with  $\xi_2 = \xi_3 = \xi_4 = -0.01$  and  $c_2 = c_3 = c_4 = 0.99$  and  $\varpi_2 = 0.01, \varpi_3 = 0.005, \varpi_4 = 0.0025$  so that the unconditional betas are the same as in Case 1. Recall that the ARMA(1,1) is a particular case of the ACB model, obtained by setting  $p = 1, q = 0$  and  $x_{1,t} = 1$ . In this case, an ACB (or ARMA) model with  $\xi_1 = -0.01$  and  $c_1 = 0.99$  would not be invertible because the invertibility condition of this model (see Example 2 in the appendix) is  $\sup_{\varphi \in \Theta_\varphi} |c_1 - \xi_1| < 1$  while at the true parameter values  $|c_1 - \xi_1| = 1$ . Our example is more general but we can infer that this model is not invertible. Indeed, even in the simpler case that  $p = 1$  where  $x_{1,t}$  follows a GARCH(1,1) with  $\mu_1 = 0, \Lambda_t(\varphi_0) = -\frac{\xi_1 x_{1,t}^2}{\mu_1^2 + g_{1,t}^2(\varphi_0)} + c_1$  and therefore  $E \log(\Lambda_t(\varphi_0)) = 0$  since  $E \left( \frac{x_{1,t}^2}{\mu_1^2 + g_{1,t}^2(\varphi_0)} \right) = 1$ . This is confirmed in Figure 2 where we see that the effect of the initial values never vanishes and therefore that the filter is not invertible at the true parameter values. Indeed, all the paths of the filtered conditional betas move in parallel and never cross each other. This is confirmed by the bottom right panel because  $\Delta_n(\varphi_0, k) > 0$  for all the chosen values of  $k$ .

To study the finite-sample properties of the ACB model, the models corresponding to Cases 1, 2 and 3 are estimated using the multistep QML estimator presented in the previous section, and the standard errors are computed using the results given in Theorem 3. While one-step QMLE estimation is possible, we focus on the multistep QML estimator, as the log-likelihood of a one-step QML estimator tends to be quite flat around the true parameter value, leading to very poor accuracy on reasonable sample sizes.

For the sake of illustration, we plot in Figure B1 of the online supplement one path of  $n = 4,000$  observations of both simulated (green dotted lines) and estimated (blue solid lines) time-varying conditional betas corresponding to Case 1. This figure suggests that the estimated conditional betas are very close to the true betas. Interestingly, we also see in this example that the estimated conditional betas are rapidly approaching the true conditional betas, even when the initialization of the three time-varying conditional betas is not very good (as for  $\beta_{2,t}$ ).

The results of the Monte Carlo simulation of Case 1 for 1,000 replications are reported in Table 1 while those of Cases 2 and 3 are reported respectively in Tables B1 and B2 of the online supplement. Each table is divided in two panels. The left panels correspond to a sample size of  $n = 2,000$  observations, while the right panels correspond to  $n = 4,000$  observations. The following information is reported in each table: the true values, bias, RMSE, average of the asymptotic standard errors and the frequency of the estimated parameters outside of the 95% confidence interval based on the asymptotic distribution.

We highlight that the biases are very small in all three cases for both sample sizes

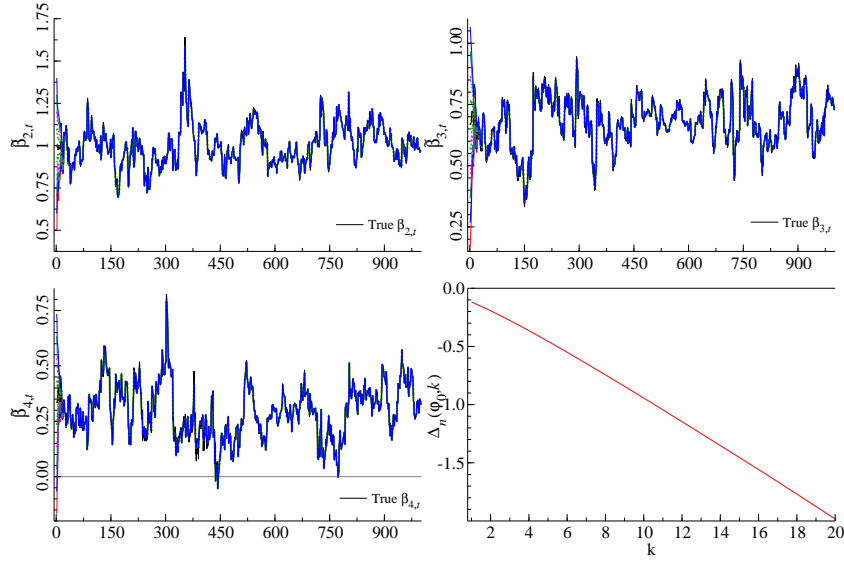


Figure 1: Invertibility condition satisfied. The plot shows 1,000 values of three simulated conditional betas (black solid line) for a DGP satisfying the invertibility condition. It displays also 11 paths of filtered betas  $\tilde{\beta}_{t,i}$  ( $i = 2, 3$  and 4) evaluated at the true parameter values but using different initial values  $\tilde{\beta}_{0,i}$  set to  $\tilde{\beta}_{0,i} = \beta_{0,i} + 0.1k$  for  $k = -5, 4, \dots, 4, 5$ , where  $\beta_{0,i}$  is the true value of the  $i$ -th beta at time  $t = 0$ . The bottom right graph plots  $\Delta_n(\varphi_0, k)$  for  $k = 1, \dots, 20$ .

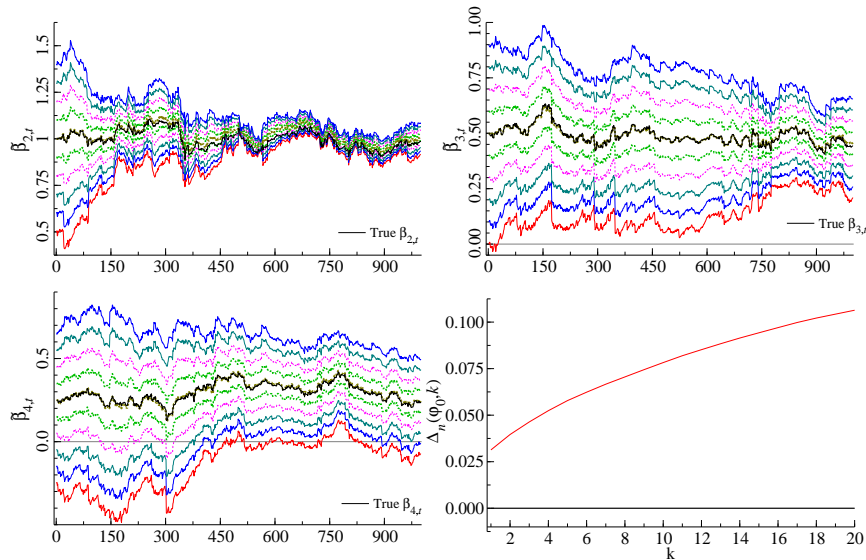


Figure 2: Invertibility condition not satisfied. Note: see Figure 1.

Table 1: True values, bias, RMSE, average of the asymptotic standard errors and frequency of the estimated parameters outside of the 95% confidence interval for a DGP and the estimated model corresponding to Case 1. Sample sizes:  $n = 2,000$  and  $4,000$ .

	$n = 2,000$					$n = 4,000$				
	True	Bias	RMSE	aSTD	1-Cov95	True	Bias	RMSE	aSTD	1-Cov95
$\varpi_1$	0.001	0.000	0.014	0.014	4.400	0.001	-0.000	0.010	0.010	4.900
$\varpi_2$	0.060	0.008	0.036	0.028	7.400	0.060	0.002	0.017	0.017	5.500
$\xi_2$	0.050	0.004	0.013	0.012	7.400	0.050	0.002	0.008	0.008	6.500
$c_2$	0.940	-0.008	0.035	0.028	7.400	0.940	-0.002	0.017	0.017	5.500
$\varpi_3$	0.040	0.004	0.019	0.018	5.600	0.040	0.002	0.012	0.012	5.500
$\xi_3$	0.050	0.004	0.013	0.012	7.700	0.050	0.001	0.008	0.008	5.200
$c_3$	0.940	-0.007	0.028	0.026	5.500	0.940	-0.004	0.019	0.018	5.100
$\varpi_4$	0.020	0.003	0.013	0.011	7.400	0.020	0.001	0.007	0.006	6.700
$\xi_4$	0.050	0.001	0.012	0.012	6.600	0.050	-0.001	0.008	0.008	6.900
$c_4$	0.940	-0.009	0.037	0.030	9.100	0.940	-0.003	0.019	0.018	7.500
$\omega$	0.005	0.002	0.004	0.003	3.100	0.005	0.001	0.002	0.002	4.000
$\alpha$	0.050	-0.000	0.010	0.010	7.500	0.050	-0.000	0.007	0.007	6.200
$\beta$	0.940	-0.003	0.014	0.013	6.600	0.940	-0.001	0.009	0.009	5.100
Average		0.000			6.592		-0.000			5.738

Note: The Monte Carlo simulation results are for  $n = 2,000$  (left panel) and  $n = 4,000$  (right panel) and 1,000 replications. The column labeled ‘True’ reports the true parameters, the column labeled ‘Bias’ shows the empirical mean of the deviation in the estimated parameters from the true value, the column labeled ‘RMSE’ provides the root mean squared error, the column labeled ‘aSTD’ shows the average of the estimated standard errors, and finally, the column labeled ‘1-Cov95’ provides the frequency of the estimated parameters outside of the 95% confidence interval based on the asymptotic distribution. The models are estimated using the multistep QML estimator. The average of the Bias and 1-Cov95 columns are also reported in the last row of each panel.

and all parameters. The average biases reported in the last row of each panel (over all parameters) are all close to 0. The RMSEs are always very close to the average of the standard errors computed from the asymptotic distribution. The frequencies of true parameters outside of the 95% confidence interval drawn from the asymptotic distribution are reasonably close to 5%, with an average between 6.592 and 5.738 for  $n = 2,000$  and 4,000 respectively; this confirms the asymptotic normality result derived in Section 3 for these sample sizes.

To show the robustness of our QMLE when the innovations are not Gaussian, we report in Table B3 of the online supplement additional Monte Carlo simulation results for Case 1 with  $n = 4,000$  when  $\eta_t$  follows respectively a standardized Student-t distributed with a degree of freedom of 5, denoted  $ST(0, 1, 5)$ , and a standardized version of the Skewed Student-t distribution of Fernández and Steel (1998) proposed by Giot and Laurent (2003) with a degree of freedom of 5 and an asymmetry parameter of  $\exp(-0.1)$ , denoted  $SK - ST(0, 1, \exp(-0.1), 5)$ , so that the innovations have fat-tails and are left skewed. Results suggest that the Gaussian QMLE performs similarly when the distribution of the innovations is Normal, when it has fat-tails and even when it is skewed.

We now investigate the finite sample properties of the three tests for the null hypothesis of a constant conditional beta presented in Section 3.6. The rejection frequencies of the null hypothesis  $H_0 : \beta_{1,t} = \beta_1$  are reported in Table 2. To study the size of the tests, we consider the model corresponding to Case 1 (i.e., constant intercept). The first row of the column labeled ‘Size (Case 1)’ contains the rejection frequency of the individual Nyblom test on the intercept  $\varpi_1$  in the model estimated under the null hypothesis (the other three conditional betas and the conditional variance of the error term are correctly speci-

fied in all simulations). The next three rows correspond to the rejection frequencies of the null hypothesis of a Wald test for the null hypothesis  $H_0 : \xi_1 = 0$  in a model specifying  $\beta_{1,t}$  as  $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t$ ,  $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t + 0.9\beta_{1,t}$  and  $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t + 0.95\beta_{1,t}$ .

To study the power of these tests, we first consider the model corresponding to Case 2, where  $\beta_{1,t+1} = 0.001 + 0.05v_t + 0.94\beta_{1,t}$ , and therefore,  $\beta_{1,t}$  is time-varying.

Table 2: Finite sample properties of the tests for constant conditional betas.

	Size (Case 1)	Power (Case 2)
Nyblom test for $\beta_{1,t+1} = \varpi_1$	3.2	75.6
Wald test for $\xi_1 = 0$ in $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t$	6.5	19.2
Wald test for $\xi_1 = 0$ in $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t + 0.90\beta_{1,t}$	5.2	99.8
Wald test for $\xi_1 = 0$ in $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t + 0.95\beta_{1,t}$	6.5	100.0

Note: The empirical size and power are studied for a DGP corresponding to Cases 1 and 2. In Case 1,  $\beta_{1,t} = 0.001$ , and therefore, the slope coefficient of the intercept ( $x_{1,t} = 1$ ) is constant. In Case 2,  $\beta_{1,t+1} = 0.001 + 0.05v_t + 0.94\beta_{1,t}$ , and therefore,  $\beta_{1,t}$  is time-varying.

The results reported in Table 2 are for a nominal size of 5% and a sample size of 4,000 observations. The four tests have relatively good sizes. Importantly, fixing  $c_i$  at either 0.9 or 0.95 does not affect much the size of the Wald test. The Nyblom test is found to have good power (more than 75%). As expected, the power of the Wald test based on the ACB specification without an autoregressive term (i.e.,  $\beta_{1,t+1} = \varpi_1 + \xi_1 v_t$ ) is quite low and even lower than that of the Nyblom test. Interestingly, the power of the Wald test based on the ACB specification when  $c_1$  is fixed either to 0.9 or 0.95 is higher than the other two tests and close to 100% in this example.

## 5 Empirical Application

In this section, we illustrate the performance of the ACB model in a risk management experiment similar to those in Engle (2016) and Darolles et al. (2018). The main objective is to find a strategy to track US stocks using a six-factor model.

The six factor model contains the risk factors of the three-factor model of Fama and French (1993), i.e., the market factor  $MKT$  proxied by the log-returns on the SP500 index, the standard Fama French size, the  $SMB$  (*Small Minus Big*) and  $HML$  (*High Minus Low*) value factors. The other three factors are the  $RMW$  (*Robust Minus Weak*) and  $CMA$  (*Conservative Minus Aggressive*) factors initially proposed by Fama and French (2015) as well as  $MOM$  (*Momentum*) as in Carhart (1997).

The starting point of this application is the following model with time-varying parameters, presented for a generic series  $y_t$ :

$$y_t = \beta_{INT,t} + \beta_{MKT,t}MKT_t + \beta_{SMB,t}SMB_t + \beta_{HML,t}HML_t + \beta_{RMW,t}RMW_t + \beta_{CMA,t}CMA_t + \beta_{MOM,t}MOM_t + v_t, \quad (32)$$

where  $\beta_{INT,t}$  is the time-varying conditional intercept (often called alpha in the financial literature) and  $\beta_{MKT,t}$ ,  $\beta_{SMB,t}$ ,  $\beta_{HML,t}$ ,  $\beta_{RMW,t}$ ,  $\beta_{CMA,t}$  and  $\beta_{MOM,t}$  are the six time-varying conditional betas. On total, there are potentially seven time-varying parameters (in addition to the time-varying variance), i.e.,  $\beta_{INT,t}, \dots, \beta_{MOM,t}$ , that we sometimes

simply denote  $\beta_{i,t}$  for  $i = 1, \dots, 7$  for ease of notation. Note also that we do not add another index to denote the forecasting model used to produce the conditional betas, but it is clear that the tracking portfolios are model dependent.

In the asset pricing context, expected returns on any asset are linear in the betas and only depend upon the risk premiums embedded in the factors so that  $\beta_{INT,t} = 0 \forall t$  (i.e., there is no alpha or intercept in (32)). We perform a tracking exercise that consists of taking a position at time  $t$  in the 6 considered factors whose weights are the one-step-ahead forecasts of the corresponding conditional betas. For each model, the conditional beta forecasts are, therefore, used to construct a hedging portfolio. The returns at time  $t + 1$  of such a tracking portfolio (denoted  $Z_{t+1|t}$ ) are obtained by setting the intercept to zero and using the conditional beta forecasts as weights of the realized returns at time  $t + 1$  for the six factors; i.e.,

$$\begin{aligned} Z_{t+1|t} = & \beta_{MKT,t+1|t}MKT_{t+1} + \beta_{SMB,t+1|t}SMB_{t+1} + \beta_{HML,t+1|t}HML_{t+1} \\ & + \beta_{RMW,t+1|t}RMW_{t+1} + \beta_{CMA,t+1|t}CMA_{t+1} + \beta_{MOM,t+1|t}MOM_{t+1}, \end{aligned} \quad (33)$$

where  $MKT_{t+1}$  is the realized return of the market on day  $t + 1$  (the other variables indexed by  $t + 1$  are defined in the same way).

The performance of a model is obtained by computing the realized tracking errors  $TE_{t+1} = y_{t+1} - Z_{t+1|t}$ , where  $y_{t+1}$  is the realized return of the considered series on day  $t + 1$ . We therefore seek the models with the smallest sample mean squared error (MSE). To do so, we rely on the model confidence set in Hansen et al. (2011) and, more particularly, the Mulcom 3.0 package in Hansen et al. (2021), which is an Ox econometric toolkit for multiple comparisons.

We consider 16 US stocks (randomly selected among the most liquid ones) whose tickers are AXP, BA, CAT, DIS, GE, JPM, KO, MCD, MRK, MSFT, PFE, PG, UTX, VZ, WMT and XOM. Several models compatible with Equation (32) are considered in this empirical application and estimated on daily data. Some models are nested by the ACB model, while others are not. All models are used to produce one-step-ahead forecasts of the betas. The first forecasts are obtained from models estimated on a window of  $n = 1,500$  observations. The sample is then expended by one observation each time a new forecast is produced until the end of the sample is reached. The parameters of the models are re-estimated every 25 days. The competing models based solely on daily data are listed below.

OLS: This model is an ACB model without dynamics in the conditional betas and a constant conditional variance for the error term. It is obtained by setting  $q = 0$  and imposing  $\xi_i = c_i = 0$  for all the betas in (32) as well as  $\alpha = \beta = 0$ . This model is therefore equivalent to an OLS regression with constant parameters.

OLS(100): This model is identical to the previous model but the model is estimated on the latest 100 observations instead of the full sample available.<sup>3</sup>

---

<sup>3</sup>Results of a linear model estimated on the latest 250 observations were qualitatively the same and therefore not reported to save space.



- ACB: This model is an ACB model with a time-varying intercept and six factors in the conditional mean so that all the betas are allowed to be time-varying using a pure (Q)SD dynamic without additional exogenous variable in the conditional betas (i.e.,  $q = 0$ ).
- SS-RW: This model allows each conditional beta (including the constant) to follow a random walk; i.e.,  $\beta_{i,t} = \beta_{i,t-1} + \delta_i u_{i,t}$ , where  $\delta_i \geq 0$ ,  $u_{i,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1) \forall i$  and  $u_{i,t} \perp u_{j,t} \forall i \neq j$ . This model is estimated by Gaussian ML via its state space formulation and requires the use of the Kalman filter. Note that when  $\delta_i = 0$ , the  $i$ -th beta is constant over time.<sup>4</sup>
- DCB: This model, proposed by Engle (2016) and called the dynamic conditional betas (DCBs) model, allows obtaining time-varying betas from an estimate of the conditional covariance matrix. As in Engle (2016), we rely on the DCC of Engle (2002) with GARCH(1,1) models for the conditional variances.

We have data for the period spanning from January 1999 to the end of December 2017 (i.e., 4,778 observations) for all series but CAT, DIS, GE, JPM and KO for which the sample starts in January 2005 (i.e., 3,269 observations). The daily data used to estimate the above models are obtained by aggregating 5-minute log-returns on the 16 US stocks (transaction prices purchased from tickdatamarket) and the six factors. The data on the factors are obtained from Dacheng Xiu’s website.<sup>5</sup> See Ait-Sahalia et al. (2020) for more details on the construction of these series.

Following the suggestion of the editor, we use intraday data to compute realized measures of the above daily conditional betas. There is indeed a growing literature on realized betas, initiated by Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2006), that uses high-frequency data to compute lower frequency realized betas. We estimate realized betas day by day using the 5-minute data under the assumptions that the drift is negligible and that the realized betas are constant over the day. In this case, realized betas of day  $t$  (denoted  $RB_{i,t}$ , for  $i = 2, \dots, 7$ ) correspond to the usual OLS estimates of a linear model regressing the 5-minute log-returns of a stock on the six factors (without intercept). Other estimators of realized betas that are robust to time-variation in the spot betas are discussed in Ait-Sahalia et al. (2020) and Andersen et al. (2021). We let this extension to future work but to attenuate the effect of large jumps when estimating the realized betas, we follow Ait-Sahalia et al. (2020) and truncate intraday log-returns of day  $t$  of the stocks and factors using a threshold value of  $3(\frac{1}{78})^{0.47} BV_t$ , where  $BV_t$  is the bipower variation on day  $t$  of the correspond series.

We, therefore, also consider two methods relying on realized betas.

- True RB: This method is infeasible in practice because it uses realized betas on day  $t + 1$  in (33) to form the tracking portfolios. Although infeasible, this method allows to

<sup>4</sup>Koundouri et al. (2016) extended the SS-RW model by allowing the time-varying betas to follow a zero-mean AR(1) process.

<sup>5</sup>Note that log-returns are reconstructed from the raw returns provided on his website and that as in Ait-Sahalia et al. (2020) the log-returns are not in excess of the risk free rate.

assess whether realized betas computed on 5-minute data are accurate proxies of the optimal weights in this tracking exercise.

ARFIMA RB: This method relies on a univariate model to forecast each realized beta series. To account for the presence of long-memory in the series of daily realized betas, we use an ARFIMA(1, $d$ ,0) model estimated by Gaussian maximum likelihood.

All models have been implemented in the Ox 8.0 programming language (see Doornik, 2012). Note that we rely on the G@RCH 8.0 software of Laurent (2018) for the DCC models, on the SsfPack 3.0 Ox library of Doornik et al. (2008) for the SS-RW model and the ARFIMA package of Doornik and Ooms (2021) for the ARFIMA model. The ACB models are estimated using the multistep QML estimator presented in Section 3 and used in all Monte Carlo simulations.

For the sake of illustration, we report in Table 3 the results of the ACB model for series XOM (Exxon Mobil) estimated on the full sample.

Some comments are in order.

1. The intercept (i.e.,  $\beta_{INT,t}$ ) and the beta of SMB (i.e.,  $\beta_{SMB,t}$ ) are restricted to be constant over-time. This decision is taken after having estimated model (32) and having applied the Wald test discussed in Section 3.6 of the null hypothesis  $H_0 : \beta_{i,t} = \beta_i$ , when fixing  $c_i$  to 0.95. The null hypothesis is rejected at the 5% nominal level for all conditional betas except the intercept and SMB. Overall, there is weak evidence of dynamics in the intercept of the 16 stocks, except for series VZ (Verizon Communications).
2. The log-likelihood value of the estimated ACB is -6107.04 while the one of a GARCH model with constant slopes is -6622.52 so that this model is strongly rejected at any conventional significance level in favor of the ACB using a likelihood ratio test. We therefore conclude that there is strong evidence of dynamics in some of the conditional betas of the six-factor model. The same conclusion holds for the 16 series.
3. The estimated  $c_i$  parameters are all very close but smaller than one.
4. The estimated coefficients of the lagged scaled score in the conditional beta equations, i.e.,  $\xi_i$ , are all very close to 0 but positive and significant.
5. Finally, the three-factor model of Fama and French (1993) with time-varying betas, obtained by setting  $\beta_{RMW,t+1|t} = \beta_{CMA,t+1|t} = \beta_{MOM,t+1|t} = 0$ , is rejected in favor of the six-factor model using a likelihood ratio test. Indeed, the log-likelihood value of this model is -6443.87 so that this model is rejected at any conventional significance level. The same conclusion also holds for the other 15 series considered in this empirical application. We therefore report only the results of the six-factor model.

Note that we have tested the inclusion of the  $i$ -th (for  $i = 2, \dots, 7$ ) daily realized beta of the previous day (i.e.,  $\gamma_{1,i}RB_{i,t}$ ) in (4). The main conclusion is that the coefficients  $\gamma_{1,i}$

of  $RB_{i,t}$  are found to be insignificant and that their inclusion deteriorates the performance of the model in the financial application, reason why we do not report the results of this model.

Table 3: Estimation results of the ACB models for series XOM

Equation	Parameters	Coefficient	Std.Error	$p$ -value
Intercept	$\varpi_1$	-0.026	0.011	0.022
MKT	$\varpi_2$	0.002	0.001	0.072
	$\xi_2$	0.008	0.002	0.001
	$c_2$	0.998	0.001	0.000
SMB	$\varpi_3$	-0.358	0.032	0.000
HML	$\varpi_4$	0.001	0.001	0.403
	$\xi_4$	0.011	0.003	0.000
	$c_4$	0.996	0.002	0.000
CMA	$\varpi_5$	0.000	0.001	0.397
	$\xi_5$	0.008	0.002	0.000
	$c_5$	0.996	0.001	0.000
RMW	$\varpi_6$	0.000	0.000	0.944
	$\xi_6$	0.009	0.002	0.000
	$c_6$	0.997	0.001	0.000
MOM	$\varpi_7$	0.000	0.001	0.597
	$\xi_7$	0.021	0.003	0.000
	$c_7$	0.999	0.001	0.000
$v_t$	$\omega$	0.007	0.004	0.065
	$\alpha$	0.067	0.013	0.000
	$\beta$	0.928	0.015	0.000
Log-likelihood	-6107.04			

Note: QML estimates of the ACB model for series XOM on the full sample spanning from January 1999 to December 2017 (i.e., 4,778 observations). Both the intercept and the beta of the factor *SMB* are restricted to be constant.

The estimated conditional betas of the ACB model for series XOM are plotted in Figure 3 together with 95% confidence bands obtained using the simulation method proposed by Blasques et al. (2016). Interestingly, we see that the time-varying conditional betas obtained with the ACB model are slowly varying over-time and quite persistent. We also see from the confidence bands that the signals are estimated in a rather precise way and allow to distinguish periods where the conditional betas are significantly above or below their average.

To empirically check the invertibility of the ACB filter at the QML estimates  $\hat{\varphi}$ , we proceed in a way similar to Section 4 and plot in the first five graphs of Figure 4 eleven paths of filtered time-varying conditional betas  $\tilde{\beta}_{i,t}(\hat{\varphi})$  for series XOM and initial values  $\tilde{\beta}_{i,0}$  set to  $\tilde{\beta}_{i,0} = \hat{\beta}_{i,0} + 0.1k$  for  $k = -5, 4, \dots, 4, 5$ , where  $\hat{\beta}_{i,0}$  is the initial value of the  $i$ -th beta used in the estimation by QML. The bottom right graph plots  $\Delta_n(\hat{\varphi}, k)$  for  $k = 1, \dots, 21$ , i.e.,  $\Delta(\varphi, N)$  evaluated at the estimated values. This figure shows that the effect of the initialization vanishes rapidly and therefore that the filter is invertible.

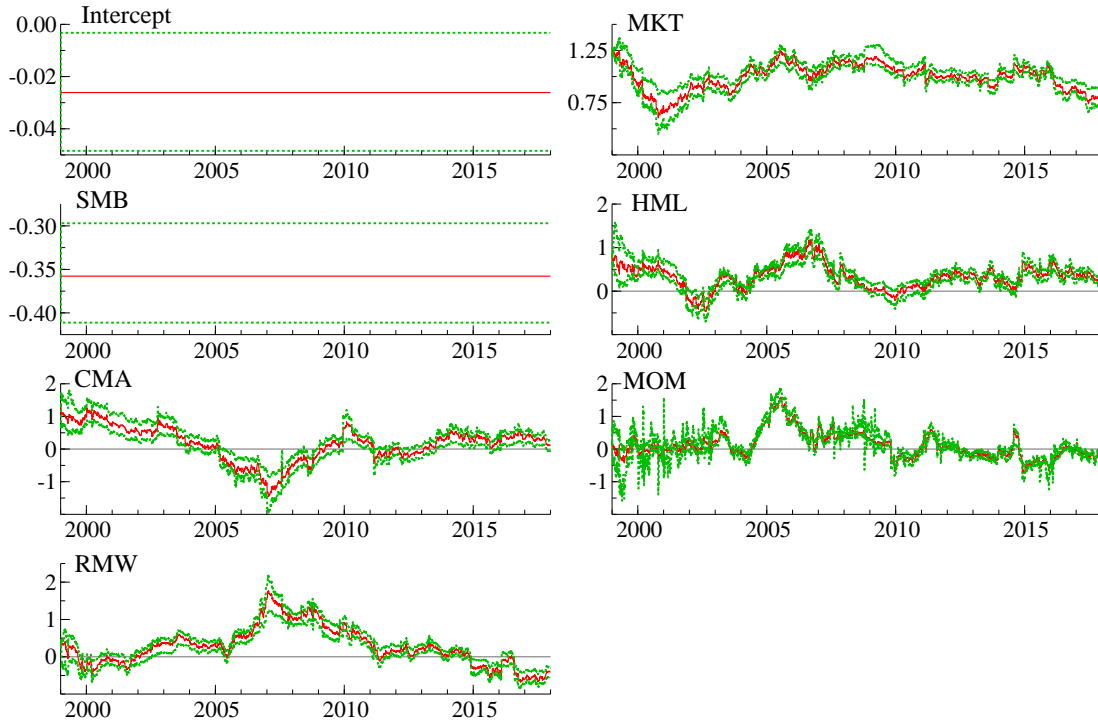


Figure 3: Time series graph of the conditional betas of series XOM (solid red lines) and 95% confidence bands (dotted green lines) estimated on the full period with the ACB model.

This is confirmed by the bottom right panel because  $\Delta_n(\hat{\varphi}, k) < 0$  for  $k > 15$ . The same conclusion holds for the 16 series considered in the application.

For the sake of comparison, we also plot in Figure 5 the estimated conditional betas of three competing models performing well in the tracking exercise. The solid red and dashed blue lines correspond respectively to the ACB and SS-RW models estimated on daily returns while the dotted gray lines correspond to the ARFIMA(1,d,0) fitted to the daily realized betas. It can be seen that the in-sample forecasts of the ARFIMA model estimated on the realized betas are the most erratic ones and that the trajectories of the conditional betas of the ACB and SS-RW models are quite close, although different. It can be concluded that the three methods give different but similar filtered paths.

Table 4 contains the results of the MCS test with an MSE loss function on the realized tracking errors (similar results are obtained using a median absolute deviation loss function). The figures reported in the first 16 rows correspond to the p-values of the MCS test implemented with 10,000 bootstrap samples and a block length of 5 observations. A p-value below the chosen critical level (e.g., 5%) means that the model does not belong to the MCS at that level and is therefore inferior to those belonging to the MCS. The figures reported in the rows labeled ‘#>0.05’ correspond to the number of times (out of 16) that the corresponding model belongs to the MCS at the 5% level.

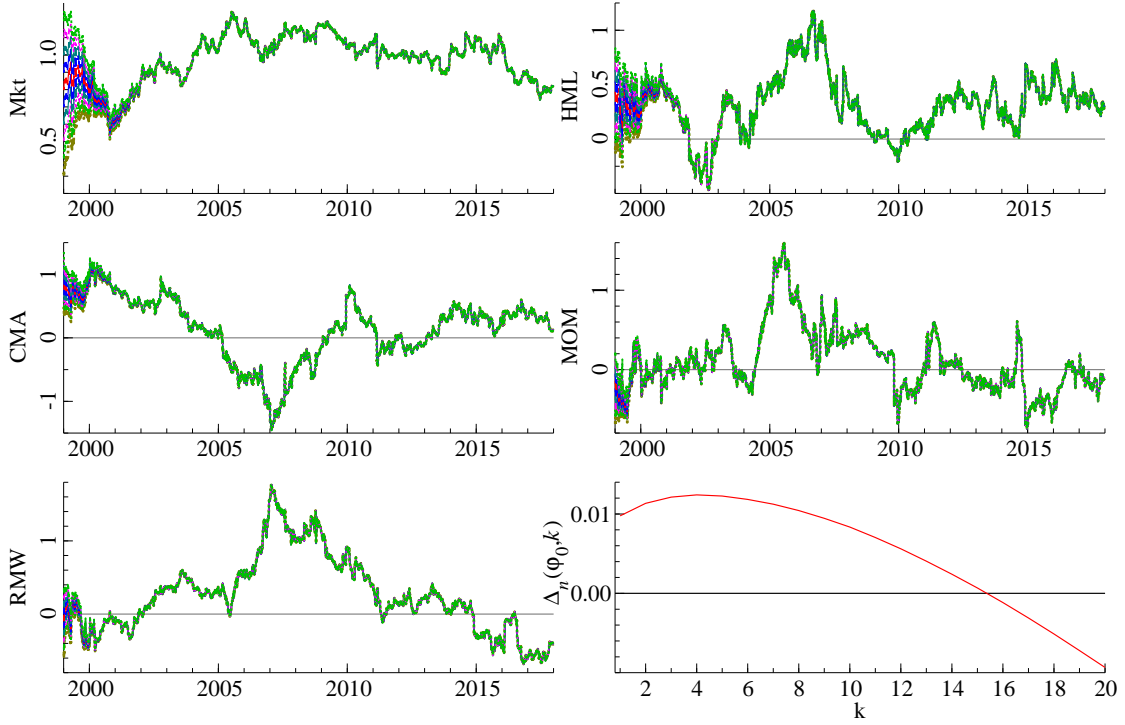


Figure 4: Empirical verification of the invertibility of the ACB filter at the QML estimates. The first 5 graphs plot 11 paths of filtered time-varying conditional betas  $\tilde{\beta}_{t,i}(\hat{\varphi})$  for series XOM and initial values  $\tilde{\beta}_{i,0}$  set to  $\tilde{\beta}_{i,0} = \hat{\beta}_{i,0} + 0.1k$  for  $k = -5, 4, \dots, 4, 5$ , where  $\hat{\beta}_{i,0}$  is the initial value of the  $i$ -th beta used in the estimation by QML. The bottom right graph plots  $\Delta_n(\hat{\varphi}, k)$  (see Section 2.3) for  $k = 1, \dots, 20$ .

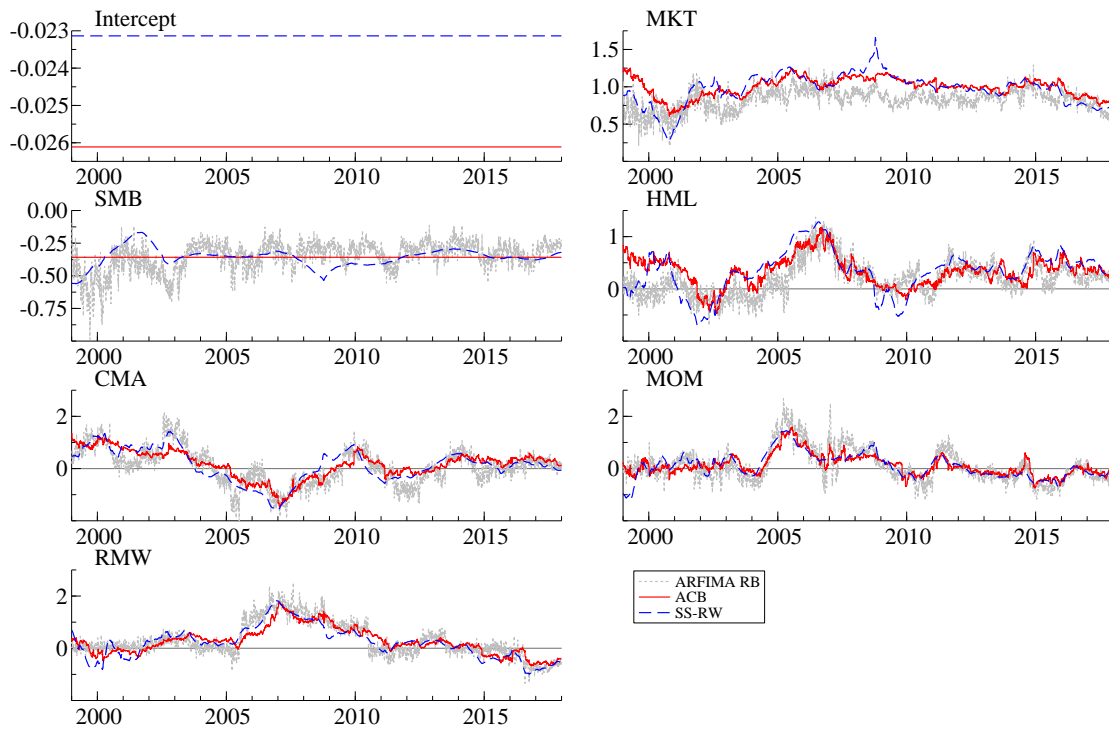


Figure 5: Time series graph of the conditional betas for the six-factor model estimated on the full period. The solid red and dashed blue lines correspond respectively to the ACB and SS-RW models estimated on daily returns while the dotted gray lines correspond to the ARFIMA(1,  $d$ , 0) fitted to the daily realized betas.

Note that the MCS test is applied on a universe of 7 models, including a strategy that is infeasible in practice. Indeed, recall that the method labeled ‘True RB’ relies on the realized betas observed on day  $t + 1$ , which are obviously not known at the end of day  $t$ . We consider this method to assess the relevance of realized betas in the context of a tracking exercise and the effect of a smoothing of the realized betas using an ARFIMA filter.

This table first confirms that if they were observed at time  $t$ , realized betas on day  $t + 1$  would deliver the best tracking portfolios for all the series considered in the application. Furthermore, univariate ARFIMA(1,  $d$ , 0) models fitted on these realized betas are almost as good as the infeasible method because this method belongs to the MCS at level 5% in 15 out of 16 cases. Importantly, the best model solely based on daily data is the ACB model, which belongs to the MCS in 13 cases, i.e. slightly less often than the ARFIMA. While belonging less often (i.e., in 10 cases) to the MCS than the ACB, the SS-RW model also performs pretty well. The DCB model and the models with constant parameters (estimated on the full sample or the latest 100 observations) clearly underperform compared to the other models.

Finally, we also compare the performance of the five models based on daily returns for forecasting conditional betas by comparing their forecasts with the corresponding realized betas. For each model and time period  $t + 1$ , the loss is measured as the sum over the six betas of the squared difference between the  $i$ -th realized beta at time  $t + 1$  and its forecast obtained on day  $t$ , i.e.,  $\sum_{i=2}^7 (RB_{i,t+1} - \hat{\beta}_{i,t+1|t})^2$ . The results reported in Table 5 suggest that the ACB model produces on average the best forecasts when the quality of the forecasts is assessed using daily realized betas (i.e., it is in the MCS at the 5% nominal level in 12 out of 16 cases while the second-best model is the SS-RW which belongs in the MCS in only 6 cases). However, when the ARFIMA model fitted on the realized betas is including in the set of competing models, this model outperforms all the other methods (results are not reported here to save space).

In Section C of the online supplement, we report the results of a similar application but for the 12 US industry portfolios. This application differs from the first one because we do not have intraday data to calculate the daily realized betas, we compare the performance of 3- and 5-factor models and also because we use exogenous variables in the conditional betas. The results of this second application allow us to confirm the good performance of the ACB model.

## 6 CONCLUSION

In this paper, we introduce a new (quasi) score-driven autoregressive conditional beta (ACB) model. This model allows conditional betas to depend on exogenous variables as well as conditional volatility in the error term. By scaling the updating step with the inverse conditional volatility of the regressors, the score-driven update achieves a smoother filtered beta, which naturally compensates for periods of high volatility in the regressors.

We establish conditions for the existence of unique stationarity solutions of the ACB model and the invertibility of the filtered time-varying parameters. We also derive an

Table 4: Results of the MCS test applied on the realized tracking errors of the 16 stocks

	OLS	OLS(100)	SS-RW	ACB	DCB	True RB	ARFIMA RB
AXP	0.000	0.000	1.000	0.752	0.000	0.752	0.752
BA	0.007	0.072	0.072	0.072	0.000	1.000	0.072
CAT	0.000	0.000	0.996	0.996	0.000	1.000	0.996
DIS	0.001	0.001	0.833	0.833	0.000	1.000	0.826
GE	0.033	0.051	0.000	0.403	0.000	1.000	0.479
JPM	0.000	0.000	0.000	0.352	0.000	0.352	1.000
KO	0.002	0.076	0.424	0.918	0.000	0.993	1.000
MCD	0.005	0.033	0.486	1.000	0.000	0.727	0.727
MRK	0.002	0.014	0.070	0.044	0.000	1.000	0.070
MSFT	0.000	0.000	0.000	0.002	0.000	1.000	0.013
PFE	0.000	0.000	0.013	0.000	0.000	0.647	1.000
PG	0.000	0.023	0.029	0.728	0.000	1.000	0.728
UTX	0.001	1.000	0.381	0.661	0.000	0.449	0.281
VZ	0.000	0.000	0.001	0.094	0.000	0.397	1.000
WMT	0.000	0.036	0.067	0.570	0.000	0.570	1.000
XOM	0.000	0.000	0.829	0.239	0.000	1.000	0.239
#>0.05	0.000	4.000	10.000	13.000	0.000	16.000	15.000

Note: The table shows the p-values of the model confidence set test applied on the the tracking errors of 7 competing models. The figures reported in the first 16 rows correspond to the p-values of the MCS test implemented with 10,000 bootstrap samples and a block length of 5 observations with a MSE loss function. The figures reported in the row labeled ‘#>0.05’ correspond to the number of times (out of 16) that each model belongs to the MCS at the 5% level. The model labeled ‘OLS(100)’ is estimated on the latest 100 observations of the available data while the method labeled ‘True RB’ is infeasible because it uses the observed realized betas of day  $t + 1$  to form the tracking portfolio. The other models are estimated by QML on expanding windows. The first sample consists of the first 1,500 available observations. The parameters are re-estimated every 25 days (i.e., approximately every month) so that the parameters are kept constant to produce 25 consecutive forecasts before being reevaluated. The total number of one-step-ahead forecasts (and therefore, realized tracking errors) is 3,278 for all series but CAT, DIS, GE, JPM and KO for which there are 1,767 forecasts.

empirically verifiable condition for invertibility. Additionally, we establish the consistency and asymptotic normality of both one-step and multistep QMLE.

A Monte Carlo simulation study reveals that multistep QMLE performs well with different sample sizes and model specifications. Furthermore, we also find good size and power properties for the proposed parameter constancy tests, which test for time variations in the betas within the specific context of the ACB model.

Empirically, we find strong evidence of time variation in the regression betas of a six-factor model. Overall, we highlight the empirical relevance of the ACB model in a financial econometrics context. In particular, we find using the MCS test with a MSE loss function of the tracking errors in a beta hedging exercise that our newly proposed ACB model is the best model solely based on daily data. This model performs slightly better than the SS-RW model (in which the conditional betas follow random walks), much



Table 5: Results of the MCS test applied on the forecasting errors of the realized beta of the 16 stocks

	OLS	OLS(100)	SS-RW	ACB	DCB
AXP	0.000	0.000	0.000	1.000	0.000
BA	0.000	1.000	0.000	0.000	0.000
CAT	0.000	0.000	0.002	1.000	0.000
DIS	0.000	0.000	0.000	1.000	0.000
GE	0.000	1.000	0.000	0.759	0.000
JPM	0.000	0.000	0.000	1.000	0.000
KO	1.000	0.000	0.000	0.810	0.000
MCD	0.000	0.004	1.000	0.984	0.000
MRK	0.000	0.000	1.000	0.002	0.000
MSFT	1.000	0.000	0.007	0.311	0.000
PFE	0.000	0.000	1.000	0.000	0.000
PG	0.000	0.055	0.000	1.000	0.000
UTX	0.000	0.184	1.000	0.000	0.000
VZ	0.000	0.000	0.394	1.000	0.000
WMT	0.000	0.000	0.000	1.000	0.000
XOM	0.000	0.000	1.000	0.053	0.000
#>0.05	2.000	4.000	6.000	12.000	0.000

Note: The table shows the p-values of the model confidence set test for the forecasting accuracy of the one-step-ahead forecasts of the conditional betas. For each model and time period  $t + 1$ , the loss is measured as the sum over the six betas of the squared difference between the realized beta at time  $t + 1$  and its one-step ahead forecast, i.e.,  $\sum_{i=2}^7 (RB_{i,t+1} - \hat{\beta}_{i,t+1|t})^2$ . The MCS test is implemented with 10,000 bootstrap samples and a block length of 5 observations in a universe consisting of the 6 models based on daily returns. The figures reported in the row labeled ‘#>0.05’ correspond to the number of times (out of 16) that each model belongs to the MCS at the 5% level. The total number of one-step-ahead forecasts is 3,278 for all series but CAT, DIS, GE, JPM and KO for which there are 1,767 forecasts.

better than the DCB model and models with constant parameters, and slightly worse than univariate ARFIMA models fitted on the realized betas obtained from 5-minute returns.

For future research, it will be valuable to study the properties of an ACB model of higher orders to allow for time-varying betas with richer dynamics. Empirically, this extension is not difficult to achieve. In theory, however, there are some challenges to overcome given the nonlinearity of the updating equation for the betas. As such, this model may require more involved and complex derivations for invertibility and ultimately for consistency and asymptotic normality of the estimators.

## APPENDIX

The assumptions are collected in this appendix. We start with the conditions on the regressors. Let  $\boldsymbol{\epsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{p,t})^\top$ . Assuming that

- A1:**  $(\boldsymbol{\epsilon}_t)$  is an *i.i.d.* sequence,  $E\varepsilon_{i,t} = 0$ ,  $E\varepsilon_{i,t}^2 = 1$ ,  $\mu_{0i} \in \mathbb{R}$ ,  $\omega_{0i} > 0$ ,  $\alpha_{0i} > 0$ ,  $\beta_{0i} \geq 0$  and  $E \log \left( \alpha_{0i} \varepsilon_{i,t}^2 + \beta_{0i} \right) < 0$  for  $i = 1, \dots, p$ .

it is well known that there exists a stationary solution  $\boldsymbol{x}_t = (x_{1t}, \dots, x_{pt})^\top$  of the GARCH equations (8)–(9). We do not need to specify a DGP for the exogenous variables, but we assume that

- A2:**  $(\boldsymbol{\eta}_t, \boldsymbol{x}_t^\top, \boldsymbol{z}_t^\top)$  is stationary and ergodic and  $E\|\boldsymbol{z}_t\|^s < \infty$  for some  $s > 0$ .

Furthermore, consider the following assumption.

- A3:**  $(\eta_t)$  is *i.i.d.*,  $E\eta_t = 0$ ,  $E\eta_t^2 = 1$ ,  $\omega_0 > 0$ ,  $\alpha_0 > 0$ ,  $\beta_0 \geq 0$ ,  $E \log(\alpha_0 \eta_1^2 + \beta_0) < 0$  and  $|c_{0i}| < 1$  for  $i = 1, \dots, p$ .

Note that the assumptions  $\omega_0 > 0$  and  $\alpha_0 > 0$  are generally made to avoid degenerate GARCH volatilities. To study the stability of the SRE (14), assume the following.

- A4:** (a) for all  $\boldsymbol{\varphi} \in \Theta_\varphi$ ,  $0 \leq \beta < 1$  and  $0 \leq \beta_i < 1$  for  $i = 1, \dots, p$ ;  
 (b) there exists an operator norm  $\|\cdot\|$  and an integer  $k \geq 1$  such that

$$E \log \sup_{\boldsymbol{\varphi} \in \Theta_\varphi} \left\| \prod_{i=1}^k \Lambda_{t-i}(\boldsymbol{\varphi}) \right\| < 0,$$

where  $\Lambda_t(\boldsymbol{\varphi})$  is the  $p \times p$  matrix whose entries are  $-\frac{\xi_i x_{i,t} x_{j,t}}{\mu_i^2 + g_{i,t}^2(\boldsymbol{\varphi})} + c_i 1_{\{i=j\}}$ .

**EXAMPLE 2** (ARMA(1,1) continued). *In Example 1, Assumption A4 (b) is reduced to  $\sup_{\boldsymbol{\varphi} \in \Theta_\varphi} |c_1 - \xi_1| < 1$ , which is the well-known invertibility condition of the ARMA(1,1) model.*

For the QMLE of the regressor GARCH models to be well defined, strongly consistent and asymptotically normal, we make use of the following assumption (see, e.g., Francq and Zakoian, 2019).

- A5:** (a) For  $i \in \{1, \dots, p\}$ , the support of the distribution of  $\varepsilon_{i,t}$  conditionally on  $\mathcal{F}_{t-1}$  contains at least 3 points. Moreover,  $\Theta$  is a compact subset of  $\mathbb{R} \times (0, \infty) \times (0, \infty) \times [0, 1)$ , which contains  $\boldsymbol{\theta}_0^{(i)}$ .  
 (b) For  $i \in \{1, \dots, p\}$ ,  $\kappa_i := E\varepsilon_{i,1}^4 < \infty$ , and  $\boldsymbol{\theta}_0^{(i)}$  belongs to the interior of  $\Theta$ .

Under **A1** and **A5**, we have the Bahadur representation

$$\sqrt{n} \left( \widehat{\boldsymbol{\theta}}_n^{(i)} - \boldsymbol{\theta}_0^{(i)} \right) \stackrel{op(1)}{=} \left\{ \boldsymbol{J}^{(i)} \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \frac{\varepsilon_{i,t}^2 - 1}{g_{i,t}^2(\boldsymbol{\theta}_0^{(i)})} \frac{\partial g_{i,t}^2(\boldsymbol{\theta}_0^{(i)})}{\partial \boldsymbol{\tau}} + \frac{2\varepsilon_{i,t}}{g_{i,t}(\boldsymbol{\theta}_0^{(i)})} \boldsymbol{e}_1 \right\}, \quad (34)$$

where  $\mathbf{e}_1 = (1, 0, 0, 0)^\top$  and  $\mathbf{J}^{(i)} = \lim_{n \rightarrow \infty} \partial^2 \tilde{O}_n^{(i)}(\hat{\boldsymbol{\theta}}_n^{(i)}) / \partial \boldsymbol{\theta}^{(i)} \partial \boldsymbol{\theta}^{(i)\top}$  a.s.

For the consistency and asymptotic normality of both the one-step and multistep estimators, we shall make use of the following assumption.

- A6:** (a) The support of the distribution of  $\eta_1$  contains at least 3 points. Moreover,  $\Theta_\vartheta$  is a compact set that contains  $\boldsymbol{\vartheta}_0$ .  
(b)  $\boldsymbol{\vartheta}_0$  belongs to the interior of  $\Theta_\vartheta$ .

Additionally, the next assumption frames the regression model as a model of the conditional expectation of  $y_t$  given  $x_t$  and  $\mathcal{F}_{t-1}$ . In particular, together with **A1-A3**, this assumption entails  $E(y_t | \mathbf{x}_t, \mathcal{F}_{t-1}) = \sum_{i=1}^p \beta_{i,t} x_{i,t}$ .

- A7:**  $\eta_t$  is independent of the sigma field  $\mathcal{F}_{t-1}^+$  generated by  $\mathcal{F}_{t-1}$  and  $\mathbf{x}_t$ .

We further impose an identifiability assumption requiring that if  $\boldsymbol{\lambda}_t^\top \mathbf{x}_t = 0$  a.s. for some  $\boldsymbol{\lambda}_t \in \mathcal{F}_{t-1}$ , then  $\boldsymbol{\lambda}_t = \mathbf{0}$ .

- A8:** For  $i \in \{1, \dots, p\}$ , the conditional distribution of  $\varepsilon_{i,t}$  given  $\mathcal{F}_{t-1}$  and  $\{\varepsilon_{j,t}, j \neq i\}$  is not degenerate.

For the exogenous variables  $\mathbf{z}_t$  entering the updating equation of the time-varying betas, we do not need any kind of strong exogeneity assumption. Specifically, conditions such as the independence between  $\mathbf{z}_t$  and  $(\boldsymbol{\varepsilon}_t^\top, \eta_t)$  are not necessary for the results that follow. For identifiability reasons, we do, however, assume that  $\mathbf{z}_t$  does not contain all the extra information conveyed by  $\eta_t$  or  $\boldsymbol{\varepsilon}_t$ . More precisely, we impose the following assumption.

- A9:** (a)  $\eta_t \notin \mathcal{F}_{t-1}^{++}$ , where  $\mathcal{F}_{t-1}^{++}$  denotes the sigma field generated by  $\mathcal{F}_{t-1}$ ,  $\mathbf{x}_t$  and  $\mathbf{z}_t$ ;  
(b) for  $i \in \{1, \dots, p\}$ ,  $\varepsilon_{i,t}$  does not belong to the sigma field generated by  $\mathcal{F}_{t-1}$  and  $\mathbf{z}_t$ ;  
(c) for  $i \in \{1, \dots, q\}$ ,  $z_{i,t}$  does not belong to the sigma field generated by  $\mathcal{F}_{t-1}$  and  $\{z_{j,t}, j \neq i\}$ .

Note that **A9** (c) precludes multicollinearity of the exogenous variables.

We now give conditions for the one-step and the multistep QML estimators to be  $\sqrt{n}$ -consistent and asymptotically normal and derive their asymptotic covariance matrix. Note that we demonstrated the consistency by assuming, for  $\mathbf{Y}_t(\boldsymbol{\varphi})$ , marginal moments only. For nonstandard GARCH (see *e.g.*, Example (3.3) in Francq and Zakoian, 2004) or multivariate GARCH (see Avarucci et al., 2013), it is often necessary to have high-order finite moments to show asymptotic normality and for information matrices to exist. We thus assume the following.

- A10:** There exist a neighborhood  $V(\boldsymbol{\varphi}_0)$  of  $\boldsymbol{\varphi}_0$  and conjugate numbers; that is,  $p_i \geq 1$

and  $q_i \geq 1$  satisfying  $1/p_i + 1/q_i = 1$  for  $i = 1, \dots, 4$  such that

$$\begin{aligned}
& \left\| \sup_{\varphi \in V(\varphi_0)} \frac{v_t^2(\varphi)}{g_t^2(\varphi)} \right\|_{p_1} < \infty, & \left\| \sup_{\varphi \in V(\varphi_0)} \left\| \frac{1}{g_t^2(\varphi)} \frac{\partial g_t^2(\varphi)}{\partial \varphi} \right\| \right\|_{2q_1} < \infty, \\
& \left\| \sup_{\varphi \in V(\varphi_0)} \frac{v_t^2(\varphi)}{g_t^2(\varphi)} \right\|_{p_2} < \infty, & \left\| \sup_{\varphi \in V(\varphi_0)} \left\| \frac{1}{g_t^2(\varphi)} \frac{\partial^2 g_t^2(\varphi)}{\partial \varphi \partial \varphi^\top} \right\| \right\|_{q_2} < \infty, \\
& \left\| \sup_{\varphi \in V(\varphi_0)} \frac{|v_t(\varphi)|}{g_t(\varphi)} \right\|_{p_3} < \infty, & \left\| \sup_{\varphi \in V(\varphi_0)} \left\| \frac{1}{g_t(\varphi)} \frac{\partial^2 v_t^2(\varphi)}{\partial \varphi \partial \varphi^\top} \right\| \right\|_{q_3} < \infty, \\
& \left\| \sup_{\varphi \in V(\varphi_0)} \frac{|v_t(\varphi)|}{g_t(\varphi)} \right\|_{p_4} < \infty, & \left\| \sup_{\varphi \in V(\varphi_0)} \left\| \frac{1}{g_t^2(\varphi)} \frac{\partial g_t^2(\varphi)}{\partial \varphi} \frac{1}{g_t(\varphi)} \frac{\partial v_t(\varphi)}{\partial \varphi^\top} \right\| \right\|_{q_4} < \infty, \\
& \left\| \sup_{\varphi \in V(\varphi_0)} \frac{1}{g_t(\varphi)} \left\| \frac{\partial v_t(\varphi)}{\partial \varphi^\top} \right\| \right\|_2 < \infty.
\end{aligned}$$

Note that by (4.26) and (4.29) in Francq and Zakoian (2004), it can be seen that (the analog of) **A10** is always satisfied in the case of standard GARCH( $p, q$ ) models that do not contain zero coefficients. The following example shows that this is not the case in our framework.

**EXAMPLE 3** (Necessity of moment conditions). *For simplicity, consider the case where  $p = 1$  and  $q = 0$  (i.e., there exist no exogenous variables in the conditional beta equations). Assume also that  $\mu_1 = 0$  and  $g_{1t}^2(\varphi) = \omega_1$ . SRE (14) can be written as*

$$\beta_{t+1}(\varphi) = \varpi + \xi \frac{y_t x_t}{\omega_1} + \left( c - \frac{\xi x_t^2}{\omega_1} \right) \beta_t(\varphi)$$

and thus,

$$\frac{\partial \beta_{t+1}(\varphi)}{\partial \xi} = \frac{v_t(\varphi) x_t}{\omega_1} + \left( c - \frac{\xi x_t^2}{\omega_1} \right) \frac{\partial \beta_t(\varphi)}{\partial \xi} = \sum_{i=0}^{\infty} \left( \frac{v_{t-i}(\varphi) x_{t-i}}{\omega_1} \right) \prod_{j=0}^{i-1} \left( c - \frac{\xi x_{t-j}^2}{\omega_1} \right),$$

with the convention  $\prod_{j=0}^{-1} \cdot = 1$  and under the condition that  $E \log \left| c - \frac{\xi \omega_{01} \varepsilon_t^2}{\omega_1} \right| < 0$ . If we assume that  $(\eta_t)$  and  $(\varepsilon_t)$  are independent, then  $v_t$  and  $x_s = \sqrt{\omega_{01}} \varepsilon_s$  are also independent. Then,

$$E \left| \frac{\partial \beta_{t+1}(\varphi_0)}{\partial \xi} \right|^r = \frac{E |v_1|^r E |\varepsilon_1|^r}{|\omega_{01}|^{r/2}} \frac{1}{1 - E |c_0 - \xi_0 \varepsilon_1^2|^r} < \infty$$

if and only if  $E |c_0 - \xi_0 \varepsilon_1^2|^r < 1$  and  $E |\alpha_0 \eta_1^2 + \beta_0|^r < 1$  (the latter condition is equivalent to  $E |v_1|^r < \infty$ ). Now, assuming for further simplicity that  $\beta = 0$  (i.e., that  $(v_t)$  follows ARCH(1)), we also have

$$\frac{1}{g_t^2(\varphi_0)} \frac{\partial g_t^2(\varphi_0)}{\partial \xi} = \frac{-2v_{t-1}x_{t-1}}{\omega + \alpha v_{t-1}^2} \frac{\partial \beta_{t-1}(\varphi_0)}{\partial \xi},$$

which shows that the second moment condition in **A10** requires  $E|c_0 - \xi_0 \varepsilon_1^2|^{2q_1} < 1$  and  $E|\alpha_0 \eta_1^2 + \beta_0|^{2q_1} < 1$ . Note also that for the existence of the information matrices  $\mathbf{I}_{1S}$  and  $\mathbf{J}_{1S}$  below, a moment of order 2 is needed for  $\frac{1}{g_t^2(\varphi_0)} \frac{\partial g_t^2(\varphi_0)}{\partial \xi}$ . This shows that the constraints  $E|c_0 - \xi_0 \varepsilon_1^2|^2 < 1$  and  $E|\alpha_0 \eta_1^2 + \beta_0|^2 < 1$  are needed for the asymptotic variances of the QML estimators to be well defined on this example.

## Acknowledgements

The authors gratefully acknowledge the editor, Torben Andersen, the associate editor and two anonymous referees for very useful comments on a previous version of this paper. We also thank Tim Bollerslev, Peter Reinhard Hansen, Andrew Harvey, Ilze Kalnina, Siem Jan Koopman and Julie Schnaitmann for helpful comments and discussions. Christian acknowledges the research support of the labex ECODEC while Sébastien acknowledges the research support of the French National Research Agency Grant ANR-17-EURE-0020 and the Excellence Initiative of Aix-Marseille University - A\*MIDEX. Christian and Sébastien also acknowledge research support by the French National Research Agency Grant ANR-21-CE26-0007-01 for their project MLEforRisk. Francisco Blasques acknowledges financial support of the Dutch Science Foundation (NWO) under the grant Vidi.195.099.

## References

- Ait-Sahalia, Y., Kalnina, I., Xiu, D., 2020. High-frequency factor models and regressions. *Journal of Econometrics* 216, 86–105.
- Amendola, A., Francq, C., 2009. Concepts and tools for nonlinear time series modelling, in: Belsley, D., Kontoghiorghes, E. (Eds.), *Handbook of Computational Econometrics*. Wiley. chapter 10, pp. 377–427.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Wu, G., 2006. Realized beta: Persistence and predictability, in: Fomby, T., Terrell, D. (Eds.), *Econometric Analysis of Financial and Economic Time Series*. Emerald Group Publishing Limited, Bingley. volume 20 of *Advances in Econometrics*, pp. 1–39.
- Andersen, T.G., Thyrgaard, M., Todorov, V., 2021. Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk. *Quantitative Economics* 12, 647–682.
- Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.
- Avarucci, M., Beutner, E., Zaffaroni, P., 2013. On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models. *Econometric Theory* 29, 545–566.

- Bai, J., 1996. Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica* 64, 597–622.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.
- Barndorff-Nielsen, O.E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72, 885–925.
- Bazzi, M., Blasques, F., Koopman, S.J., Lucas, A., 2017. Time varying transition probabilities for markov regime switching models. *Journal of Business and Economic Statistics* 38, 458–478.
- Blasques, F., Francq, C., Laurent, S., 2023. Quasi score-driven models. *Journal of Econometrics* 234, 251–275.
- Blasques, F., Gorgi, P., Koopman, S.J., Wintenberger, O., 2018a. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* 12, 1019–1052.
- Blasques, F., Koopman, S., Lasak, K., Lucas, A., 2016. In-sample bounds for time-varying parameters of observation driven models. *International Journal of Forecasting* 32, 875–887.
- Blasques, F., Koopman, S., Lucas, A., 2018b. Amendments and corrections: ‘information-theoretic optimality of observation-driven time series models for continuous responses’. *Biometrika* 105, 753.
- Blasques, F., Koopman, S.J., Lucas, A., 2015. Information theoretic optimality of observation driven time series models. *Biometrika* 102, 325–343.
- Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96, 116–131.
- Boudt, K., Lunde, A., Laurent, S., Quaedvlieg, R., Sauri, O., 2017. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics* 196, 347–367.
- Bougerol, P., 1993. Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31, 942–959.
- Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)* 37, 149–192.
- Carhart, M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.

- Creal, D., Koopman, S., A., L., 2012. A general framework for observation driven time-varying parameter models. *Journal of Applied Econometrics* 28, 777–795.
- Creal, D., Koopman, S.J., Lucas, A., 2010. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. Working paper.
- Creal, D., Koopman, S.J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* 29, 552–563.
- Darolles, S., Francq, C., Laurent, S., 2018. Asymptotics of cholesky GARCH models and time-varying conditional betas. *Journal of Econometrics* 204, 223–247.
- Doornik, J., 2012. Object-oriented matrix programming using Ox. Timberlake Consultants Press.
- Doornik, J., Koopman, S.J., Shephard, N., 2008. *Statistical Algorithms for Models in State Space Form: SsfPack 3.0*. Timberlake Consultants Press.
- Doornik, J.A., Ooms, M., 2021. A package for estimating, forecasting and simulating arfima models: Arfima. Package manual.
- Durbin, J., Koopman, S.J., 2001. *Time Series Analysis by State Space Methods (Oxford Statistical Science Series)*. Oxford University Press.
- Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20, 339–350.
- Engle, R.F., 2016. Dynamic conditional beta. *Journal of Financial Econometrics* 14, 643–667.
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1–22.
- Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy* 81, 607–636.
- Fernández, C., Steel, M., 1998. On bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* 93, 359–371.
- Francq, C., Zakoian, J.M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Francq, C., Zakoian, J.M., 2019. *GARCH Models - Structure, Estimation and Finance Applications*. John Wiley.

- Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity data sets. *Econometrica* 84, 985–1046.
- Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. *Journal of Applied Econometrics* 18, 641–663.
- Grassi, S., Violante, F., 2022. Asset pricing using block-cholesky garch and time-varying betas. Working paper.
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press.
- Hansen, B.E., 1992. Testing for parameter instability in linear models. *Journal of Policy Modeling* 14, 517–533.
- Hansen, P., Lunde, A., Laurent, S., 2021. *Econometric Toolkit for Multiple Comparisons: Mulcum 3.0*. Timberlake Consultants Press.
- Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. *Econometrica* 79, 453–497.
- Harvey, A.A., Chakravarty, T., 2008. Beta-t(E)GARCH. Discussion Paper 08340. University of Cambridge.
- Koundouri, P., Kourogenis, N., Pittis, N., Samartzis, P., 2016. Factor models of stock returns: Garch errors versus time-varying betas. *Journal of Forecasting* 36, 445–461.
- Lange, R.J., van Os, B., van Dijk, D., 2022. Robust observation-driven models using proximal-parameter updates. Technical Report. Tinbergen Institute Discussion Paper 2022-066/III.
- Laurent, S., 2018. G@RCH 8. Estimating and Forecasting GARCH Models. Timberlake Consultants Ltd.
- Nyblom, J., 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.
- Patton, A.J., Verardo, M., 2012. Does beta move with news? Firm-specific information flows and learning about profitability. *The Review of Financial Studies* 25, 2789–2839.
- Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34, 2449–2495.
- Wintenberger, O., 2013. Continuous invertibility and stable qml estimation of the egarch (1, 1) model. *Scandinavian Journal of Statistics* 40, 846–867.