Contemporaneous Aggregation of GARCH Models and Evaluation of the Aggregation Bias

Eric Jondeau[a]

First draft: February 2008

Abstract

It is well known that the class of strong (Generalized) AutoRegressive Conditional Heteroskedasticity (or GARCH) processes is not closed under contemporaneous aggregation. This paper provides the dynamics followed by the aggregate process when the individual persistence parameters are drawn from the same (unknown) distribution. Assuming heterogeneity across individual parameters, the dynamics of the aggregate volatility involves additional lags that reflect the moments of the distribution of the individual persistence parameters. Then the paper describes a consistent estimator of the aggregate process, based on nonlinear least squares. A simulation study reveals that this aggregation-corrected estimator performs very well under realistic sets of parameters. Last, this approach is extended to a multi-sector context. This extension is used to evaluate the importance of the aggregation bias. Using size and book-to-market portfolios, I show that the investor is willing to pay one fifth of her expected return to switch from the standard GARCH(1,1) estimator to the aggregation-corrected estimator.

Keywords: Contemporaneous aggregation, Heterogeneity, Volatility, GARCH model.
JEL classification: C13, C21, G11.

[a]Swiss Finance Institute and University of Lausanne, Faculty of Business and Economics, CH 1015 Lausanne, Switzerland. E-mail: eric.jondeau@unil.ch.

Financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK IP D2 "Financial Econometrics for Risk Management") is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

Since (Generalized) AutoRegressive Conditional Heteroskedasticity (or GARCH) models have been introduced in finance (Engle, 1982, Bollerslev, 1986), they are commonly used to model the volatility of financial returns, such as stock returns, interest rates, or exchange rates. Multivariate extensions have been proposed that allow to model several volatility processes simultaneously. These models are now regularly used as key tools for asset and risk management. In these areas, a particularly pregnant issue is the level of aggregation that should be used for modeling the dynamics of portfolio returns. Two alternative approaches are readily available (see Andersen et al., 2005): on the one hand, the asset-level approach requires the estimation of the joint behavior of all the assets in the portfolio, thus allowing to capture the interactions between asset returns and to evaluate their implications for portfolio return. In the case of large portfolios, however, the computational burden may render this approach simply infeasible. On the other hand, the portfolio-level approach only requires the modeling of the portfolio return, but it is often inappropriate for scenario analysis. This approach is in general preferred for computational reasons. Another instance where the portfolio-level approach is naturally adopted is the modeling of the dynamics of sectoral indices, asset classes, or risk factors, which are themselves constructed as portfolios. In such cases, it is often impossible to identify the underlying individual assets, so that the properties of the resulting portfolio can be inferred from the aggregate dynamics only.

An important, yet often neglected, issue raised by the portfolio-level approach is that GARCH models are not closed under aggregation. In a well-known paper, Nijman and Sentana (1996) investigate the aggregation of two independent assets with volatility processes characterized by the same level of persistence. They show that the aggregate portfolio does not share the same properties as the individual assets, meaning that the strong GARCH model is not closed under aggregation. The resulting process, namely the weak GARCH model, has been precisely studied by Drost and Nijman (1993), Nijman and Sentana (1996), and Meddahi and Renault (2004). Some issues raised by its estimation have been addressed by Francq and Zakoian (2000) and Komunjer (2001). One particularly striking result is that even in the overly simple case of two independent assets with the same volatility persistence, standard estimation techniques fail to provide consistent estimates of the dynamics of the observed aggregate process (see Nijman and Sentana, 1996, and Komunjer, 2001).

In this paper, I investigate the dynamic properties of a portfolio composed of several assets, whose volatilities are driven by GARCH processes. More precisely, the individual assets have the following characteristics: (1) each asset is driven by a GARCH$(1,1)$ variance process, (2) the assets are possibly correlated, with covariances driven by a GARCH-type process, (3) the parameters driving the persistence of the variance and covariance processes are drawn from the same probability distribution function. The first contribution of the paper is to show that aggregate squared returns follow an ARMA-type process, reflecting the weak GARCH nature of the aggregate returns. In principle, an infinite number of lags is required to fully capture the effect of the aggregation of heterogenous processes on the dynamics of portfolio squared returns. In addition, parameters pertaining to lags of squared returns can be interpreted in terms of moments of the cross-section distribution of the volatility persistence parameters.[1]

Given the resulting dynamics of the aggregate volatility, it is clear that Quasi Maximum Likelihood Estimator (QMLE) (for strong GARCH$(1,1)$ process) and even the Least-Square Estimator (LSE) of Francq and Zakoian (2000) (for weak GARCH$(1,1)$ process) will be asymptotically biased. As a second contribution, the paper provides a consistent estimator of the aggregate volatility dynamics. This estimator takes advantage of the relation between the parameters of the volatility dynamics and the moments of the cross-section distribution of the volatility persistence to reduce the estimation burden. In the spirit of Francq and Zakoian (2000), I define the Aggregation-Corrected Estimator (ACE), which alleviates problems related to the mis-specification of the conditional volatility. Using Monte-Carlo simulations, this estimator is shown to perform very well in finite sample, for realistic parameterizations. This methodology takes advantage of both portfolio-level and asset-level approaches. On the one hand, an unbiased estimate of the portfolio return dynamics is obtained using aggregate data only. On the other hand, the estimation of the aggregate process reveals the characteristics of the cross-section distribution of the individual persistence parameters.

The third contribution of the paper consists in an extension of the model to a multi-sector set-up. When the investment opportunity set is composed of industry

[1]Some issues raised by the contemporaneous aggregation of GARCH processes are clearly related to those raised by the temporal aggregation of such processes. The latter problem has been addressed by Drost and Nijman (1993), Meddahi and Renault (2004), and more recently by Hafner (2008). One major difference between the two problems is that, under contemporaneous aggregation, one has to allow some heterogeneity across individual parameters. This clearly complicates the analysis since the resulting aggregate process will reflect this underlying heterogeneity.

indices, asset classes (such as stock and bond indices for a given country), or risk factors, the aggregation bias is likely to affect the variances of all the components of the portfolio as well as their covariances. First I describe a multi-sector model that accounts for the aggregation bias. The aggregation-corrected process for cross-covariances is similar to the one obtained for sectoral variances. Then, the multi-sector model is used to estimate the joint dynamics of the six size and book-to-market portfolios, as constructed by Fama and French (1995). This data is used to evaluate the consequences of the aggregation bias on optimal asset allocation. This bias is shown to have a dramatic effect on the volatility dynamics and on the optimal portfolio weights, resulting in a large opportunity cost of using the suboptimal QMLE of the strong GARCH$(1,1)$ process instead of the proposed ACE.

This paper is related to the recent work of Zaffaroni (2007), who investigates some aspects of the contemporaneous aggregation of GARCH processes in the context of an infinite number of assets. He provides the asymptotic behavior of the variance of the aggregate process. In particular, he shows that conditional heteroskedasticity is preserved provided the degree of cross-sectional dependence between assets is sufficiently strong. One major difference with that framework is that here the number of assets is not assume to be infinite. This set-up provides a more convenient framework to address issues related to actual asset and risk management.[2]

The organization of the paper is as follows. Section 2 briefly describes the multivariate GARCH model used for individual assets and provides conditions required for the model to be well behaved. Section 3 considers the aggregation of the individual volatility processes. The main results on the dynamics of aggregate squared returns are provided. Section 4 describes a set of estimators designed to correctly estimate the dynamics of aggregate volatility. Section 5 provides some empirical evidence on the distribution of individual parameters for U.S. stocks. Monte-Carlo simulations based on these parameter characteristics show that the estimator proposed in the paper is able to correctly reproduce the behavior of aggregate squared returns. Section 6 extends the model to the multi-sector set-up, compares the forecasting ability of the competing estimation techniques, and provides an evaluation of the aggregation bias for asset allocation. Section 7 concludes. Proofs are relegated in the Appendix.

[2] For instance, with an infinite number of assets, the volatility of the aggregate portfolio is in general driven by the common innovation only. However, in standard asset or risk management, with moderate number of assets, the idiosyncratic innovation is also likely to play a role.

2 The Model for Individual Assets

Let consider the following multivariate model for individual assets. Each asset has a standard strong GARCH$(1,1)$ conditional variance

$$\begin{aligned} \varepsilon_{i,t} &= \sigma_{i,t} z_{i,t}, \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2, \end{aligned} \tag{1}$$

with $i = 1, \cdots, N$, where $\varepsilon_{i,t}$ denotes the unexpected return, $\sigma_{i,t}^2$ the conditional volatility and $z_{i,t}$ the innovation process, with $z_{i,t} \sim iid\ N(0,1)$.[3] Conditional covariances are described in a similar way

$$\sigma_{ij,t} = \omega_{ij} + \alpha_{ij} \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta_{ij} \sigma_{ij,t-1}, \tag{2}$$

with $i, j = 1, \cdots, N$, $j \neq i$. This formulation is a diagonal vec multivariate model as introduced by Bollerslev, Engle, and Wooldridge (1988), in which each element of the covariance matrix $(\sigma_{ij,t})$ depends on the corresponding lagged terms $\sigma_{ij,t-1}$ and $\varepsilon_{i,t-1}\varepsilon_{j,t-1}$. In matrix format, it is written as

$$\Sigma_t = \Omega + A \odot \left(\varepsilon_{t-1} \varepsilon_{t-1}'\right) + B \odot \Sigma_{t-1},$$

where matrices $A = \{\alpha_{ij}\}$, $B = \{\beta_{ij}\}$, and $\Omega = \{\omega_{ij}\}$ are (n,n) positive definite and symmetric matrix, and $\odot$ denotes the Hadamard product.[4]

Equations (1) and (2) can be written more conveniently as follows

$$\begin{aligned} \sigma_{i,t}^2 &= \sigma_i^2 + \alpha_i \left(\varepsilon_{i,t-1}^2 - \sigma_{i,t-1}^2\right) + \gamma_i \left(\sigma_{i,t-1}^2 - \sigma_i^2\right), \\ \sigma_{ij,t} &= \sigma_{ij} + \alpha_{ij} \left(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \sigma_{ij,t-1}\right) + \gamma_{ij} \left(\sigma_{ij,t-1} - \sigma_{ij}\right), \end{aligned}$$

where $\gamma_i = \alpha_i + \beta_i$, $\gamma_{ij} = \alpha_{ij} + \beta_{ij}$, $\sigma_i^2 = \omega_i/(1-\gamma_i)$, and $\sigma_{ij} = \omega_{ij}/(1-\gamma_{ij})$. Parameters γ_i and γ_{ij} denote the persistence of the conditional variance and covariance processes, and σ_i^2 and σ_{ij} denote unconditional variances and covariances respectively.

Positivity of the variance processes requires that $\omega_i > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$. In addition, stationarity of the individual variance and covariance processes requires that $\gamma_i < 1$ and $\gamma_{ij} < 1$. Since variances and covariances are modeled separately, additional restrictions are required to ensure that the covariance matrix Σ_t is positive semi-definite at each date t. Such an issue was already addressed in papers dealing

[3]Normality is assumed here to obtain some analytical results regarding the moments of the innovations on squared returns. This assumption will be relaxed in Section 5.

[4]The Hadamard product defines the element-wise product of two matrices, $\{A \odot B\}_{ij} = A_{ij} B_{ij}$.

with the decentralized estimation of a covariance matrix.Conditions ensuring that the conditional covariance matrix Σ_t is positive semi-definite have been established by Ledoit, Santa-Clara, and Wolf (2003). For convenience, these conditions are reported in the following proposition.

Proposition 1 *Sufficient conditions for positive semi-definite covariance matrix are that* $\Omega \div (1 - B)$*,* A*, and* B *are positive semi-definite and* $\gamma_i < 1$*,* $\forall i = 1, \cdots, N$*, where* 1 *is the* $(N \times N)$ *matrix of ones.*[5]

Proof: See Ledoit, Santa-Clara, and Wolf (2003, p. 738).

As shown by Nijman and Sentana (1996), the contemporaneous aggregation of strong GARCH processes gives rise to the so-called weak GARCH process. In this set-up, σ_t^2 cannot be interpreted as the conditional variance of ε_t, but only as the linear projection of ε_t^2 on $\mathcal{F}_t = \left\{1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \cdots\right\}$. Consequently, one typically focuses on the dynamics of the aggregated squared returns, instead of conditional volatility. For this reason, it is useful to give the expression for individual squared returns and cross-product of returns. Straightforward manipulation of equation (1) gives

$$\varepsilon_{i,t}^2 = \omega_i + \gamma_i \varepsilon_{i,t-1}^2 + v_{i,t} - \beta_i v_{i,t-1},$$

where $v_{i,t} = \varepsilon_{i,t}^2 - \sigma_{i,t}^2 = \sigma_{i,t}^2 (z_{i,t}^2 - 1)$ is the innovation process for squared returns. Similarly, equation (2) gives for cross-products of returns,

$$\varepsilon_{i,t}\varepsilon_{j,t} = \omega_{ij} + \gamma_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + v_{ij,t} - \beta_{ij} v_{ij,t-1},$$

where $v_{ij,t} = \varepsilon_{i,t}\varepsilon_{j,t} - \sigma_{ij,t} = \sigma_{i,t}\sigma_{j,t}\left(z_{i,t}z_{j,t} - \rho_{ij,t}\right)$. I use the notation $\sigma_{ij,t} = \sigma_{i,t}\sigma_{j,t}\rho_{ij,t}$, where $\rho_{ij,t}$ denotes the conditional correlation between assets i and j at date t.

In the following, equations (1) and (2) are assumed to be based on realizations of random variables (ω, α, β). The following assumptions describe the characteristics of the innovation processes and the random coefficients:

Assumption 1 *(1) Innovations* $z_{i,t}$ *are serially uncorrelated normal random variables with* $E[z_{i,t}] = 0$*, and* $V[z_{i,t}] = 1$*. Cross-correlation between* $z_{i,t}$ *and* $z_{j,t}$ *is* $E[z_{i,t}z_{j,t}] = \rho_{ij}$*. (2)* $z_{i,t}$ *and* $(\omega_i, \alpha_i, \gamma_i)$ *are mutually independent.* $(z_{i,t}, z_{j,t})$ *and* $(\omega_{ij}, \alpha_{ij}, \gamma_{ij})$ *are mutually independent.*

[5] $\div$ denotes the element-wise division of two matrices $\{A \div B\}_{ij} = A_{ij}/B_{ij}$.

Assumption 2 *(1)* $0 \leq \omega_i \leq \infty$ *and* $-\infty \leq \omega_{ij} \leq \infty$*. (2)* α_i *and* α_{ij} *are iid realizations of random variable* α*, with* $0 < \alpha < 1$*. (3)* γ_i *and* γ_{ij} *are iid realizations of random variable* γ*, with* $0 < \gamma < 1$*. (4) Realizations of* $(\omega_i, \alpha_i, \gamma_i)$ *and* $(\omega_{ij}, \alpha_{ij}, \gamma_{ij})$ *satisfy the restrictions of Proposition 1.*

Assumption 1 indicates that the time dependency in the model is fully captured by the covariance matrix Σ_t, since innovations $z_{i,t}$ are iid. Innovations can be contemporaneously correlated with constant correlation[6]

$$\rho_{ij} = \frac{\omega_{ij}/(1-\gamma_{ij})}{\sqrt{\omega_i/(1-\gamma_i)\,\omega_j/(1-\gamma_j)}}.$$

Assumption 2 indicates that the parameters α and γ pertaining to the conditional variances and covariances are drawn from the same distribution. For the moment, I do not assume any particular distribution for the random variables α and γ. Assumption 2(4) is required to have a well-behaved multivariate model, with a positive semi-definite covariance matrix at each date t.

Other important characteristics of the model are the properties of the innovation processes $v_{i,t}$ and $v_{ij,t}$.[7] The following proposition holds:

Proposition 2 *(1) Innovations* $v_{i,t}$ *and* $v_{ij,t}$ *are serially uncorrelated random variables with* $E[v_{i,t}] = E[v_{ij,t}] = 0$*,* $V[v_{i,t}] = \sigma^2_{v,i}$*, and* $V[v_{ij,t}] = \sigma^2_{v,ij}$*.*

(2) Provided some restrictions on the parameter set given in the Appendix, the unconditional variances and covariances of the innovation processes $v_{i,t}$ *and* $v_{ij,t}$ *are*

[6] An alternative way to introduce correlation across assets is to assume that innovations $z_{i,t}$ are composed of a common innovation u_t and an idiosyncratic term $\eta_{i,t}$, i.e. $z_{i,t} = u_t + \eta_{i,t}$. The perfect correlation case with $z_{i,t} = u_t$ has been studied by Zaffaroni (2007) for an infinite number of assets.

[7] Komunjer (2001) further investigated the time series properties of $v_{i,t}$ in particular its higher moments, for Gaussian innovations $z_{i,t}$.

given by

$$
\begin{aligned}
E[v_{i,t}^2] &= \sigma_{v,i}^2 = 2E[\sigma_{i,t}^4], \\
E[v_{ij,t}^2] &= \sigma_{v,ij}^2 = E[\sigma_{i,t}^2\sigma_{j,t}^2] + E[\sigma_{ij,t}^2], \\
E[v_{i,t}v_{j,t}] &= 2E[\sigma_{ij,t}^2], \\
E[v_{i,t}v_{ij,t}] &= 2E[\sigma_{i,t}^2\sigma_{ij,t}], \\
E[v_{i,t}v_{jk,t}] &= 2E[\sigma_{ij,t}\sigma_{ik,t}], \\
E[v_{ij,t}v_{ik,t}] &= E[\sigma_{i,t}^2\sigma_{jk,t}] + E[\sigma_{ij,t}\sigma_{ik,t}], \\
E[v_{ij,t}v_{kl,t}] &= E[\sigma_{ik,t}\sigma_{jl,t}] + E[\sigma_{il,t}\sigma_{jk,t}].
\end{aligned}
$$

(3) Innovations $v_{i,t}$ and $v_{ij,t}$ are orthogonal to any random coefficient driving the dynamics of conditional variances and covariances.

Proof: See Appendix 1. All the moments and co-moments of conditional variances and covariances are defined in the Appendix.

This proposition states that the variances of the innovation processes $v_{i,t}$ are fully defined by the set of micro-parameters $(\omega_i, \alpha_i, \gamma_i)$, while the variances of the innovation processes $v_{ij,t}$ and the covariances are defined by the parameters describing the individual processes $(\omega_i, \alpha_i, \gamma_i)$ as well as the ones describing the joint processes $(\omega_{ij}, \alpha_{ij}, \gamma_{ij})$.

3 Aggregation

I now consider the characteristics of a portfolio composed of N risky assets whose dynamics have just been described. There is no riskless asset and short sales are not allowed. Portfolio weights are denoted $w = (w_1, \cdots, w_N)'$, with $w_i \geq 0$ and $\sum_{i=1}^N w_i = 1$.[8] The aggregate unexpected return is defined as $\varepsilon_{p,t} = \sum_{i=1}^N w_i \varepsilon_{i,t}$, i.e.

[8]Portfolio weights are assumed to be constant over time. This assumption loosely corresponds to a strategic allocation problem. The case with time-varying weights may raise new issues, in particular if their dynamics are dependent from the parameters of the individual squared returns. The investigation of this extension is beyond the scope of this paper and left for further research.

the cross-sectional average of the $\varepsilon_{i,t}$'s. Aggregate squared return is then defined as

$$\varepsilon_{p,t}^2 = \left(\sum_{i=1}^N w_i \varepsilon_{i,t} \right)^2 = \sum_{i=1}^N w_i^2 \varepsilon_{i,t}^2 + 2 \sum_{i=1}^N \sum_{j>i}^N w_i w_j \varepsilon_{i,t} \varepsilon_{j,t}.$$

This expression reminds that the properties of aggregate squared returns will be dominated by those of the cross-product of individual returns rather than by the individual squared returns.

In the framework described above and under Assumptions 1 and 2, the following result holds regarding the dynamics of aggregate squared returns.

Proposition 3 *Let* $\{\varepsilon_{i,t}\}$, $i = 1, \cdots, N$, *be a multivariate GARCH*$(1,1)$ *process defined by equations (1) and (2). Under Assumptions 1 and 2, the aggregate process* $\varepsilon_{p,t} = \sum_{i=1}^N w_i \ \varepsilon_{i,t}$ *satisfies*

$$\varepsilon_{p,t}^2 = \sigma_{p,t}^2 + v_{p,t}, \tag{3}$$

$$\sigma_{p,t}^2 = \Omega_p + \sum_{k=1}^\infty \Psi_k \varepsilon_{p,t-k}^2 + \sum_{k=1}^\infty \Phi_k \sigma_{p,t-k}^2, \tag{4}$$

where the innovation process

$$v_{p,t} = \sum_{i=1}^N w_i^2 v_{i,t} + 2 \sum_{i=1}^N \sum_{j>i}^N w_i w_j v_{ij,t} \tag{5}$$

is a white noise with mean zero and variance σ_v^2. *Aggregate squared returns also have the following dynamics*

$$\varepsilon_{p,t}^2 = \Omega_p + \sum_{k=1}^\infty \Lambda_k \varepsilon_{p,t-k}^2 + v_{p,t} - \sum_{k=1}^\infty \Phi_k v_{p,t-k}. \tag{6}$$

Parameters $\Psi_k = E[\psi_k]$, $\Phi_k = E[\phi_k]$, *and* $\Lambda_k = E[\lambda_k]$, $k = 1, \cdots,$ *in equations (4) and (6) are functions of the moments of the distribution of individual parameters:*

$$\begin{array}{lll} \psi_1 = \alpha & \psi_{k+1} = (\lambda_k - \Lambda_k)\,\alpha & k = 1, 2, \cdots \\ \phi_1 = \beta & \phi_{k+1} = (\lambda_k - \Lambda_k)\,\beta & \\ \lambda_1 = \gamma & \lambda_{k+1} = (\lambda_k - \Lambda_k)\,\gamma = \psi_{k+1} + \phi_{k+1}, & \end{array}$$

and σ_v^2 *and* Ω_p *are defined in the Appendix.*

Proof: See Appendix 2.

Several comments are of interest regarding Proposition 3. First, equation (6) clearly shows that the aggregate return is a weak GARCH process with an infinite number of lags. This allows to better understand why estimating a weak GARCH$(1,1)$ process is not enough to correct for heterogeneity across assets. It would provide consistent estimates of the aggregate parameters $(\Omega_p,\ \gamma,\ \beta)$ only if the additional lags Λ_k, $k>1$, were equal to 0, i.e. if the distribution of γ were degenerate to a single value, $E[\gamma]$. Equation (6) simplifies to equation (10) in Nijman and Sentana (1996) when all individual persistence parameters are equal to $E[\gamma]$. In this case, one has $\Lambda_k = \Phi_k = 0$ for all $k>1$.

An interesting extreme case is the Integrated GARCH process, where

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + (1-\alpha_i)\sigma_{i,t-1}^2,$$

or equivalently

$$\varepsilon_{i,t}^2 = \omega_i + \varepsilon_{i,t-1}^2 + v_{i,t} - (1-\alpha_i) v_{i,t-1}.$$

In this case, the aggregate squared return simply writes

$$\varepsilon_{p,t}^2 = \Omega_p + \varepsilon_{p,t-1}^2 + v_{p,t} - (1-E[\alpha]) v_{p,t-1}.$$

This example clearly shows that the additional lags in equation (6) come from the heterogeneity in the persistence parameters.

Second, estimating equation (6) allows to recover the moments of the cross-section distribution of the persistence parameter γ.[9] Contemplating this equation reveals that all the parameters Λ_k pertaining to lags $\varepsilon_{p,t-k}^2$ are related to cross-sectional moments of γ. After straightforward computations, one obtains

$$\Lambda_{k+1} = E\left[\lambda_{k+1}\right] = E\left[\left(\lambda_k - \Lambda_k\right)\gamma\right] = E[\gamma^{k+1}] - \sum_{r=1}^{k} \Lambda_r\ E[\gamma^{k-r+1}]. \tag{7}$$

[9] A similar interpretation has been proposed by Granger (1980) for the process resulting from the aggregation of autoregressive processes or by Lewbel (1994) for the process resulting from the aggregation of linear dynamic processes.

In particular, the four first moments of γ are given by[10]

$$
\begin{aligned}
E[\gamma] &= \Lambda_1, \\
V[\gamma] &= \Lambda_2, \\
sk[\gamma] &= (\Lambda_3 - \Lambda_1\Lambda_2) / (\Lambda_2)^{3/2}, \\
ku[\gamma] &= \left(\Lambda_4 - 2\Lambda_1\Lambda_3 + \Lambda_1^2\Lambda_2 + \Lambda_2^2\right) / (\Lambda_2)^2 .
\end{aligned}
$$

While Φ_1 is the expected value of β, other parameters Φ_k pertaining to lags $v_{p,t-k}$ correspond to cross-sectional co-moments between powers of γ and β. For instance, $\Phi_2 = cov[\gamma, \beta]$. As a consequence, the characteristics of the cross-section distribution of β (or α) cannot be recovered. Last, the aggregate constant term Ω_p summarizes the characteristics of all the individual parameters ω_i and ω_{ij}, but it does not allow to recover any of their characteristics.

Third, equation (6) suggests that there are two extreme cases of particular interest, where the expressions for Φ_k, $k > 1$, simplify. In the first case, γ and α are independent from each other, so that $\Phi_2 = E[\phi_2] = E[(\lambda_1 - \Lambda_1)\beta] = E[(\gamma - \Lambda_1)(\gamma - \alpha)] = V[\gamma] = \Lambda_2$ and more generally $\Phi_k = \Lambda_k$, $\forall k > 1$. In the second case, γ and β are independent from each other, so that $\Phi_k = 0$, $\forall k > 1$. The expressions for the dynamics of aggregate squared returns in these two cases are given in the following corollary:

Corollary 1 *(1) When γ and α are independent from each other, the dynamics of $\varepsilon_{p,t}^2$ is given by*

$$
\begin{aligned}
\varepsilon_{p,t}^2 &= \Omega_p + \sum_{k=1}^{\infty} \Lambda_k \varepsilon_{p,t-k}^2 + v_{p,t} - E[\beta]\, v_{p,t-1} - \sum_{k=2}^{\infty} \Lambda_k v_{p,t-k} && (8) \\
&= \frac{\Omega_p}{1 - \sum_{k=1}^{\infty} \Lambda_k} + v_{p,t} + E[\alpha] \sum_{k=1}^{\infty} E[\gamma^k] v_{p,t-k}. && (9)
\end{aligned}
$$

[10] More generally, the non-central moment conditions are given by

$$
E\left[\gamma^{k+1}\right] = \Lambda_{k+1} + \sum_{r=1}^{k} \Lambda_r E\left[\gamma^{k-r+1}\right].
$$

(2) When γ and β are independent from each other, the dynamics of $\varepsilon_{p,t}^2$ is given by

$$\varepsilon_{p,t}^2 = \Omega_p + \sum_{k=1}^{\infty} \Lambda_k \varepsilon_{p,t-k}^2 + v_{p,t} - E[\beta]\, v_{p,t-1} \tag{10}$$

$$= \frac{\Omega_p}{1 - \sum_{k=1}^{\infty} \Lambda_k} + \sum_{k=0}^{\infty} \left(E[\gamma^k] - E[\beta]\, E[\gamma^{k-1}] \right) v_{p,t-k}. \tag{11}$$

Proof: See Appendix 3.

Equations (9) and (11) describe the dynamics of squared returns when γ is independent of either α or β. In these expressions, the parameters pertaining to the lags are only related to the moments of the persistence parameter γ and to the expected value of α (or β). This is very likely to reduce the estimation burden dramatically, in particular if one adopts a parametric distribution for the persistence parameter (see below). Clearly, whether γ is independent from either α or β is an empirical issue that should be addressed for each particular dataset. If individual parameters are estimated in a first step using individual asset returns, it is possible to decide ex ante which specification is more likely to yield efficient estimates. However, if individual estimates are not available, it is preferable to rest on specification (6).

Last, for estimation purpose, one needs to truncate the infinite sums in equation (6) to estimate the dynamics of aggregate squared returns. It turns out that, even for a rather small number of lags, the innovation $v_{p,t}$ is expected to be close to an iid process. Indeed, when the persistence parameter is large, as it is typically the case in financial applications, the cross-section variance of this parameter ($\Lambda_2 = V[\gamma]$) is likely to be small, and more generally the convergence of the Λ_k parameters toward 0 is expected to be fast. As the subsequent Monte-Carlo simulations will show, the estimation of equation (6) provides essentially unbiased estimates of the mean value of α and γ even for a small number of lags and for a small number of assets.

4 Estimation of the Aggregate Model

In this section, I review some of the issues raised by the estimation of aggregate volatility processes and then propose a new estimation procedure for the case of contemporaneous aggregation. A first issue relies on the somewhat disappointing properties of the QMLE of the strong GARCH$(1,1)$ process, even in the case of two independent individual processes. Following Drost and Nijman (1993), who obtained very good performances of the QMLE under temporal aggregation, Nijman and Sentana (1996) investigated its performances in the case of contemporaneous aggregation. However, Monte-Carlo experiments led these authors to conclude that "the QML estimator is approximately consistent in some cases and clearly inconsistent in others".

The reasons for the inconsistency of the QMLE have been investigated by Komunjer (2001). She showed that three main assumptions required for consistency and asymptotic normality of the QMLE are not met in weak GARCH models: (1) the conditional variance of $\varepsilon_{p,t}^2$ is not correctly specified, since $\sigma_{p,t}^2$ is no longer the conditional expectation of $\varepsilon_{p,t}^2$ given $(1, \varepsilon_{p,t-1}, \varepsilon_{p,t-2}, \cdots)$ as in the strong GARCH process; (2) the standardized innovation $\varepsilon_{p,t}/\sigma_{p,t}$ is not iid; and (3) the high moments (beyond mean and variance) of the variance innovation $v_{p,t}$ may not exist for some parameter sets. Komunjer (2001) proposed a new QMLE for weak GARCH processes that accounts for the deficiencies of the standard estimator in this context.[11] Unfortunately, the large-sample properties of this new estimator are still quite disappointing. For instance, for the set of parameters $(\alpha, \beta) = (0.145, 0.705)$ which is close to the parameters found in empirical estimates, the standard Gaussian QMLE is $(0.072, 0.758)$ even with $T = 5,000$ observations, indicating a severe under-estimation of α, similar to the one obtained with the Nijman and Sentana (1996) approach. Using an alternative distribution allowing for fat tails does not reduce these biases. Clearly, assuming a more general aggregation problem with several processes and random coefficients is very likely to worsen the properties of the QMLE, even with large samples. Our simulation experiment, reported in the next section, shows that this is actually the case.

Another approach has been followed by Francq and Zakoian (2000) for the estimation of a weak GARCH(1,1) process. These authors explicitly acknowledge the possible mis-specification of the first two moments and therefore do not rely on QMLE. Instead they define the least-square estimator as the set of parameters that minimizes the sum of squared residuals in the aggregate squared return dynamics. In the context of this paper, it is defined as follows.

Definition 1 *The Least-Square estimator (LSE), denoted $\theta_{LS} = (\Omega_p, \Lambda_1, \Phi_1)'$, is defined as*

$$\theta_{LS} \in \arg\min \frac{1}{T} \sum_{t=1}^{T} v_{p,t}^2,$$

[11] This procedure lies on an innovation algorithm that generates estimates of the variance innovation $\{v_{p,t}\}$ and then computes the QMLE under the Gaussian assumption.

where

$$v_{p,t} = \varepsilon_{p,t}^2 - \Omega_p - \Lambda_1 \varepsilon_{p,t-1}^2 + \Phi_1 v_{p,t-1},$$

with $\Lambda_1 = E[\gamma]$ *and* $\Phi_1 = E[\beta]$.

This estimator is consistent and asymptotically normal only under homogeneity of the persistence parameters across assets, i.e. $\gamma = E[\gamma]$ and $V[\gamma] = 0$.

In the case of contemporaneous aggregation of heterogeneous processes, the least-square estimator has to be adapted to capture the specific features of the resulting weak GARCH process. As shown in Proposition 3, under parameter heterogeneity, aggregate squared returns have an infinite ARMA representation. However, estimating equation (6) with a large number of lags and unrestricted parameters would clearly be inefficient, since the sequence of parameters Λ_k is known to be directly related to the moments of the cross-section distribution of the persistence parameter γ. To take advantage of these relations, I adopt a flexible parametric representation for the cross-sectional distribution. Following Granger (1980), Gonçalves and Gourieroux (1988), and Zaffaroni (2007), I assume that the persistence parameters γ_i and γ_{ij} are independently drawn from a Beta distribution

$$f(\gamma) = \frac{\gamma^{p-1}(1-\gamma)^{q-1}}{B(p,q)},$$

where $p, q \in (0, \infty)$ and $B(p,q)$ is the beta function. This parametric distribution covers the range of values $[0,1]$ that is the admissible interval for γ. The Beta distribution is able to reproduce a wide range of distribution shapes. Two shapes are of particular interest for the persistence parameter in GARCH models. For $0 < q < 1 < p$, the distribution is continuously increasing. For $1 < q < p$, the distribution is leftward asymmetric bell-shaped. The non-central moments of the Beta distribution with parameters p and q are given by

$$E\left[\gamma^k\right] = \frac{B(p+k,q)}{B(p,q)} = \frac{\Gamma(p+q)\,\Gamma(p+k)}{\Gamma(p+q+k)\,\Gamma(p)}, \tag{12}$$

where $\Gamma(p)$ is the gamma function. For any values of p and q, one can compute non-central moments of γ using equation (12) and directly deduce the Λ_k parameters using equation (7), so that one obtains all the autoregressive parameters required in equation (6).

On the other hand, the sequence of aggregate parameters Φ_k is related to the moments of the joint distribution of γ and β. Restricting the shape of this sequence would require additional assumptions on the dependence between γ and β. For instance, assuming full dependence or independence would lead to the dynamics of squared returns summarized in Corollary 1. To avoid such additional assumptions, I directly estimate the expected value of β, $\Phi_0 = E[\beta]$, and the sequence of co-moments, Φ_k, as free parameters. This approach avoids to define a specific joint distribution for γ and β. Given that $\Lambda_k = \Phi_k + \Psi_k$, the only restriction imposed on parameters Φ_k is that $\Phi_k \in [-|\Lambda_k|;|\Lambda_k|]$. This restriction means that the co-moments between γ and either α and β are non-negative. It ensures that the aggregate squared returns defined in equation (6) are positive for each date t. The definition of the resulting Aggregation-Corrected Estimator (ACE) is given below.

Definition 2 *The Aggregation-Corrected Estimator (ACE), denoted $\theta_{AC} = (\Omega_p,\ p,\ q,\ \Phi_1,\ \cdots,\ \Phi_{K_\Phi})'$, is defined as*

$$\theta_{AC} \in \arg\min \frac{1}{T}\sum_{t=1}^{T} \upsilon_{p,t}^2,$$

where

$$\upsilon_{p,t} = \varepsilon_{p,t}^2 - \Omega_p - \sum_{k=1}^{K_\Lambda} \Lambda_k \varepsilon_{p,t-k}^2 + \sum_{k=1}^{K_\Phi} \Phi_k \upsilon_{p,t-k}, \tag{13}$$

with K_Λ and K_Φ the number of lags for Λ_k and Φ_k respectively. The expressions for Λ_k, Φ_k, and Ω_p are given in Proposition 3.

As argued before, when γ is independent from either α or β , the dynamics of $\varepsilon_{p,t}^2$ simplifies as given by equations (8) and (10) in Corollary 1. This gives rise to the following definitions of the Constrained Aggregation-Corrected Estimator (CACE).

Definition 3 *(1) When γ and α are independent from each other, the estimator is constrained by $\Phi_1 = E[\beta]$ and $\Phi_k = \Lambda_k$, $\forall k > 1$, so that the Constrained Aggregation-Corrected Estimator (CACE), denoted $\theta_{CAC}^{(1)} = (\Omega_p, p, q, \Phi_1)'$, is defined as*

$$\begin{aligned} \upsilon_{p,t} &= \varepsilon_{p,t}^2 - \Omega_p - \sum_{k=1}^{K_\Lambda} \Lambda_k \varepsilon_{p,t-k}^2 + \Phi_1 \upsilon_{p,t-1} + \sum_{k=2}^{K_\Lambda} \Lambda_k \upsilon_{p,t-k} \\ &= \varepsilon_{p,t}^2 - \Omega_p - \sum_{k=1}^{K_\Lambda} \Lambda_k \left(\varepsilon_{p,t-k}^2 - \upsilon_{p,t-k}\right) + E[\alpha]\, \upsilon_{p,t-1}. \end{aligned} \tag{14}$$

with K_Λ the number of lags for Λ_k. The expressions for Λ_k and Ω_p are given in Proposition 3 and Appendix 2 respectively.

(2) When γ and β are supposed to be independent from each other, the estimator is constrained by $\Phi_1 = E[\beta]$ and $\Phi_k = 0$, $\forall k > 1$. In this case, the CACE, denoted $\theta_{CAC}^{(2)} = (\Omega_p, p, q, \Phi_1)'$, is defined as

$$v_{p,t} = \varepsilon_{p,t}^2 - \Omega_p - \sum_{k=1}^{K_\Lambda} \Lambda_k \varepsilon_{p,t-k}^2 + \Phi_1 v_{p,t-1}, \tag{15}$$

with K_Λ the number of lags for Λ_k. The expressions for Λ_k and Ω_p are given in Proposition 3 and Appendix 2 respectively.

Whether γ is more likely to be independent from α or β is an empirical question I will discuss in the next section. Additional information on the practical implementation of these estimators is given in Section 5.2.

5 Performance of the Estimators

In this section, I evaluate the finite-sample properties of the various estimators. For this purpose, I first report some empirical evidence on the characteristics of micro-parameters based on U.S. stocks. Then I perform some Monte-Carlo simulations based on these characteristics to evaluate the relative performances of the various estimators. The QMLE and LSE estimators of the GARCH$(1,1)$ process are shown to be severely biased under aggregation with parameter heterogeneity, given their inability to capture to full dynamics of the aggregate squared return. In contrast, the (unrestricted and constrained) Aggregation-Corrected estimators display much smaller biases.

5.1 Calibration Based on U.S. Equities

In order to get some intuition about the individual parameters, I consider a sample of U.S. companies between 1973 and 2004. As the October 1987 crash was found to affect the estimation of some of the bivariate models, I eventually estimated the models over the 1988-2004 sample.[12] The individual variance parameters $(\omega_i, \alpha_i, \gamma_i)$ are estimated for each of the 66 individual stocks of the sample, and the covariance parameters $(\omega_{ij}, \alpha_{ij}, \gamma_{ij})$ are estimated for each of the 2,145 pairs of stocks.

[12]There are 66 companies belonging to the S&P 100 at the end of 2004 for which prices are available on Datastream over the 1973-2004 period.

I begin with the characteristics of the parameter estimates of GARCH$(1,1)$ processes.[13] **Table 1** reports summary statistics on the estimated parameters of the conditional variance equation (1) estimated by ML. The mean estimates of α_i, β_i, and γ_i are 0.049, 0.938, and 0.987 respectively (Panel A). Since the standard error of the mean estimate is smaller than 0.004 in all cases, these mean values are very precisely estimated. The large value of the persistence parameter γ_i has been often mentioned in the empirical literature. Although the estimated value is generally close to 1, the constraint $\gamma_i < 1$ is never binding, suggesting that volatility of returns is adequately described by a highly persistent, yet stationary, process.

Turning to the estimation of the conditional covariance equation (2), the mean estimates of α_{ij}, β_{ij}, and γ_{ij} are 0.026, 0.948, and 0.974 respectively (Panel B). These numbers reveal that the estimate of α_{ij} is in general slightly smaller than the estimate of α_i. On the opposite, the estimation of β_{ij} is in general slightly higher than the estimate of β_i. All in all, the persistence of conditional covariances is very close to the one of conditional variances.

Figure 1 displays the histogram of these parameters. On the left hand side are reported the parameters of the conditional variances, whereas the right hand side corresponds to the parameters of the conditional covariances. As it appears clearly, the distribution of the variance parameters α_i and β_i are close to, yet more dispersed than, the distribution of the covariance parameters α_{ij} and β_{ij}. The shape of the distributions (as characterized by the skewness and kurtosis) is very similar for the variance and covariance parameters. As already mentioned, the persistence of the variance processes is slightly higher than the persistence of the covariance processes. The other characteristics of both empirical distributions look very similar.

Additional information required for the calibration of the unconditional variances and covariances can be found in Table 1 and in **Figure 2**, which displays the histogram of the constant parameters ω_i and ω_{ij} in the GARCH processes, the unconditional standard deviations σ_i and the unconditional correlations ρ_{ij}. First the mean value of σ_i is 0.021, with a minimum of 0.01 and a maximum of 0.053. These values are used in the next section to calibrate the minimum and maximum values for the variances, $\underline{\sigma}^2$ and $\overline{\sigma}^2$. Another interesting result is the rather small unconditional correlation between stock returns. The mean estimate of ρ_{ij} is 0.181, with a minimum of 0.017 and a maximum of 0.642, with only 2% of the estimated

[13] Variance and covariance parameters are estimated using the flexible GARCH approach developed by Ledoit, Santa-Clara, and Wolf (2003). In some cases, these estimates of $(\omega_{ij}, \alpha_{ij}, \gamma_{ij})$ have been corrected to ensure the positive definiteness of the bivariate covariance matrices.

correlations higher than 0.4. These values are used to calibrate $\underline{\rho}$ and $\overline{\rho}$.

The last issue I address in this section is the dependency between parameters. This issue is important from a simulation perspective, since some assumptions on these dependencies are required to define the experiment. **Table 2** reports the correlation between variance parameters and between covariance parameters. Regarding first the dependency between α, β, and γ, there is a major difference between variance and covariance parameters: for the variance processes, γ is positively correlated with β but negatively correlated with α (0.87 and -0.69, respectively). In contrast, for the covariance processes, γ is again positively correlated with β, but it is uncorrelated with α (0.88 and -0.05, respectively). This suggests that in the simulation experiment, it would be more natural to simulate α and γ as independent random variables and to deduce $\beta_i = \gamma_i - \alpha_i$ and $\beta_{ij} = \gamma_{ij} - \alpha_{ij}$.

Another important relation between the parameters is given by the unconditional variance $\sigma_i^2 = \omega_i/(1-\gamma_i)$. It turns out that variance parameters ω_i and γ_i have a high negative correlation (-0.83), while the correlation between γ_i and σ_i^2 is rather low (0.21). For covariance parameters $(\omega_{ij}, \gamma_{ij}, \sigma_{ij})$, one obtains a similar correlation between ω_{ij} and γ_{ij} (-0.71) and between γ_{ij} and σ_{ij} (-0.14). This suggests to simulate γ and σ independently from each other, and to deduce $\omega_i = (1 - \gamma_i)\,\sigma_i^2$ and $\omega_{ij} = (1 - \gamma_{ij})\,\sigma_{ij}$.

5.2 Simulation Experiment

I begin with a brief description of the experiment design. As argued before, individual parameters α_i and γ_i are drawn from independent Beta distributions with parameters (p_α, q_α) and (p_γ, q_γ) and β_i is deduced from $\gamma_i - \alpha_i$. The unconditional variance σ_i^2 is drawn from a symmetric Beta distribution with $p = q = 3$ and the simulated parameters are truncated in the range $[\underline{\sigma}^2, \overline{\sigma}^2]$. Then ω_i is deduced using $\omega_i = (1 - \gamma_i)\sigma_i^2$. Parameters pertaining to covariances are defined in the same way. The unconditional correlation ρ_{ij} is assumed to be independent of γ_i and γ_{ij}. It is drawn from a Beta distribution and truncated in the range $[\underline{\rho}, \overline{\rho}]$. Then, the unconditional covariance is defined as $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ and $\omega_{ij} = (1 - \gamma_{ij})\sigma_{ij}$.

The characteristics of the parameter distributions for the baseline experiment are based on the distributions and calibrations described in Section 5.1 for the U.S. stocks. These characteristics are summarized in **Table 3**. For each simulation, I draw a sample of individual and joint parameters $(\alpha_i, \gamma_i, \sigma_i^2)$ and $(\alpha_{ij}, \gamma_{ij}, \sigma_{ij})$ for $i, j = 1, \cdots, N$. Then I draw a time series of N individual unexpected returns $\varepsilon_{i,t}$

for $i = 1, , \cdots, N$ and $t = 1, \cdots, T$.

Baseline case. **Table 4** reports summary statistics on parameter estimates for the various estimation procedures described above. Each experiment is based on 1,000 samples. In the baseline case, the number of observations per sample is $T = 1,000$ and the number of assets varies from $N = 10$ to 40. For the aggregation-correcting estimator θ_{AC}, I use $K_\Lambda = 10$ to 40 lags and $K_\Phi = 4$, so that the first three covariances Φ_2 to Φ_4 are freely estimated.

I begin with the case $N = 10$, which is probably a minimum for a realistic number of assets in a portfolio. As expected, the QMLE of the GARCH$(1, 1)$ process (denoted QMLE$(1, 1)$) provides biased estimates of the volatility parameters. The parameter α is over-estimated (0.049, while the expected value is 0.043). In contrast, there is a significant downward bias on the persistence parameter γ. The median estimate is 0.63, while the expected value is 0.976. Increasing the number of assets to $N = 40$ results in a decrease in the estimate of α, so that it is eventually severely under-estimated. The persistence parameter γ is also severely under-estimated, indicating that the QMLE is not able to generate the high persistence found in the aggregate volatility dynamics. This result comes from the mis-specification of the QMLE under contemporaneous aggregation of heterogeneous processes.

The LSE of the weak GARCH$(1, 1)$ process (denoted LSE$(1, 1)$) is able to correct some of the problems induced by the aggregation bias. In particular, it acknowledges that the aggregate return is a weak GARCH process, i.e. that the conditional volatility is mis-specified. However, the lag structure of this estimator is correct only when the individual processes are driven by the same parameter, i.e. under homogeneity of the individual volatility processes. The LSE provides a more persistent aggregated squared returns than the QMLE. For $N = 10$, the value of $E[\gamma]$ is estimated to be 0.88, instead of 0.63 for the QMLE. But as for the QMLE, the larger the number of assets, the higher the bias in the persistence parameter.

Turning to the aggregation-corrected estimators, the table reveals that for $N = 10$ the persistence parameter is correctly and precisely estimated to be 0.976 for the three possible values of K_Λ from 10 to 40. The mean value of α is also very well estimated, with estimates ranging from 0.041 to 0.044, depending on the value of K_Λ. Increasing the number of assets in the portfolio generates small biases in the estimation of the mean value of γ and α. However, they remain small relative to the biases implied by the other reported estimators. For instance, the mean value of γ is around 0.972, a value still close to the expected value 0.976. One also notices that the choice of the number of lags K_Λ does not significant affect the parameter estimates. This suggests that additional lags are indeed required (since the LSE$(1,1)$ is severely biased), yet only a moderate number of lags is enough to correct for the aggregation bias.

Increasing the number of observations to $T = 5,000$ results in a significant reduction of the bias of the persistence parameter for the QMLE$(1,1)$ and LSE$(1,1)$. The estimated mean value of γ is as high as 0.884 for the QMLE and 0.966 for the LSE, although the estimates are highly dispersed. The inconsistency of the LSE estimator transpires in the severe under-estimation of the mean value of α. The ACEs display essentially the same patterns as before, suggesting that they are not particularly affected by finite-sample biases.

Unreported results reveal that the constrained ACEs, $\theta^{(1)}_{CAC}$ and $\theta^{(2)}_{CAC}$, yield a sizeable under-estimation of the mean value of α, around 0.029. This reveals that the constraints imposed on the dynamics of the aggregate process are overly restrictive. In the simulation experiment, α and γ are drawn independently from each other, so that $\theta^{(1)}_{CAC}$ should perform rather well. However, as already mentioned, the covariance matrix Σ_t is imposed to be symmetric and semi-definite positive, for each date t. This correction has been found to generate some unwarranted dependency between γ and α. This may explain why the constrained aggregation-corrected estimator does not perform as well as expected.

Robustness check. To evaluate the robustness of these results, I now consider several simulation experiments based on alternative assumptions. In the first set of experiments, the range for the unconditional correlations is changed, while in the second set alternative distributions for the innovation process are used. All these simulation results are reported in **Table 5**.[14]

While the baseline case is calibrated on U.S. stocks with moderate positive correlation, the first experiment consider the case of uncorrelated assets ($\rho_{ij} \in [-0.1; 0.1]$) and the case of highly correlated assets ($\rho_{ij} \in [0.75; 0.9]$). On the one hand, with uncorrelated assets, the QMLE$(1,1)$ and the LSE$(1,1)$ provide estimates of the

[14] Experiments based on alternative numbers of lags in the ACEs (K_Φ and K_Λ) confirm that these estimators have very good finite-sample properties. Other experiments essentially left unaltered the patterns already described. In particular, there is no sizeable effect on the parameter estimates, when γ is assumed to be independent from β (with α deduced from $\gamma - \beta$) or when the magnitude of the unconditional variances and covariances are changed. The results, not reported to save space, are available in the Technical Appendix.

persistence parameter γ that are much higher than under the benchmark case (to 0.84 and 0.95 respectively). These estimates are still far below the expected value however. On the other hand, in the case of highly correlated assets, the parameter estimates obtained from the QMLE$(1,1)$ and LSE$(1,1)$ are much lower than in the baseline case (to 0.52 and 0.46 respectively). Regarding the ACEs, they turn out to be very robust to changes in the range of correlations across assets.

The second set of experiments relies on the distribution of the innovation process. Although $z_{i,t}$ has been assumed to be normally distributed from the beginning, it is well known that the empirical distribution of asset returns is often asymmetric and/or fat-tailed. Arguably, these features may affect the ability of the correcting approaches to provide unbiased estimators. To evaluate this possibility, I ran several experiments with innovations drawn from distributions with asymmetry and/or fat tails. Table 5 reports the results for a t distribution (with 4 degrees of freedom) and a skewed t distribution (with $\nu = 4$ degrees of freedom and an asymmetric parameter equal to -0.5).[15] It clearly affects the magnitude of the bias in the QMLE$(1,1)$ and LSE$(1,1)$. The median estimates of the persistence parameter γ produced by the QMLE and the LSE increase from 0.63 and 0.88 for the normal innovations to 0.71 and 0.92 for the $t(4)$ innovations, while the expected value is 0.976. Asymmetry in the innovation distribution does not appear to affect the parameter estimates significantly. Again, the characteristics of the ACEs are not affected by the change in the conditional distribution, whatever the number of lags K_Λ. This result is consistent with the fact that the ACEs do not rely on any particular distributional assumption, since they are based on least-square estimation.

These results may explain why estimating a (strong or weak) GARCH$(1,1)$ process for the aggregate dynamics often produces large persistence parameters in empirical work, while the simulation results reported in Table 4 for the baseline case

[15] The skewed t distribution is defined as

$$g(z|\nu,\lambda) = bc\left(1 + \frac{\zeta^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}},$$

where

$$\zeta = \begin{cases} (bz + a)\,/\,(1-\lambda) & \text{if } z < -a/b, \\ (bz + a)\,/\,(1+\lambda) & \text{if } z \geq -a/b, \end{cases}$$

with $a = 4\lambda c\,(\nu - 2)\,/(\nu - 1)$, $b = \sqrt{1 + 3\lambda^2 - a^2}$, and $c = \Gamma\left((\nu+1)/2\right)/\left[\sqrt{\pi(\nu-2)}\,\Gamma\left(\nu/2\right)\right]$, and ν and λ denote the degree of freedom and the asymmetric parameter respectively. λ is defined over the range $[-1;1]$, with a negative value denoting a negative skewness.

suggest that the estimation of γ is severely under-estimated when a GARCH$(1,1)$ process is estimated. In principle, the aggregation bias should induce a severe under-estimation of the persistence estimated by the GARCH$(1,1)$ process due to the omitted dynamics. However, in presence of non-normal innovations not accounted in the estimated model, the persistence term is over-estimated to capture some of the implications of the non-normality therefore reducing the overall bias.[16]

6 Evaluation of the Aggregation Bias

The simulation results reported above suggest that aggregation bias matters for the estimation of the aggregate squared returns. Under realistic parameterizations, the biases obtained with the (mis-specified) QMLE$(1,1)$ and LSE$(1,1)$ can be very large. This is likely to have dramatic consequences on asset and risk management. To investigate this issue, I now describe a multi-sector extension of the model above, which is then used to evaluate the consequences of the aggregation bias on the optimal asset allocation between size and book-to-market portfolios.

6.1 A Multi-sector Model

The model (1)–(2) can be extended to provide a multi-sector model designed for asset allocation across industry indices.[17] For this purpose, I consider an investment set composed of H sectors (or classes), indexed $h = 1, \cdots, H$, with N_h assets in sector h. All the assets in sector h have a standard strong GARCH$(1,1)$ conditional variance

$$\begin{aligned} \varepsilon_{i,t}^{(h)} &= \sigma_{i,t}^{(h)} z_{i,t}^{(h)}, \\ \sigma_{i,t}^{(h)2} &= \omega_i^{(h)} + \alpha_i^{(h)} \varepsilon_{i,t-1}^{(h)2} + \beta_i^{(h)} \sigma_{i,t-1}^{(h)2}, \end{aligned} \qquad (16)$$

with $i = 1, \cdots, N_h$. Conditional covariances between assets i and j in the same sector h are described as

$$\sigma_{ij,t}^{(h)} = \omega_{ij}^{(h)} + \alpha_{ij}^{(h)} \varepsilon_{i,t-1}^{(h)} \varepsilon_{j,t-1}^{(h)} + \beta_{ij}^{(h)} \sigma_{ij,t-1}^{(h)}, \qquad (17)$$

with $i, j = 1, \cdots, N_h$, $j \neq i$. As before, the parameters ($\alpha_i^{(h)}$ and $\alpha_{ij}^{(h)}$) and ($\beta_i^{(h)}$ and $\beta_{ij}^{(h)}$) in sector h are realizations of the same random coefficients $\alpha^{(h)}$ and $\beta^{(h)}$.

[16] The interaction between the volatility dynamics and the distributional properties has been highlighted by Engle (1982).

[17] Alternatively, one may view this set-up as a model for several asset classes (in a strategic allocation framework) or risk factors.

It is not assumed that the parameters in two different sectors are drawn from the same distribution.

Conditional covariances between asset i in sector h and asset j in sector h' are described as

$$\sigma_{ij,t}^{(h,h')} = \omega_{ij}^{(h,h')} + \alpha_{ij}^{(h,h')} \varepsilon_{i,t-1}^{(h)} \varepsilon_{j,t-1}^{(h')} + \beta_{ij}^{(h,h')} \sigma_{ij,t-1}^{(h,h')}, \tag{18}$$

with $i = 1, \cdots, N_h$ and $j = 1, \cdots, N_{h'}$, where parameters $\alpha_{ij}^{(h,h')}$ and $\beta_{ij}^{(h,h')}$ are iid and drawn from the same distribution of the random variables $\alpha^{(h,h')}$ and $\beta^{(h,h')}$.

As already shown in Section 3, the dynamics of the portfolio h return is given by

$$\varepsilon_{p,t}^{(h)2} = \Omega_p^{(h)} + \sum_{k=1}^{\infty} \Lambda_k^{(h)} \varepsilon_{p,t-k}^{(h)2} + v_{p,t}^{(h)} - \sum_{k=1}^{\infty} \Phi_k^{(h)} v_{p,t-k}^{(h)}, \tag{19}$$

where $\Omega_p^{(h)}$, $\Lambda_k^{(h)}$, and $\Phi_k^{(h)}$ reflect the characteristics of the asset returns in sector h. The only missing component is the joint dynamics of two sectors h and h'. It can be described as follows:

$$\varepsilon_{p,t}^{(h)} \varepsilon_{p,t}^{(h')} = \left(\sum_{i=1}^{N_h} w_i^{(h)} \varepsilon_{i,t}^{(h)} \right) \left(\sum_{j=1}^{N_{h'}} w_j^{(h')} \varepsilon_{j,t}^{(h')} \right) = \sum_{i=1}^{N_h} \sum_{j=1}^{N_{h'}} w_i^{(h)} w_j^{(h')} \varepsilon_{i,t}^{(h)} \varepsilon_{j,t}^{(h')},$$

where $w_i^{(h)}$ denotes the weight of asset i in sector h. Given the dynamics of the individual assets, one obtains

$$\begin{aligned} \varepsilon_{p,t}^{(h)} \varepsilon_{p,t}^{(h')} &= \sum_{i=1}^{N_h} \sum_{j=1}^{N_{h'}} w_i^{(h)} w_j^{(h')} \omega_{ij}^{(h,h')} + \sum_{i=1}^{N_h} \sum_{j=1}^{N_{h'}} w_i^{(h)} w_j^{(h')} \gamma_{ij}^{(h,h')} \varepsilon_{i,t-1}^{(h)} \varepsilon_{j,t-1}^{(h')} \\ &\quad + \sum_{i=1}^{N_h} \sum_{j=1}^{N_{h'}} w_i^{(h)} w_j^{(h')} \left(v_{ij,t}^{(h,h')} - \beta_{ij}^{(h,h')} v_{ij,t-1}^{(h,h')} \right). \end{aligned}$$

Using the same approach as before, the expression for the cross-product of unexpected returns in sectors h and h' is

$$\varepsilon_{p,t}^{(h)} \varepsilon_{p,t}^{(h')} = \Omega_p^{(h,h')} + \sum_{k=1}^{\infty} \Lambda_k^{(h,h')} \varepsilon_{p,t-k}^{(h)} \varepsilon_{p,t-k}^{(h')} + v_{p,t}^{(h,h')} - \sum_{k=1}^{\infty} \Phi_k^{(h,h')} v_{p,t-k}^{(h,h')}, \tag{20}$$

where the parameters $\Phi_k^{(h,h')} = E\left[\phi_k^{(h,h')}\right]$ and $\Lambda_k^{(h,h')} = E\left[\lambda_k^{(h,h')}\right]$, $k \geq 1$, in equation (20) are functions of the moments of the distribution of individual parameters:

$$\begin{array}{lll} \phi_1^{(h,h')} = \beta^{(h,h')} & \phi_{k+1}^{(h,h')} = \left(\lambda_k^{(h,h')} - \Lambda_k^{(h,h')} \right) \beta^{(h,h')} & k = 1, 2, \cdots \\ \lambda_1^{(h,h')} = \gamma^{(h,h')} & \lambda_{k+1}^{(h,h')} = \left(\lambda_k^{(h,h')} - \Lambda_k^{(h,h')} \right) \gamma^{(h,h')} & \end{array}$$

and $\Omega_p^{(h,h')}$ is defined in equation (24) in the Appendix.

6.2 Aggregation Bias in Size and Book-to-market Portfolios

I now evaluate the magnitude of the aggregation bias and the loss for an investor of using the mis-specified strong GARCH$(1,1)$ process instead of the aggregation-corrected model for aggregate returns. For this purpose, I consider the six size and book-to-market portfolios proposed by Fama and French (1995).[18] The data starts in January 1963 and ends in June 2007, at the weekly frequency. The sample is divided into two subperiods. The first one (1963-2002) is used for the estimation of the model (2,058 observations), and the last five years (2003-07) are used for the (out-of-sample) variance and covariance forecasts, portfolio allocation, and performance evaluation (244 observations).

Summary statistics on the six portfolios are presented in **Table 6**. As expected, the average return is higher for small firms and firms with high book-to-market ratio (or value firms). In addition, the volatility is higher for small firms and firms with low book-to-market ratio (or growth firms). Further investigation of these series reveals that they are negatively skewed and fat-tailed. Finally, the correlation matrix indicates that the portfolios are highly correlated (the minimum correlation is 0.7).

Table 7 reports parameter estimates for each portfolio and each pair of portfolios, concentrating on the QMLE$(1,1)$ and the ACEs with $K_\Lambda = 10$ or 20 lags and $K_\Phi = 4$ lags.[19] The table reveals systematic differences between these estimators. On the one hand, the parameter α is always larger with the QMLE$(1,1)$ process, for the variance as well as for the covariance processes. Indeed the average value is 0.115 for the QMLE versus 0.09 for the ACEs. This pattern suggests that under the

[18]The portfolios are available on Kenneth French's website. They result from the intersection of two portfolios formed on size (small / big) and three portfolios formed on the ratio of book equity to market equity (low / medium / high). Each portfolio results from the aggregation of a large number of U.S. equities.

[19]Estimation is performed over the 1963-2002 subperiod. Returns are defined as log-returns. Given the significant persistence found in returns for small portfolios, I first fit an autoregressive process of order one for the portfolio returns. Then, I pursue the analysis with the unexpected returns. Since the next step is based on asset allocation, the covariance matrix at date t across the six portfolio returns needs to be invertible. Consequently, the parameter estimates are again adjusted to ensure that the covariance matrix is semi-definite positive (see Ledoit, Santa-Clara, and Wolf, 2003). For the ACEs, only the mean values of the cross-section distribution of the parameters are reported to save space.

QMLE the conditional volatility is more reactive to shocks on lagged returns. On the other hand, the persistence parameter γ is in general larger with the QMLE on average 0.96 versus 0.93 for the ACEs), suggesting that variances and covariances are also more persistent with this estimation procedure.

The consequences of the aggregation bias can be evaluated statistically by comparing the relative performances of the two estimation techniques for forecasting the conditional variances and covariances of unexpected returns. One difficulty is that variances are not observable, but measured with noise by squared unexpected returns $\hat{\varepsilon}_t^2$. As shown by Patton (2006), most standard loss functions used to rank competing variance forecasts are not robust to the use of noisy proxies and therefore result in biased variance forecasts. The variance forecasts are thus compared on the basis of robust loss functions, chosen among the family described by Patton (2006):

$$\begin{aligned}
L_{1,t}(\hat{\varepsilon}_t^2, h_t) &= \left(h_t - \hat{\varepsilon}_t^2\right)^2, \\
L_{2,t}(\hat{\varepsilon}_t^2, h_t) &= \hat{\varepsilon}_t^2/h_t - \log(\hat{\varepsilon}_t^2/h_t) - 1, \\
L_{3,t}(\hat{\varepsilon}_t^2, h_t) &= (\hat{\varepsilon}_t^6 - h_t^2)/6 - h_t^2(\hat{\varepsilon}_t^2 - h_t)/2,
\end{aligned}$$

where h_t denotes the variance forecast. L_1 is the usual squared error, L_2 and L_3 are asymmetric loss measures that penalize under-predictions and over-predictions respectively. The test of the difference between the loss function for two specifications is based on the test developed by Diebold and Mariano (1995) and West (1996). The loss difference between the forecasts based on the QMLE and the ACE is defined as, for $k = 1, 2, 3$

$$d_{k,t} = L_{k,t}(\hat{\varepsilon}_t^2, \hat{\sigma}^2_{QMLE,t}) - L_{k,t}(\hat{\varepsilon}_t^2, \hat{\sigma}^2_{ACE,t}),$$

where $\hat{\sigma}^2_{QMLE}$ and $\hat{\sigma}^2_{ACE}$ denote the variance forecast based on the QMLE and the ACE respectively. The Diebold-Mariano test is simply based on the t-stat associated to the loss difference

$$DM_k = \frac{\bar{d}_k}{\bar{\sigma}_k},$$

where $\bar{d}_k$ is the sample mean and $\bar{\sigma}_k^2$ the sample variance of the loss difference. Under the null hypothesis $E[d_{k,t}] = 0$, the t-stat is asymptotically distributed as a $N(0, 1)$. Following Diebold and Mariano (1995), the sample variance is measured as $\bar{\sigma}_k^2 = 2\pi\ f_d(0)/\tau$ where $f_d(0)$ is a consistent estimate of the loss difference spectral density at frequency 0, and τ is the number of observations over the out-of-sample period. For the covariances, the loss functions are simply defined as $L_{k,t}(\hat{\varepsilon}_{i,t}\hat{\varepsilon}_{j,t}, h_{ij,t})$, where $h_{ij,t}$ denotes the covariance forecast between assets i and j.

Table 8 presents the results of the Diebold-Mariano/West test over the out-of-sample period for all variances and covariances. Several results are of importance. First, almost all the test statistics are positive, meaning that in most cases, the aggregation-correcting estimators produce better variance and covariance forecasts than the QMLE. This result was expected, since the ACE corrects for the aggregation bias, which is likely to affect the size and book-to-market portfolios. In many cases, the statistics are significantly positive, while none is significantly negative.

In most cases, the null hypothesis is rejected for variances and covariances involving small-size portfolios. Interestingly, these portfolios are characterized by large QML estimates of α. This result suggests that the main source of over-performance of the ACE comes from its ability to correctly measure the reaction of variances and covariances to past shocks.

6.3 The Opportunity Cost of the Aggregation Bias

To evaluate the opportunity cost of the aggreagation bias, I now consider an investor allocating her wealth between the six size and book-to-market portfolios. The investor is allowed to borrow and lend at the risk-free rate r_f and she can also short sell the available risky assets. For sake of simplicity, the investor is assumed to have a single-period horizon.[20] The optimal portfolio is a vector of portfolio weights $\varpi_t = \{\varpi_{1,t}, \cdots, \varpi_{N,t}\}$ and $\varpi_0 = 1 - \sum_{i=1}^{N} \varpi_{i,t}$ is the weight allocated to the risk-free asset. It is obtained by maximizing the mean-variance criterion

$$\max_{\{\varpi_t\}} \quad \mu_{p,t} - \frac{\lambda}{2}\sigma^2_{p,t},$$

where λ denotes the risk aversion parameter, $\mu_{p,t} = \varpi_t' \mu_t$ the portfolio expected return and $\sigma^2_{p,t} = \varpi_t' \Sigma_t \varpi_t$ the portfolio variance. Expected returns μ_t are simply estimated as the average returns over the period preceding the allocation. The covariance matrix Σ_t is estimated using the QMLE$(1,1)$ or the ACEs with $K_\Lambda = 10$ or 20 lags and $K_\Phi = 4$ lags.

The allocation is performed over the second subperiod (2003-07) to avoid any overfitting. For each week of the allocation sample, expected returns and the covariance matrix are computed and the optimal portfolio weights are determined.

[20] Given the specification of the aggregation-corrected dynamics, which includes in principle an infinite number of lags, the usual techniques solving the multiple-period investment problem cannot be implemented in this context (see Barberis, 2000). The extension to multiple-horizon investment is left for further research.

The evaluation of the cost of using the QMLE$(1,1)$ in presence of aggregation bias is based on two measures. The first measure of performance is the Sharpe ratio, which is computed using the ex-post average return μ_p and volatility σ_p, as $SR_p = (\mu_p - r_f)/\sigma_p$. Since the Sharpe ratio does not provide a measure of outperformance over alternative strategies, I also consider a second tool, namely the *performance fee* measure proposed by West, Edison, and Cho (1993) and Fleming, Kirby, and Ostdiek (2001, 2003). It measures the management fee an investor is willing to pay to switch from the suboptimal strategy to the optimal strategy. If we denote $\mu^*_{p,t}$ and $\sigma^{*2}_{p,t}$ the expected return and variance of the portfolio return obtained under the optimal strategy, and $\hat{\mu}_{p,t}$ and $\hat{\sigma}^2_{p,t}$ the expected return and variance of the portfolio return obtained using the suboptimal strategy, the performance fee (or opportunity cost), denoted ϑ, is defined as[21]

$$\vartheta = (\mu^*_{p,t} - \hat{\mu}_{p,t}) - \frac{\lambda}{2}(\sigma^{*2}_{p,t} - \hat{\sigma}^2_{p,t}). \tag{21}$$

Summary statistics on realized portfolio returns and performance measures are reported in **Table 9**, for various levels of risk aversion. As it appears clearly, the optimal allocations based on the ACE systematically yield a higher expected return than the allocations based on the QMLE$(1,1)$. There is also an increase in the risk of the portfolio, but this increase remains moderate, suggesting that the investor adopting the ACE strategy is able to benefit from investment opportunities in a more efficient way, thanks to a better evaluation of assets' risk. As a result, the Sharpe ratio is much higher for the ACE strategies than for the QMLE strategy, whatever the level of risk aversion.

The table also reveals that the performance fee is positive and large for all levels of risk aversion. For instance for $\lambda = 2$, the investor is willing to pay a premium of 97 basis points to benefit from the ACE$(20,4)$ strategy in place of the QMLE$(1,1)$ strategy. This premium has to be compared to the expected return of the QMLE$(1,1)$ strategy, which is equal to 4.6%. The performance fee therefore represents a significant portion of the expected return. Turning to a higher risk aversion does not alter this result. For $\lambda = 10$, the performance fee is equal to 20 basis points to compare with an expected return of 0.96% for the QMLE$(1,1)$ strategy. For all risk aversions, the performance fee represents about one fifth of the expected return of the suboptimal strategy.

[21] In this context, the performance fee and the certainty equivalent, previously adopted among others by Kandel and Stambaugh (1996) and Campbell and Viceira (1999), are equivalent.

7 Conclusion

This paper describes the dynamics of the aggregate squared returns in presence of several assets with volatility driven by a strong GARCH$(1,1)$ process. It is shown to be a weak GARCH process with an infinite number of lags, i.e. an infinite ARMA process for aggregate squared returns. This result establishes a relation between the characteristics of the persistence parameter across assets and the dynamics of the aggregate squared returns. The proposed estimation procedure is consistent with the infinite weak GARCH structure and explicitly acknowledges this relation between the parameters of the ARMA process and the moments of the cross-section distribution of the persistence parameter. This estimation procedure provides an unbiased estimate of the dynamics of aggregate squared returns and performs very well in finite sample.

In this paper, I also evaluate the cost of assuming a strong GARCH$(1,1)$ process to model the aggregate volatility instead of the aggregation-corrected estimator. For the well-known and widely-used size and book-to-market portfolios, this cost is shown to be both statistically significant and economically sizeable. This result suggests that it is of great importance for both asset and risk management to use the appropriate dynamics of returns when they result from aggregation. This is the case in particular for sectoral indices, asset classes or even portfolios based on asset characteristics (size, book-to-market ratio, momentum, etc.).

Some issues remain unresolved for the moment. In particular, it is not clear how to handle the aggregation of asymmetric GARCH processes. The main difficulty that arises in this case is that the characteristics of the individual asymmetries are lost when only aggregate returns are considered. Another interesting extension is the case of time-varying weights. For sectoral indices or asset classes, the weights of the various components are likely to vary over time, in general in correlation with the relative market capitalization of the individual assets. When this is the case, the parameters of the aggregate squared return process are themselves time-varying, since they reflect the cross-section average of the individual parameters based on time-dependent weights. In such an instance, there are some composition effects that need to be explicitly addressed. These extensions are left for further research.

References

[1] Andersen, T.G., T. Bollerslev, P.F. Christoffersen and F.X. Diebold (2005), Practical Volatility and Correlation Modeling for Financial Market Risk Management, in M. Carey and R. Stulz (eds.), Risks of Financial Institutions. Chicago: University of Chicago Press.

[2] Barberis, N. (2000), Investing for the Long Run when Returns are Predictable, *Journal of Finance*, 55:225–264.

[3] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31:307–327.

[4] Bollerslev, T., R.F. Engle, J.M. Wooldridge (1988), A Capital Asset Pricing Model with Time-varying Covariance, *Journal of Political Economy*, 96:116–131.

[5] Campbell, J.Y., and L.M. Viceira (1999), Consumption and Portfolio Decisions When Expected Returns Are Time Varying, *Quarterly Journal of Economics*, 114:433–495.

[6] Diebold, F.X., and R. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business and Economics Statistics*, 13:253–264.

[7] Drost, F.C., and T.E. Nijman (1993), Temporal Aggregation of GARCH Processes, *Econometrica*, 61:909–927.

[8] Engle, R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50:987–1007.

[9] Fama, E.F., and K.R. French (1995), Size and Book-to-market Factors in Earnings and Returns, *Journal of Finance*, 50:131–155.

[10] Fleming, J., C. Kirby, and B. Ostdiek (2003), The Economic Value of Volatility Timing Using "Realized" Volatility, *Journal of Financial Economics*, 67:473–509.

[11] Fleming, J., C. Kirby, and B. Ostdiek (2001), The Economic Value of Volatility Timing, *Journal of Finance*, 56:329–352.

[12] Francq, C., and J.M. Zakoian (2000), Estimating Weak GARCH Representations, *Econometric Theory*, 16:692–728.

[13] Gonçalves, E., and C. Gouriéroux (1988), Agrégation de Processus Autorégressifs d'Ordre 1, *Annales d'Économie et de Statistique*, 12: 127–149.

[14] Granger, C.W.J. (1980), Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics*, 14:227–238.

[15] Hafner, C.M. (2008), Temporal Aggregation of Multivariate GARCH Models, forthcoming *in Journal of Econometrics*.

[16] Kandel, S., and R. Stambaugh (1996), On the Predictability of Stock Returns: An Asset Allocation Perspective, *Journal of Finance* 51:385–424.

[17] Komunjer, I. (2001), Consistent Estimation for Aggregated GARCH Processes, UCSD Working Paper.

[18] Ledoit, O., P. Santa-Clara, and M. Wolf (2003), Flexible Multivariate GARCH Modeling with an Application to International Stock Markets. *Review of Economics and Statistics*, 85:735–747.

[19] Lewbel, A. (1994), Aggregation and Simple Dynamics, *American Economic Review*, 84:905–918.

[20] Meddahi, N., and E. Renault (2004), Temporal Aggregation of Volatility Models, *Journal of Econometrics*, 119:355–379.

[21] Nijman, T.E., and E. Sentana (1996), Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes, *Journal of Econometrics*, 71:71–87.

[22] Patton, A. (2006), Volatility Forecast Comparison Using Imperfect Volatility Proxies, Working Paper, London School of Economics.

[23] West, K.D. (1996), Asymptotic Inference About Predictive Ability, *Econometrica*, 64:1067–1084.

[24] West, K.D., H.J. Edison, and D. Cho (1993), A Utility-Based Comparison of some Models of Exchange Rate Volatility, *Journal of International Economics*, 35:23–45.

[25] Zaffaroni, P. (2007), Aggregation of GARCH Processes, *Journal of Time Series Analysis*, 28:521–544.

8 Appendix

8.1 Proof of Proposition 2

The proof of Proposition 2 is based on a set of results about the properties of the various innovation processes of the model. First, assuming that innovations $z_{i,t}$ are normally distributed, one has the following moments and co-moments of $z_{i,t}$

$$\begin{aligned}
&E[z_{i,t}] = 0\\
&E[z_{i,t}^2] = 1\\
&E[z_{i,t}^3] = 0\\
&E[z_{i,t}^4] = 3\\
&E[z_{i,t}z_{j,t}] = \rho_{ij}\\
&E[z_{i,t}^2 z_{j,t}] = E[z_{i,t}z_{j,t}z_{k,t}] = 0\\
&E[z_{i,t}^2 z_{j,t}^2] = 1 + 2\rho_{ij}^2\\
&E[z_{i,t}^3 z_{j,t}] = 3\rho_{ij}\\
&E[z_{i,t}^2 z_{j,t} z_{k,t}] = \rho_{jk} + 2\rho_{ij}\rho_{ik}\\
&E[z_{i,t}z_{j,t}z_{k,t}z_{l,t}] = \rho_{ij}\rho_{kl} + \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk}.
\end{aligned}$$

Regarding unexpected returns, one has the following relations, denoting $\sigma_i^2 = E[\sigma_{i,t}^2] = \omega_i/(1-\gamma_i)$ and $\sigma_{ij} = E[\sigma_{ij,t}] = \omega_{ij}/(1-\gamma_{ij}) = \sigma_i\sigma_j\rho_{ij}$,

$$\begin{aligned}
&E[\varepsilon_{i,t}] = 0\\
&E[\varepsilon_{i,t}^2] = E[\sigma_{i,t}^2] = \sigma_i^2\\
&E[\varepsilon_{i,t}^3] = 0\\
&E[\varepsilon_{i,t}^4] = 3E[\sigma_{i,t}^4]\\
&E[\varepsilon_{i,t}\varepsilon_{j,t}] = E[\sigma_{ij,t}] = \sigma_{ij}\\
&E[\varepsilon_{i,t}^2\varepsilon_{j,t}] = E[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{k,t}] = 0\\
&E[\varepsilon_{i,t}^2\varepsilon_{j,t}^2] = E[\sigma_{i,t}^2\sigma_{j,t}^2] + 2E[\sigma_{ij,t}^2]\\
&E[\varepsilon_{i,t}^3\varepsilon_{j,t}] = 3E[\sigma_{i,t}^2\sigma_{ij,t}]\\
&E[\varepsilon_{i,t}^2\varepsilon_{j,t}\varepsilon_{k,t}] = E[\sigma_{i,t}^2\sigma_{jk,t}] + 2E[\sigma_{ij,t}\sigma_{jk,t}]\\
&E[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{k,t}\varepsilon_{l,t}] = E[\sigma_{ij,t}\sigma_{kl,t}] + E[\sigma_{ik,t}\sigma_{jl,t}] + E[\sigma_{il,t}\sigma_{jk,t}].
\end{aligned}$$

With these results, it is possible to derive the expressions given in Proposition 2.

(1) The expected value of $v_{i,t}$ is given by

$$E[v_{i,t}] = E[\varepsilon_{i,t}^2 - \sigma_{i,t}^2] = E[\sigma_{i,t}^2(z_{i,t}^2 - 1)] = E[\sigma_{i,t}^2(E[z_{i,t}^2/I_{t-1}] - 1)] = 0,$$

where the last equality holds because $E[z_{i,t}^2/I_{t-1}] = 1$. Similarly,

$$\begin{aligned}
E[v_{ij,t}] &= E[\varepsilon_{i,t}\varepsilon_{j,t} - \sigma_{ij,t}] = E[\sigma_{i,t}\sigma_{j,t}z_{i,t}z_{j,t}] - E[\sigma_{ij,t}]\\
&= E[\sigma_{i,t}\sigma_{j,t}E[z_{i,t}z_{j,t}/I_{t-1}]] - E[\sigma_{ij,t}] = \sigma_i\sigma_j\rho_{ij} - E[\sigma_{ij,t}] = 0.
\end{aligned}$$

Innovations $v_{i,t}$ are serially uncorrelated, because for $s < t$

$$E[v_{i,s}v_{i,t}] = E[(\varepsilon_{s,t}^2 - \sigma_{s,t}^2)(\varepsilon_{i,t}^2 - \sigma_{i,t}^2)] = E[(\varepsilon_{s,t}^2 - \sigma_{s,t}^2)(E[\varepsilon_{i,t}^2/I_s] - E[\sigma_{i,t}^2/I_s])],$$

with $E[\varepsilon_{i,t}^2/I_s] = E[\sigma_{i,t}^2/I_s]$ $\forall s < t$. As a consequence, one has

$$E[v_{i,s}v_{i,t}] = 0 \quad \forall s < t.$$

(2) The variance of $v_{i,t}$ is given by

$$\begin{aligned} E[v_{i,t}^2] &= E[\left(\varepsilon_{i,t}^2 - \sigma_{i,t}^2\right)^2] = E[\varepsilon_{i,t}^4 - 2\varepsilon_{i,t}^2\sigma_{i,t}^2 + \sigma_{i,t}^4] = 3E[\sigma_{i,t}^4] - 2E[\sigma_{i,t}^4] + E[\sigma_{i,t}^4] \\ &= 2E[\sigma_{i,t}^4], \end{aligned}$$

because $E[\varepsilon_{i,t}^2\sigma_{i,t}^2] = E[\sigma_{i,t}^4]$. In addition, since $E[\varepsilon_{i,t}^4] = 3E[\sigma_{i,t}^4]$, one has

$$\begin{aligned} E[\sigma_{i,t}^4] &= E[\left(\omega_i + \alpha_i\varepsilon_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2\right)^2] \\ &= \omega_i^2 + 2\omega_i E[\alpha_i\varepsilon_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2] + E[\sigma_{i,t-1}^4(\alpha_i^2 z_{i,t-1}^4 + \beta_i^2 + 2\alpha_i\beta_i z_{i,t-1}^2)] \\ &= \omega_i^2 + 2\omega_i(\alpha_i + \beta_i)\sigma_i^2 + E[\sigma_{i,t}^4](3\alpha_i^2 + \beta_i^2 + 2\alpha_i\beta_i), \end{aligned}$$

where $\sigma_i^2 = \omega_i/(1 - \alpha_i - \beta_i) = \omega_i/(1 - \gamma_i)$. Therefore, one has

$$E[\sigma_{i,t}^4] = \frac{\sigma_i^4[(1-\gamma_i)^2 + 2(1-\gamma_i)\gamma_i]}{1 - 3\alpha_i^2 + \beta_i^2 + 2\alpha_i\beta_i} = \frac{\sigma_i^4(1-\gamma_i^2)}{1 - \gamma_i^2 - 2\alpha_i^2}.$$

Finally, one obtains, provided $\alpha_i^2 < (1 - \gamma_i^2)/2$

$$E[v_{i,t}^2] = \frac{2\sigma_i^4(1-\gamma_i^2)}{1 - \gamma_i^2 - 2\alpha_i^2}.$$

The variance of $v_{ij,t}$ is given by

$$\begin{aligned} E[v_{ij,t}^2] &= E[(\varepsilon_{i,t}\varepsilon_{j,t} - \sigma_{ij,t})^2] = E[\varepsilon_{i,t}^2\varepsilon_{j,t}^2 - 2\varepsilon_{i,t}\varepsilon_{j,t}\sigma_{ij,t} + \sigma_{ij,t}^2] \\ &= E[\sigma_{i,t}^2\sigma_{j,t}^2] + 2E[\sigma_{ij,t}^2] - 2E[z_{i,t}z_{j,t}\sigma_{i,t}\sigma_{j,t}\sigma_{ij,t}] + E[\sigma_{ij,t}^2] \\ &= E[\sigma_{i,t}^2\sigma_{j,t}^2] + E[\sigma_{ij,t}^2]. \end{aligned}$$

In addition,

$$\begin{aligned} E[\sigma_{i,t}^2\sigma_{j,t}^2] &= E[\left(\omega_i + \alpha_i\varepsilon_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2\right)\left(\omega_j + \alpha_j\varepsilon_{j,t-1}^2 + \beta_j\sigma_{j,t-1}^2\right)] \\ &= \omega_i\omega_j + \omega_i(\alpha_j + \beta_j)\sigma_j^2 + \omega_j(\alpha_i + \beta_i)\sigma_i^2 + \alpha_i\alpha_j E[z_{i,t-1}^2 z_{j,t-1}^2\sigma_{i,t-1}^2\sigma_{j,t-1}^2] \\ &\quad + \alpha_i\beta_j E[z_{i,t-1}^2\sigma_{i,t-1}^2\sigma_{j,t-1}^2] + \beta_i\alpha_j E[z_{j,t-1}^2\sigma_{i,t-1}^2\sigma_{j,t-1}^2] + \beta_i\beta_j E[\sigma_{i,t-1}^2\sigma_{j,t-1}^2] \\ &= \sigma_i^2\sigma_j^2(1 - \alpha_i\alpha_j - \alpha_i\beta_j - \beta_i\alpha_j - \beta_i\beta_j) \\ &\quad + E[\sigma_{i,t-1}^2\sigma_{j,t-1}^2](\alpha_i\alpha_j(1 + 2\rho_{ij}^2) + \alpha_i\beta_j + \beta_i\alpha_j + \beta_i\beta_j), \end{aligned}$$

so that, provided $\alpha_i\alpha_j < (1-\gamma_i\gamma_j)/(2\rho_{ij}^2)$

$$\begin{aligned}
E[\sigma_{i,t}^2\sigma_{j,t}^2] &= \frac{\sigma_i^2\sigma_j^2(1-\alpha_i\alpha_j-\alpha_i\beta_j-\beta_i\alpha_j-\beta_i\beta_j)}{1-(\alpha_i\alpha_j(1+2\rho_{ij}^2)+\alpha_i\beta_j+\beta_i\alpha_j+\beta_i\beta_j)} \\
&= \frac{\sigma_i^2\sigma_j^2(1-\gamma_i\gamma_j)}{1-\gamma_i\gamma_j-2\alpha_i\alpha_j\rho_{ij}^2}.
\end{aligned}$$

One also has

$$\begin{aligned}
E[\sigma_{ij,t}^2] &= E[(\omega_{ij}+\alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1}+\beta_{ij}\sigma_{ij,t-1})^2] \\
&= \omega_{ij}^2+2\omega_{ij}(\alpha_{ij}+\beta_{ij})\sigma_{ij}+\alpha_{ij}^2E[z_{i,t-1}^2z_{j,t-1}^2\sigma_{i,t-1}\sigma_{j,t-1}] \\
&\quad +2\alpha_{ij}\beta_{ij}E[z_{i,t-1}z_{j,t-1}\sigma_{i,t-1}\sigma_{j,t-1}\sigma_{ij,t-1}]+\beta_{ij}^2E[\sigma_{ij,t-1}^2] \\
&= \sigma_{ij}^2[(1-\gamma_{ij})^2+2\gamma_{ij}(1-\gamma_{ij})] \\
&\quad +E[\sigma_{ij,t-1}^2](2\alpha_{ij}^2+2\alpha_{ij}\beta_{ij}+\beta_{ij}^2)+\alpha_{ij}^2E[\sigma_{i,t-1}^2\sigma_{j,t-1}^2],
\end{aligned}$$

so that, provided $\alpha_{ij}^2 < (1-\gamma_{ij}^2)$

$$\begin{aligned}
E[\sigma_{ij,t}^2] &= \frac{\sigma_{ij}^2[(1-\gamma_{ij})^2+2\gamma_{ij}(1-\gamma_{ij})]+\alpha_{ij}^2E[\sigma_{i,t-1}^2\sigma_{j,t-1}^2]}{1-(2\alpha_{ij}^2+2\alpha_{ij}\beta_{ij}+\beta_{ij}^2)} \\
&= \frac{\sigma_{ij}^2(1-\gamma_{ij}^2)+\alpha_{ij}^2E[\sigma_{i,t-1}^2\sigma_{j,t-1}^2]}{1-\gamma_{ij}^2-\alpha_{ij}^2}.
\end{aligned}$$

Finally, one obtains

$$\begin{aligned}
E[v_{ij,t}^2] &= E[\sigma_{i,t}^2\sigma_{j,t}^2]+E[\sigma_{ij,t}^2] \\
&= \frac{\sigma_{ij}^2(1-\gamma_{ij}^2)}{1-\gamma_{ij}^2-\alpha_{ij}^2}+\left(1+\frac{\alpha_{ij}^2}{1-\gamma_{ij}^2-\alpha_{ij}^2}\right)E[\sigma_{i,t-1}^2\sigma_{j,t-1}^2] \\
&= \frac{\sigma_{ij}^2(1-\gamma_{ij}^2)}{1-\gamma_{ij}^2-\alpha_{ij}^2}+\frac{1-\gamma_{ij}^2}{1-\gamma_{ij}^2-\alpha_{ij}^2}\frac{\sigma_i^2\sigma_j^2(1-\gamma_i\gamma_j)}{1-\gamma_i\gamma_j-2\alpha_i\alpha_j\rho_{ij}^2}.
\end{aligned}$$

Similar computations show that

$$\begin{aligned}
E[v_{i,t}v_{j,t}] &= E[\left(\varepsilon_{i,t}^2-\sigma_{i,t}^2\right)\left(\varepsilon_{j,t}^2-\sigma_{j,t}^2\right)] = E[\varepsilon_{i,t}^2\varepsilon_{j,t}^2-\varepsilon_{i,t}^2\sigma_{j,t}^2-\varepsilon_{j,t}^2\sigma_{i,t}^2+\sigma_{i,t}^2\sigma_{j,t}^2] \\
&= E[\sigma_{i,t}^2\sigma_{j,t}^2]+2E[\sigma_{ij,t}^2]-2E[\sigma_{i,t}^2\sigma_{j,t}^2]+E[\sigma_{i,t}^2\sigma_{j,t}^2] = 2E[\sigma_{ij,t}^2] \\
E[v_{i,t}v_{ij,t}] &= E[\left(\varepsilon_{i,t}^2-\sigma_{i,t}^2\right)(\varepsilon_{i,t}\varepsilon_{j,t}-\sigma_{ij,t})] = 2E[\sigma_{i,t}^2\sigma_{ij,t}] \\
E[v_{i,t}v_{jk,t}] &= E[\left(\varepsilon_{i,t}^2-\sigma_{i,t}^2\right)(\varepsilon_{j,t}\varepsilon_{k,t}-\sigma_{jk,t}) = 2E[\sigma_{ij,t}\sigma_{ik,t}] \\
E[v_{ij,t}v_{ik,t}] &= E[(\varepsilon_{i,t}\varepsilon_{j,t}-\sigma_{ij,t})(\varepsilon_{i,t}\varepsilon_{k,t}-\sigma_{ik,t})] = E[\sigma_{i,t}^2\sigma_{jk,t}]+E[\sigma_{ij,t}\sigma_{ik,t}] \\
E[v_{ij,t}v_{kl,t}] &= E[(\varepsilon_{i,t}\varepsilon_{j,t}-\sigma_{ij,t})(\varepsilon_{k,t}\varepsilon_{l,t}-\sigma_{kl,t})] = E[\sigma_{ik,t}\sigma_{jl,t}]+E[\sigma_{il,t}\sigma_{jk,t}],
\end{aligned}$$

where, provided the denominators are strictly positive

$$
\begin{aligned}
E[\sigma_{i,t}^2\sigma_{ij,t}] &= \frac{\sigma_i^2\sigma_{ij}(1-\gamma_i\gamma_{ij})}{1-\gamma_i\gamma_{ij}-2\alpha_i\alpha_{ij}} \\
E[\sigma_{i,t}^2\sigma_{jk,t}] &= \frac{\sigma_i^2\sigma_{jk}(1-\gamma_i\gamma_{jk})+2\alpha_i\alpha_{jk}E[\sigma_{ij,t-1}\sigma_{jk,t-1}]}{1-\gamma_i\gamma_{jk}} \\
E[\sigma_{ij,t}\sigma_{ik,t}] &= \frac{\sigma_{ij}\sigma_{ik}(1-\gamma_{ij}\gamma_{ik})+\alpha_{ij}\alpha_{ik}E[\sigma_{i,t-1}^2\sigma_{jk,t-1}]}{1-\gamma_{ij}\gamma_{ik}-\alpha_{ij}\alpha_{ik}} \\
E[\sigma_{ij,t}\sigma_{kl,t}] &= \frac{\sigma_{ij}\sigma_{kl}(1-\gamma_{ij}\gamma_{kl})+\alpha_{ij}\alpha_{kl}E[\sigma_{ik,t-1}\sigma_{jl,t-1}+\sigma_{il,t-1}\sigma_{jk,t-1}]}{1-\gamma_{ij}\gamma_{kl}}.
\end{aligned}
$$

(3) Processes $v_{i,t}$ and $v_{ij,t}$ are orthogonal to any random coefficient driving the dynamics of conditional variances and covariances, since they are defined as the innovations of the variance and covariance processes.

8.2 Proof of Proposition 3

The dynamics of the aggregate process $\varepsilon_{p,t}^2$ is obtained from all the individual GARCH processes (1) and (2) as follows

$$
\begin{aligned}
\varepsilon_{p,t}^2 &= \left(\sum_{i=1}^N w_i\varepsilon_{i,t}\right)^2 = \sum_{i=1}^N w_i^2\left(\omega_i+\gamma_i\varepsilon_{i,t-1}^2+v_{i,t}-\beta_i v_{i,t-1}\right) \\
&\quad +2\sum_{i=1}^N\sum_{j>i}^N w_iw_j\left(\omega_{ij}+\gamma_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1}+v_{ij,t}-\beta_{ij}v_{ij,t-1}\right) \\
&= \left[\sum_{i=1}^N w_i^2\omega_i+2\sum_{i=1}^N\sum_{j>i}^N w_iw_j\omega_{ij}\right]+\left[\sum_{i=1}^N w_i^2\gamma_i\varepsilon_{i,t-1}^2+2\sum_{i=1}^N\sum_{j>i}^N w_iw_j\gamma_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1}\right] \\
&\quad +\left[\sum_{i=1}^N w_i^2\left(v_{i,t}-\beta_i v_{i,t-1}\right)+2\sum_{i=1}^N\sum_{j>i}^N w_iw_j\left(v_{ij,t}-\beta_{ij}v_{ij,t-1}\right)\right]. \qquad (22)
\end{aligned}
$$

The second term in brackets in equation (22) writes

$$\begin{aligned}
&\sum_{i=1}^{N} w_i^2 \gamma_i \varepsilon_{i,t-1}^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \gamma_{ij} \varepsilon_{i,t-1}\varepsilon_{j,t-1} = \\
&\qquad \Lambda_1 \sum_{i=1}^{N} w_i^2 \varepsilon_{i,t-1}^2 + 2\Lambda_1 \sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \varepsilon_{i,t-1}\varepsilon_{j,t-1} \\
&\qquad + \left[\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \varepsilon_{i,t-1}^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \varepsilon_{i,t-1}\varepsilon_{j,t-1}\right] \\
&\qquad = \Lambda_1 \varepsilon_{p,t-1}^2 \\
&\qquad + \left[\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \varepsilon_{i,t-1}^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \varepsilon_{i,t-1}\varepsilon_{j,t-1}\right].
\end{aligned}$$

where $\Lambda_1 = E\left[\gamma\right]$ denotes the cross-sectional mean of γ. Given Proposition 2(3), the third term in equation (22) writes

$$\sum_{i=1}^{N} w_i^2 \left(v_{i,t} - \beta_i v_{i,t-1}\right) + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \left(v_{ij,t} - \beta_{ij} v_{ij,t-1}\right) = v_{p,t} - \Phi_1 v_{p,t-1},$$

where $\Phi_1 = E\left[\beta\right]$ denotes the cross-sectional mean of β and $v_{p,t} = \sum_{i=1}^{N} w_i^2 v_{i,t} + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j v_{ij,t}$. Since $v_{i,t}$ and $v_{ij,t}$ are iid, the aggregate innovation $v_{p,t}$ is a white noise with mean 0 and variance σ_v^2, where $\sigma_v^2 = w'Vw$, with $w = (w_1, \cdots, w_N)'$ and $V = \left\{\sigma_{v,ij}^2\right\}_{ij}$.

Equation (22) therefore rewrites

$$\begin{aligned}
\varepsilon_{p,t}^2 &= \omega_p^{(1)} + \Lambda_1 \varepsilon_{p,t-1}^2 + v_{p,t} - \Phi_1 v_{p,t-1} \\
&\quad + \left[\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \varepsilon_{i,t-1}^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \varepsilon_{i,t-1}\varepsilon_{j,t-1}\right],
\end{aligned}$$

with

$$\omega_p^{(1)} = \sum_{i=1}^{N} w_i^2 \omega_i + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \omega_{ij}.$$

The second step is as follows:

$$\begin{aligned}
\varepsilon_{p,t}^2 &= \omega_p^{(1)} + \Lambda_1 \varepsilon_{p,t-1}^2 + v_{p,t} - \Phi_1 v_{p,t-1} \\
&\quad + \left[\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \left(\omega_i + \gamma_i \varepsilon_{i,t-2}^2 + v_{i,t-1} - \beta_i v_{i,t-2}\right)\right. \\
&\quad \left. + 2\sum_{i=1}^{N}\sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \left(\omega_{ij} + \gamma_{ij} \varepsilon_{i,t-2}\varepsilon_{j,t-2} + v_{ij,t-1} - \beta_{ij} v_{ij,t-2}\right)\right].
\end{aligned}$$

Again, one has

$$\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \left(v_{i,t-1} - \beta_i v_{i,t-2}\right) + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \left(v_{ij,t-1} - \beta_{ij} v_{ij,t-2}\right)$$
$$= E\left[\gamma - \Lambda_1\right] v_{p,t-1} - E\left[\left(\gamma - \Lambda_1\right)\beta\right] v_{p,t-2} = -\Phi_2 v_{p,t-2},$$

with $\Phi_2 = E\left[\left(\gamma - \Lambda_1\right)\beta\right]$. This gives

$$\begin{aligned}
\varepsilon_{p,t}^2 &= \omega_p^{(2)} + \Lambda_1 \varepsilon_{p,t-1}^2 + v_{p,t} - \Phi_1 v_{p,t-1} - \Phi_2 v_{p,t-2} \\
&\quad + \left[\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \gamma_i \varepsilon_{i,t-2}^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \gamma_{ij} \varepsilon_{i,t-2} \varepsilon_{j,t-2}\right] \\
&= \omega_p^{(2)} + \Lambda_1 \varepsilon_{p,t-1}^2 + \Lambda_2 \varepsilon_{p,t-2}^2 + v_{p,t} - \Phi_1 v_{p,t-1} - \Phi_2 v_{p,t-2} \\
&\quad + \left[\sum_{i=1}^{N} w_i^2 \left\{\left(\gamma_i - \Lambda_1\right) \gamma_i - E\left[\left(\gamma - \Lambda_1\right)\gamma\right]\right\} \varepsilon_{i,t-2}^2\right. \\
&\quad \left. + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} w_i w_j \left\{\left(\gamma_{ij} - \Lambda_1\right) \gamma_{ij} - E\left[\left(\gamma - \Lambda_1\right)\gamma\right]\right\} \varepsilon_{i,t-2} \varepsilon_{j,t-2}\right],
\end{aligned}$$

with

$$\omega_p^{(2)} = \omega_p^{(1)} + \left[\sum_{i=1}^{N} w_i^2 \left(\gamma_i - \Lambda_1\right) \omega_i + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} w_i w_j \left(\gamma_{ij} - \Lambda_1\right) \omega_{ij}\right].$$

Proceeding iteratively, one obtains

$$\varepsilon_{p,t}^2 = \Omega_p + \sum_{k=1}^{\infty} \Lambda_k \varepsilon_{p,t-k}^2 + v_{p,t} - \sum_{k=1}^{\infty} \Phi_k v_{p,t-k}, \tag{23}$$

with $\Lambda_k = E\left[\lambda_k\right]$ with $\lambda_1 = \gamma$ and $\lambda_{k+1} = \left(\lambda_k - \Lambda_k\right)\gamma$, $\Phi_k = E\left[\varphi_k\right]$ with $\varphi_1 = \beta$ and $\varphi_{k+1} = \left(\lambda_k - \Lambda_k\right)\beta$. The constant term is defined as

$$\Omega_p = \omega_p^{(1)} + \sum_{k=1}^{\infty} \left[\sum_{i=1}^{N} w_i^2 \left(\lambda_k - \Lambda_k\right) \omega_i + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} w_i w_j \left(\lambda_k - \Lambda_k\right) \omega_{ij}\right]. \tag{24}$$

Equation (23) can be rewritten as

$$\sigma_{p,t}^2 = \Omega_p + \sum_{k=1}^{\infty} \Psi_k \varepsilon_{p,t-k}^2 + v_{p,t} - \sum_{k=1}^{\infty} \Phi_k \sigma_{p,t-k}^2,$$

with $\Psi_k = \Lambda_k - \Phi_k$.

Equation (23) clearly simplifies when $\gamma_i = \gamma_{ij} = E[\gamma]$, $\forall i, j$, since one obtains an ARMA$(1,1)$ dynamics, similar to the one proposed by Nijman and Sentana (1996) in a simpler set-up, with

$$\varepsilon_{p,t}^2 = \omega_p^1 + E[\gamma] \varepsilon_{p,t-1}^2 + v_{p,t} - E[\beta] v_{p,t-1}.$$

8.3 Proof of Corollary 1

When γ and α are independent from each other, one has $\Phi_1 = \Lambda_1 - \Psi_1$ and $\Phi_k = \Lambda_k$, $\forall k > 1$, so that

$$\begin{aligned}
\varepsilon_{p,t}^2 &= \Omega_p + \sum_{k=1}^{\infty} \Lambda_k \varepsilon_{p,t-k}^2 + \upsilon_{p,t} + \Psi_1 \upsilon_{p,t-1} - \sum_{k=1}^{\infty} \Lambda_k \upsilon_{p,t-k} \\
\varepsilon_{p,t}^2 - \upsilon_{p,t} &= \Omega_p + \sum_{k=1}^{\infty} \Lambda_k \left(\varepsilon_{p,t-k}^2 - \upsilon_{p,t-k} \right) + \Psi_1 \upsilon_{p,t-1} \\
\varepsilon_{p,t}^2 - \upsilon_{p,t} &= \left(1 - \sum_{k=1}^{\infty} \Lambda_k L^k \right)^{-1} \left(\Omega_p + \Psi_1 \upsilon_{p,t-1} \right) \\
\varepsilon_{p,t}^2 &= \frac{\Omega_p}{1 - \sum_{k=1}^{\infty} \Lambda_k} + \upsilon_{p,t} + \Psi_1 \sum_{k=1}^{\infty} E\left[\gamma^k\right] \upsilon_{p,t-k},
\end{aligned}$$

with $\Psi_1 = E[\gamma] - E[\beta] = E[\alpha]$.

When γ and β are independent from each other, one has $\Phi_1 = \Lambda_1 - \Psi_1$ and $\Phi_k = 0$, $\forall k > 1$, so that

$$\begin{aligned}
\varepsilon_{p,t}^2 &= \Omega_p + \sum_{k=1}^{\infty} \Lambda_k \varepsilon_{p,t-k}^2 + \upsilon_{p,t} - \Phi_1 \upsilon_{p,t-1} \\
\varepsilon_{p,t}^2 &= \frac{\Omega_p}{1 - \sum_{k=1}^{\infty} \Lambda_k} + \sum_{k=0}^{\infty} E\left[\gamma^k\right] \left(\upsilon_{p,t-k} - \Phi_1 \upsilon_{p,t-k-1} \right) \\
\varepsilon_{p,t}^2 &= \frac{\Omega_p}{1 - \sum_{k=1}^{\infty} \Lambda_k} + \sum_{k=0}^{\infty} \left(E[\gamma^k] - E[\beta]\, E[\gamma^{k-1}] \right) \upsilon_{p,t-k}.
\end{aligned}$$

Captions

Table 1: This table provides summary statistics on the parameter estimates of conditional variance and covariance processes. Parameters are (α_i, β_i, γ_i, ω_i, σ_i) for conditional variances and (α_{ij}, β_{ij}, γ_{ij}, ω_{ij}, ρ_{ij}) for conditional covariances. Summary statistics are the mean, standard deviation, skewness, kurtosis, minimum, and maximum of the empirical distribution.

Table 2: This table provides cross-correlation between parameter estimates of conditional variance and covariance processes. Parameters are (α_i, β_i, γ_i, ω_i, σ_i^2) for conditional variances and (α_{ij}, β_{ij}, γ_{ij}, ω_{ij}, σ_{ij}) for conditional covariances.

Table 3: This table provides the main characteristics of the baseline simulation experiment. It reports the distribution, the parameters, and the range used to simulate the various parameters of the GARCH$(1,1)$ individual processes.

Table 4: This table provides estimates of the mean value of the GARCH parameters. The expected values of the simulated random parameters are $\Omega_p = 0.01$, $E[\alpha] = 0.0434$, $E[\beta] = 0.932$, and $E[\gamma] = 0.976$. In parentheses is reported the standard deviation of the parameter distribution. Aggregated-corrected estimators are based on $K_\Lambda = 10$ to 40 lags and $K_\Phi = 4$ lags.

Table 5: This table provides estimates of the mean value of the GARCH parameters under various changes of the baseline experiment. The expected values of the simulated random parameters are $\Omega_p = 0.01$, $E[\alpha] = 0.0434$, $E[\beta] = 0.932$, and $E[\gamma] = 0.976$. In parentheses is reported the standard deviation of the parameter distribution. Aggregated-corrected estimators are based on $K_\Lambda = 10$ to 40 lags and $K_\Phi = 4$ lags.

Table 6: This table reports some summary statistics on size and book-to-market portfolios. Panel A displays univariate statistics, i.e. the mean, standard deviation, skewness, and kurtosis of portfolio returns. Panel B displays the covariances (upper part of the matrix) and correlations (lower part) between portfolio returns. The statistics have been estimated over the 1963-2007 sample.

Table 7: This table provides the parameter estimates of the QMLE$(1,1)$ and two ACEs, for each of the portfolios and for each pair of portfolios. The ACEs include 4 lags Φ_k and 10 lags (respectively, 20 lags) Λ_k. For the ACEs, the table only reports to mean value of the distributions of α and γ, to save space.

Table 8: This table provides statistics on the forecasting ability of the ACEs relative to the QMLE$(1,1)$, for each of the portfolios and for each pair of portfolios. DM_1, DM_2, and DM_3 are the Diebold and Mariano (1995) test statistics associated to the loss functions L_1, L_2, and L_3 described in the paper. Under the null hypothesis that the two estimation techniques have the same forecasting ability, the test statistics are distributed as a $N(0,1)$. ^{a} and ^{b} indicate that the test statistic is significant at the 1% and 5% significance level, respectively.

Table 9: This table provides the moments of the realized portfolio return and the performance measures associated for the optimal portfolio obtained using three estimation techniques: the QMLE$(1,1)$ and two ACEs. The ACEs include 4 lags Φ_k and 10 lags (respectively, 20 lags) Λ_k. Levels of risk aversion are $\lambda = 2$, 5, and 10. The performance fee is defined in equation (21).

Figure 1: This figure displays the empirical distribution of the parameter estimates for the individual GARCH$(1,1)$ models (on the left hand side) and the bivariate GARCH$(1,1)$ models (on the right hand side) for 66 U.S. stocks over the 1988-2004 period. Parameters are $(\alpha_i, \beta_i, \gamma_i)$ on the left hand side and $(\alpha_{ij}, \beta_{ij}, \gamma_{ij})$ on the right hand side.

Figure 2: This figure displays the empirical distribution of the parameter estimates for the individual GARCH$(1,1)$ models (on the left hand side) and the bivariate GARCH$(1,1)$ models (on the right hand side) for 66 U.S. stocks over the 1988-2004 period. Parameters are (ω_i, σ_i) on the left hand side and (ω_{ij}, ρ_{ij}) on the right hand side.

Table 1: Summary statistics on the parameter estimates of conditional variance and covariance processes

	α_i	β_i	γ_i	ω_i	σ_i
Panel A: Conditional variances					
Mean	0.0487	0.9385	0.9872	0.0050	0.0208
Std dev	0.0213	0.0315	0.0128	0.0053	0.0072
Skewness	0.8215	−1.3518	−2.3704	3.2819	2.4185
Kurtosis	4.2900	5.4133	8.9874	16.4591	10.3222
Minimum	0.0074	0.8258	0.9332	0.0007	0.0102
Maximum	0.1213	0.9901	0.9975	0.0334	0.0528
	α_{ij}	β_{ij}	γ_{ij}	ω_{ij}	ρ_{ij}
Panel B: Conditional covariances					
Mean	0.0258	0.9479	0.9736	0.0016	0.1805
Std dev	0.0081	0.0171	0.0146	0.0010	0.0880
Skewness	0.3007	−1.1873	−1.4610	1.7674	1.0218
Kurtosis	3.2520	5.3413	6.1380	8.0563	5.1218
Minimum	0.0000	0.8725	0.9003	0.0003	0.0171
Maximum	0.0576	0.9842	0.9959	0.0070	0.6422

Table 2: Cross-correlation matrix between parameter estimates of conditional variance and covariance processes

	α_i	β_i	γ_i	ω_i
Panel A: Conditional variances				
β_i	−0.9577			
γ_i	−0.6982	0.8748		
ω_i	0.6329	−0.7647	−0.8361	
σ_i^2	0.1175	0.0036	0.2072	0.2485
	α_{ij}	β_{ij}	γ_{ij}	ω_{ij}
Panel B: Conditional covariances				
β_{ij}	−0.5162			
γ_{ij}	−0.0476	0.8800		
ω_{ij}	0.2913	−0.7427	−0.7047	
σ_{ij}	−0.1805	−0.0358	−0.1418	0.2018

Table 3: Characteristics of the baseline experiment

	Distribution	Parameters		Range	
		p	q	Minimum	Maximum
α	Beta	8	150	0	1
γ	Beta	150	2	0	1
σ^2 ($\times 100$)	Beta	3	3	0.01	0.27
ρ	Beta	3	3	0	0.6

Table 4: Estimates of the mean value of the GARCH parameters

	QMLE(1,1)	LSE(1,1)	ACE(10,4)	ACE(20,4)	ACE(40,4)
$N = 10$, $T = 1,000$					
Ω_p	0.1048	0.0328	0.0064	0.0062	0.0060
	(0.082)	(0.050)	(0.003)	(0.002)	(0.002)
$E[\alpha]$	0.0487	0.0381	0.0440	0.0412	0.0411
	(0.025)	(0.017)	(0.031)	(0.031)	(0.030)
$E[\beta]$	0.5710	0.8395	0.9309	0.9334	0.9349
	(0.277)	(0.182)	(0.036)	(0.033)	(0.033)
$E[\gamma]$	0.6353	0.8814	0.9758	0.9758	0.9758
	(0.285)	(0.181)	(0.009)	(0.006)	(0.006)
$N = 20$, $T = 1,000$					
Ω_p	0.1180	0.0336	0.0049	0.0047	0.0046
	(0.060)	(0.049)	(0.002)	(0.002)	(0.002)
$E[\alpha]$	0.0349	0.0318	0.0360	0.0350	0.0342
	(0.022)	(0.017)	(0.030)	(0.029)	(0.028)
$E[\beta]$	0.3588	0.8111	0.9361	0.9386	0.9388
	(0.294)	(0.248)	(0.035)	(0.032)	(0.032)
$E[\gamma]$	0.4200	0.8442	0.9744	0.9743	0.9742
	(0.304)	(0.245)	(0.010)	(0.006)	(0.007)
$N = 40$, $T = 1,000$					
Ω_p	0.0791	0.0478	0.0039	0.0037	0.0037
	(0.041)	(0.043)	(0.002)	(0.001)	(0.001)
$E[\alpha]$	0.0245	0.0231	0.0265	0.0334	0.0294
	(0.024)	(0.020)	(0.025)	(0.028)	(0.032)
$E[\beta]$	0.4247	0.6468	0.9446	0.9364	0.9418
	(0.280)	(0.309)	(0.034)	(0.032)	(0.032)
$E[\gamma]$	0.4671	0.6909	0.9725	0.9723	0.9724
	(0.287)	(0.311)	(0.016)	(0.008)	(0.004)
$N = 10$, $T = 5,000$					
Ω_p	0.0341	0.0095	0.0063	0.0061	0.0059
	(0.035)	(0.032)	(0.002)	(0.002)	(0.002)
$E[\alpha]$	0.0439	0.0276	0.0361	0.0344	0.0347
	(0.010)	(0.009)	(0.023)	(0.023)	(0.020)
$E[\beta]$	0.8395	0.9354	0.9397	0.9409	0.9411
	(0.130)	(0.118)	(0.024)	(0.024)	(0.021)
$E[\gamma]$	0.8836	0.9663	0.9760	0.9760	0.9759
	(0.132)	(0.122)	(0.003)	(0.003)	(0.003)

Note: The expected values of the simulated random parameters are $\Omega_p = 0.01$, $E[\alpha] = 0.0434$, $E[\beta] = 0.932$, and $E[\gamma] = 0.976$.

Table 5: Estimates of the mean value of the GARCH parameters

	QMLE(1, 1)	LSE(1, 1)	ACE(10, 4)	ACE(20, 4)	ACE(40, 4)
$\rho_{ij} \in [-0.1; 0.1]$					
Ω_p	0.0234	0.0066	0.0032	0.0031	0.0030
	(0.034)	(0.011)	(0.002)	(0.001)	(0.001)
$E[\alpha]$	0.0648	0.0424	0.0476	0.0486	0.0484
	(0.024)	(0.016)	(0.029)	(0.029)	(0.029)
$E[\beta]$	0.7665	0.9058	0.9274	0.9267	0.9273
	(0.234)	(0.090)	(0.034)	(0.031)	(0.030)
$E[\gamma]$	0.8394	0.9518	0.9762	0.9763	0.9763
	(0.238)	(0.084)	(0.010)	(0.006)	(0.004)
$\rho_{ij} \in [0.75; 0.9]$					
Ω_p	0.1611	0.1804	0.0081	0.0079	0.0077
	(0.102)	(0.128)	(0.005)	(0.003)	(0.004)
$E[\alpha]$	0.0422	0.0280	0.0403	0.0383	0.0360
	(0.026)	(0.026)	(0.034)	(0.032)	(0.031)
$E[\beta]$	0.4631	0.4136	0.9317	0.9350	0.9362
	(0.301)	(0.364)	(0.042)	(0.034)	(0.035)
$E[\gamma]$	0.5172	0.4574	0.9743	0.9744	0.9743
	(0.309)	(0.382)	(0.015)	(0.007)	(0.009)
***t* distribution** $(\nu = 4)$					
Ω_p	0.0773	0.0220	0.0062	0.0061	0.0059
	(0.079)	(0.076)	(0.003)	(0.003)	(0.003)
$E[\alpha]$	0.0567	0.0337	0.0432	0.0409	0.0409
	(0.028)	(0.023)	(0.037)	(0.037)	(0.032)
$E[\beta]$	0.6431	0.8772	0.9311	0.9341	0.9330
	(0.295)	(0.303)	(0.041)	(0.041)	(0.041)
$E[\gamma]$	0.7102	0.9224	0.9756	0.9756	0.9756
	(0.303)	(0.304)	(0.008)	(0.008)	(0.009)
Sk-*t* distribution $(\nu = 4,\ \lambda = -0.5)$					
Ω_p	0.0652	0.0258	0.0058	0.0059	0.0055
	(0.073)	(0.073)	(0.004)	(0.005)	(0.002)
$E[\alpha]$	0.0592	0.0319	0.0395	0.0395	0.0375
	(0.031)	(0.022)	(0.038)	(0.033)	(0.034)
$E[\beta]$	0.6800	0.8683	0.9355	0.9357	0.9372
	(0.293)	(0.307)	(0.047)	(0.046)	(0.035)
$E[\gamma]$	0.7504	0.9146	0.9758	0.9756	0.9757
	(0.301)	(0.313)	(0.014)	(0.020)	(0.004)

Note: The expected values of the simulated random parameters are $\Omega_p = 0.01$, $E[\alpha] = 0.0434$, $E[\beta] = 0.932$, and $E[\gamma] = 0.976$.

Table 6: Summary statistics on size and book-to-market portfolios

	Panel A: Univariate statistics on weekly returns					
	Mean (×100)	Volatility (×100)	Skewness	Kurtosis		
Small-Low (S-L)	0.248	2.685	−0.969	9.576		
Small-Medium (S-M)	0.351	1.984	−1.143	10.897		
Small-High (S-H)	0.381	1.947	−1.155	11.333		
Big-Low (B-L)	0.249	2.264	−0.282	5.673		
Big-Medium (B-M)	0.285	1.969	−0.355	5.709		
Big-High (B-H)	0.304	1.928	−0.361	5.644		
	Panel B: Covariance/correlation matrices					
	S-L	S-M	S-H	B-L	B-M	B-H
Small-Low (S-L)	0.072	0.050	0.047	0.049	0.040	0.037
Small-Medium (S-M)	0.943	0.039	0.037	0.035	0.032	0.030
Small-High (S-H)	0.895	0.968	0.038	0.032	0.030	0.030
Big-Low (B-L)	0.812	0.782	0.732	0.051	0.039	0.035
Big-Medium (B-M)	0.758	0.808	0.787	0.881	0.039	0.034
Big-High (B-H)	0.708	0.781	0.799	0.792	0.904	0.037

Table 7: Estimates of the mean value of the GARCH parameters

		QMLE(1,1)			ACE(10,4)			ACE(20,4)		
		Ω_p	$E[\alpha]$	$E[\gamma]$	Ω_p	$E[\alpha]$	$E[\gamma]$	Ω_p	$E[\alpha]$	$E[\gamma]$
S-L	–	0.037	0.148	0.950	0.050	0.106	0.921	0.050	0.107	0.919
S-L	S-M	0.024	0.137	0.952	0.043	0.106	0.903	0.042	0.107	0.902
S-L	S-H	0.021	0.137	0.955	0.039	0.099	0.904	0.038	0.100	0.903
S-L	B-L	0.020	0.131	0.961	0.036	0.099	0.924	0.035	0.099	0.923
S-L	B-M	0.017	0.119	0.958	0.026	0.080	0.928	0.026	0.081	0.926
S-L	B-H	0.022	0.103	0.944	0.024	0.078	0.933	0.023	0.078	0.932
S-M	–	0.017	0.130	0.958	0.030	0.098	0.910	0.029	0.099	0.910
S-M	S-H	0.015	0.129	0.961	0.030	0.112	0.962	0.026	0.093	0.914
S-M	B-L	0.013	0.120	0.964	0.025	0.093	0.924	0.025	0.094	0.921
S-M	B-M	0.011	0.111	0.964	0.016	0.073	0.943	0.016	0.073	0.941
S-M	B-H	0.015	0.097	0.951	0.016	0.072	0.945	0.015	0.071	0.945
S-H	–	0.014	0.129	0.965	0.025	0.087	0.919	0.025	0.087	0.918
S-H	B-L	0.011	0.118	0.966	0.022	0.087	0.927	0.022	0.088	0.924
S-H	B-M	0.010	0.109	0.966	0.015	0.070	0.944	0.012	0.069	0.951
S-H	B-H	0.014	0.096	0.954	0.016	0.072	0.944	0.015	0.071	0.944
B-L	–	0.011	0.124	0.985	0.038	0.126	0.923	0.038	0.127	0.920
B-L	B-M	0.009	0.111	0.980	0.026	0.100	0.929	0.026	0.101	0.926
B-L	B-H	0.012	0.094	0.965	0.025	0.100	0.926	0.025	0.101	0.924
B-M	–	0.008	0.105	0.983	0.017	0.079	0.952	0.018	0.081	0.947
B-M	B-H	0.011	0.088	0.967	0.016	0.076	0.949	0.016	0.077	0.946
B-H	–	0.016	0.076	0.957	0.018	0.074	0.949	0.017	0.073	0.948

Table 8: Test of forecasting ability

		ACE(10,4) vs. QMLE(1,1)			ACE(20,4) vs. QMLE(1,1)		
		DM_1	DM_2	DM_3	DM_1	DM_2	DM_3
S-L	–	2.585 ^{a}	2.290 ^{b}	1.514 ^{b}	2.566 ^{a}	2.249 ^{b}	1.509 ^{b}
S-L	S-M	1.708 ^{a}	2.348 ^{a}	0.739 ^{a}	1.682 ^{a}	2.369 ^{a}	0.725 ^{a}
S-L	S-H	1.763 ^{a}	2.899 ^{b}	0.766 ^{b}	1.695 ^{a}	2.597 ^{b}	0.756 ^{b}
S-L	B-L	0.569 ^{b}	1.904	0.181 ^{b}	0.551 ^{b}	1.875	0.174 ^{b}
S-L	B-M	0.788 ^{a}	2.809 ^{b}	0.207 ^{a}	0.774 ^{a}	2.417 ^{a}	0.210 ^{a}
S-L	B-H	0.320 ^{a}	0.656	0.096 ^{a}	0.328 ^{a}	0.590	0.099 ^{a}
S-M	–	0.927 ^{b}	1.367	0.363 ^{b}	0.899 ^{b}	1.279	0.356 ^{b}
S-M	S-H	0.932 ^{b}	1.562	0.385 ^{b}	0.875	1.259	0.376 ^{b}
S-M	B-L	0.390 ^{a}	1.628 ^{b}	0.096 ^{a}	0.387 ^{a}	1.733 ^{b}	0.094 ^{a}
S-M	B-M	0.432 ^{a}	1.059	0.108 ^{a}	0.405 ^{a}	0.580	0.107 ^{a}
S-M	B-H	0.057	−1.584	0.042	0.048	−1.752	0.041
S-H	–	0.994	1.185	0.513	0.971	1.107	0.507
S-H	B-L	0.391 ^{a}	2.042 ^{b}	0.087 ^{a}	0.383 ^{a}	1.651 ^{b}	0.090 ^{a}
S-H	B-M	0.439 ^{a}	1.524 ^{b}	0.104 ^{a}	0.323	−0.571	0.098 ^{a}
S-H	B-H	0.119	−0.818	0.051	0.092	−1.230	0.050
B-L	–	−0.438	−2.062	−0.005	−0.416	−1.972	−0.001
B-L	B-M	−0.100	0.092	0.020	−0.078	0.201	0.024
B-L	B-H	−0.302	−1.412	−0.038	−0.287 ^{b}	−1.366	−0.035
B-M	–	0.303	1.521	0.089	0.301	1.533	0.088
B-M	B-H	0.060	0.406	0.026	0.078	0.462	0.029
B-H	–	−0.031	−0.140	−0.003	0.017	0.001	0.008

Table 9: Summary statistics on portfolio returns and performance measures

	Strategy based on QMLE(1,1)	Strategy based on ACE(10,4)	Strategy based on ACE(20,4)
Risk aversion $\lambda = 2$			
Mean (×100)	4.595	6.586	5.775
Std dev. (×100)	26.209	28.436	26.601
Skewness	0.345	0.391	0.314
Kurtosis	3.615	4.022	3.637
Cumul. return	10.063	14.424	12.648
Sharpe ratio	0.171	0.227	0.213
Perf. fee (×100)	–	0.774	0.973
Risk aversion $\lambda = 5$			
Mean (×100)	1.870	2.666	2.342
Std dev. (×100)	10.482	11.371	10.637
Skewness	0.344	0.389	0.312
Kurtosis	3.616	4.023	3.638
Cumul. return	4.095	5.839	5.128
Sharpe ratio	0.167	0.224	0.209
Perf. fee (×100)	–	0.311	0.390
Risk aversion $\lambda = 10$			
Mean (×100)	0.961	1.359	1.197
Std dev. (×100)	5.239	5.683	5.316
Skewness	0.343	0.386	0.310
Kurtosis	3.616	4.024	3.638
Cumul. return	2.105	2.977	2.622
Sharpe ratio	0.160	0.217	0.202
Perf. fee (×100)	–	0.156	0.195

Figure 1: Empirical distribution of individual parameters

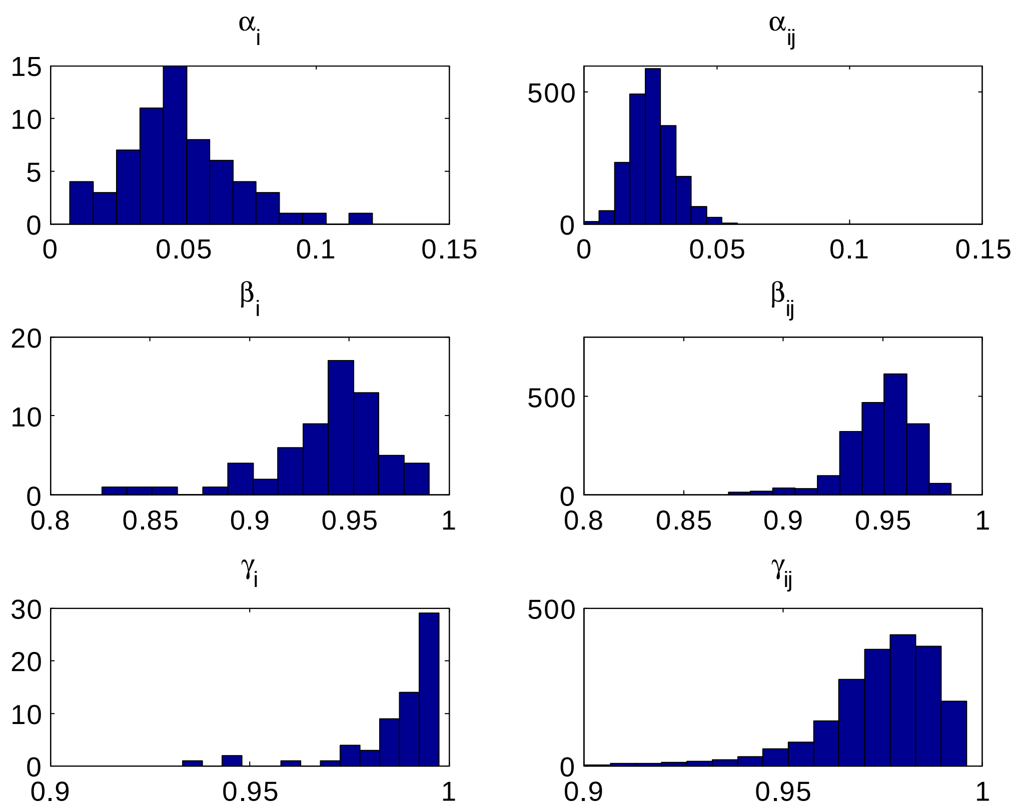

Figure 2: Empirical distribution of individual parameters (continued)