The Dynamics of Squared Returns Under Contemporaneous Aggregation of GARCH Models

Eric Jondeau*

(This version: June 2012)

Abstract

The paper investigates the properties of a portfolio composed of a large number of assets driven by a strong multivariate GARCH(1,1) process with heterogeneous parameters. The aggregate return is shown to be a weak GARCH process with a (possibly large) number of lags, which reflect the moments of the distribution of the individual persistence parameters. The paper describes consistent estimators of the aggregate return dynamic, based either on nonlinear least squares or on minimum distance, when only aggregate data is available. Monte-Carlo simulations demonstrate that the proposed aggregation-corrected estimator (ACE) performs very well under realistic sets of parameters drawn from U.S. equity data. Finally, the ACE is shown to outperform some competing estimators in forecasting the daily variance of U.S. portfolios.

Keywords: Aggregation, Heterogeneity, GARCH model, Volatility.

JEL classification: C13, C21, G17.

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

Generalized AutoRegressive Conditional Heteroskedasticity (or GARCH) models are commonly used to model the volatility of financial returns, such as stock returns, interest rates, or currency returns. Multivariate extensions allow to model the joint dynamic of several processes. These models are now used as standard tools for asset and risk management. In these areas, one key issue is the level of aggregation that should be considered for modeling the dynamics of the portfolio return. Two alternative approaches are readily available (see Andersen et al., 2005). On the one hand, the asset-level approach requires estimation of the joint behavior of all the assets in the portfolio, thus allowing one 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 renders this approach barely feasible. On the other hand, the portfolio-level approach requires only modeling of the portfolio return, but it is often inappropriate for scenario analysis. This approach is in general preferred for computational reasons. Another instance in which the portfolio-level approach is naturally adopted is the modeling of sectoral indices, asset classes, or risk factors. In such cases, it is often impossible to identify the underlying individual assets precisely, and therefore the properties of the resulting portfolio can be inferred from the aggregate process only.

An important, yet often overlooked, issue raised by the portfolio-level approach is that standard strong GARCH models are not closed under aggregation. This issue was first addressed by Nijman and Sentana (1996) for the aggregation of two independent processes with the same level of volatility persistence. They showed that the aggregate portfolio does not share the same properties as the individual assets. The properties of the resulting process, namely the weak GARCH model, have been studied by Drost and Nijman (1993), Nijman and Sentana (1996), and Meddahi and Renault (2004). Some issues raised by the estimation of a weak GARCH process have been addressed by Francq and Zakoian (2000) and Komunjer (2001). One particularly striking result is that, even in the simple case of two independent assets with the same volatility persistence, standard
estimation techniques fail to provide consistent estimates of the parameters driving the observed aggregate process (see Nijman and Sentana, 1996).

The contemporaneous aggregation of a large number of series has been already investigated in a closely related field, namely the ARMA processes. Earlier results have been provided by Granger and Morris (1976) and Granger (1980). They have established that aggregating ARMA(1,1) processes results in an ARMA process whose number of lags increases with the number of processes under aggregation. Granger (1980) shows that in general the aggregate process exhibits long-memory features. Since a weak GARCH process can be viewed as an ARMA process for squared returns, a similar result holds for the contemporaneous aggregation of GARCH(1,1) processes. Ding and Granger (1996) and Kazakevičius, Leipus, and Viano (2004) have investigated the aggregation of \( n \)-component GARCH processes, in which the individual volatility processes are driven by a common innovation. Zaffaroni (2007) extends this research to the aggregation of standard GARCH(1,1) processes in the context of an infinite number of series and demonstrates the asymptotic behavior of the variance of the aggregate process. He shows that conditional heteroskedasticity is preserved provided the degree of cross-sectional dependence between the assets is sufficiently strong.

This paper provides some new results on the dynamic properties of a portfolio composed of a large number of assets. The individual assets are assumed to be driven by a multivariate GARCH(1,1) variance process with heterogenous parameters. As expected, aggregate squared returns follow an ARMA-type process, reflecting the weak GARCH nature of the aggregate returns. In general, a large number of lags is required to fully capture the effect of the aggregation of heterogenous processes on the dynamic of the portfolio squared return. The first contribution of the paper is to show that the parameters driving the dynamic of the aggregate squared returns depend on the properties of the individual parameters. They can be interpreted in terms of moments of the cross-section distribution of the volatility persistence parameters. This result has important implications from

---

\(^1\)The main difference with Zaffaroni’s (2007) set-up relies on the form of the dependence across the assets. Zaffaroni considers two extreme cases, with purely idiosyncratic innovations or with common innovations. This paper considers the case of cross-correlated innovations, in a standard multivariate GARCH framework.
a modeling perspective. The fact that the aggregate squared return does not share the same specification as the individual components is known for a long time. However, there was no clear description of how the aggregate specification should look like. This paper is the first one to clearly establish the aggregate dynamic as a function of the individual ones.

Given the resulting dynamic of the aggregate process, the usual estimator, i.e., the Quasi Maximum Likelihood Estimator (QMLE) for a strong GARCH(1,1) process, is inconsistent. The paper describes two consistent estimators for the parameters of the aggregate process when only aggregate data are available. These estimators, based either on nonlinear least square or on minimum distance, take advantage of the relationship between the parameters of the aggregate volatility dynamics and the moments of the cross-section distribution of the volatility persistence in order to reduce the estimation burden. The proposed Aggregation-Corrected Estimators (ACE) do not rely on QMLE in order to circumvent the misspecification of the conditional volatility. Using Monte-Carlo simulations, the new estimators are shown to perform very well in finite samples for several realistic parameterizations. They provide unbiased estimates of the parameters driving the portfolio return dynamic. They also allow inferring the properties of the cross-section distribution of volatility persistence parameters, although they rely on aggregate data only.

Finally, the forecasting performances of competing estimators of the aggregate squared returns are compared to the so-called disaggregated estimator (DISE) based on the aggregation of individual volatility forecasts. The ACE significantly outperforms the DISE and the QMLE in forecasting the variance of a portfolio of U.S. equities.

The remainder of the paper is organized as follows: Section 2 briefly describes the multivariate GARCH model used for the individual assets and provides conditions required for the model to be well behaved. Section 3 considers the aggregation of individual volatility processes and provides the main results for the dynamics of aggregate squared returns. Section 4 describes the ACE, which correctly captures the dynamic of the aggregate volatility process. Section 5 evaluates the performances of the proposed estimator using both simulated data and actual data on U.S equities. Section 6 concludes.
2 The Model for Individual Assets

This section describes the set-up. The investment set is composed of a large number of individual assets, driven by correlated heterogenous GARCH processes. We assume a diagonal vec multivariate model, introduced by Bollerslev, Engle, and Wooldridge (1988), in which the dynamic of the unexpected returns, \( \mathbf{\varepsilon}_t = \{\varepsilon_{i,t}\}, \ t \in \mathbb{Z}, \ i = 1, \ldots, \mathbb{N}, \) is given by:

\[
\begin{align*}
\varepsilon_t &= H_{t}^{1/2} z_t, \\
X_t &= \varepsilon_t \varepsilon_t', \\
H_t &= \Omega + A \odot X_{t-1} + B \odot H_{t-1},
\end{align*}
\]

where \( z_t \) is the \((\mathbb{N}, 1)\) vector of idiosyncratic innovations with \( E[z_t] = 0 \) and \( V[z_t] = I_{\mathbb{N}}, \)
\( H_t = \{h_{ij,t}\} \) is the \((\mathbb{N}, \mathbb{N})\) conditional covariance matrix, and \( X_t \) is the \((\mathbb{N}, \mathbb{N})\) matrix of cross-products of unexpected returns with elements \( X_{ij,t} = \varepsilon_{i,t} \varepsilon_{j,t}. \)

Matrices \( \Omega = \{\omega_{ij}\}, \ A = \{\alpha_{ij}\}, \) and \( B = \{\beta_{ij}\} \) are \((\mathbb{N}, \mathbb{N})\) matrices of parameters, and \( \odot \) denotes the Hadamard product, such that \( \{A \odot B\}_{ij} = \alpha_{ij} \beta_{ij} \). In this specification, the individual variances and the covariances are all described by a GARCH(1,1) process with heterogenous parameters.

Straightforward manipulation of equations (1)–(3) gives a multivariate ARMA(1,1) model for the cross-product of unexpected returns:

\[
X_t = \Omega + \Gamma \odot X_{t-1} + v_t - B \odot v_{t-1},
\]

where \( v_t = X_t - H_t \) denotes the \((\mathbb{N}, \mathbb{N})\) matrix of covariance innovations and \( \Gamma = \{\gamma_{ij}\} = A + B \) is the \((\mathbb{N}, \mathbb{N})\) matrix of persistence parameters. The properties of \( X_t \) are rather well known. They have been studied in detail by Comte and Lieberman (2003) and Hafner.

---

2The “square root” matrix \( H_{t}^{1/2} \) in equation (1) is defined as \( H_t = H_{t}^{1/2} H_{t}^{1/2}'. \) There are alternative ways of defining the “square root” of a covariance matrix, such as the Cholesky decomposition or the spectral decomposition. We use the Cholesky decomposition in the empirical part of the paper.

3Alternative specifications to the diagonal vec model could be considered as well. The results of this paper could also be easily adapted to the vech model (Bollerslev, Chou, and Kroner, 1988) or BEKK model (Engle and Kroner, 1995).
(2003) in a similar framework. In particular, the innovation \( v_t \) is a weak white noise, with \( E[v_t] = 0, E[vech(v_t)vech(v_t)'] = \Sigma_v \), and \( E[vech(v_t)vech(v_s)'] = 0, \forall s \neq t \), although \( v_t \) is not an i.i.d. sequence.

In the sequel, we make the following set of assumptions, closely related to those adopted in Hafner (2003).

**Assumption 1 (Positivity)** *The matrices \( \Omega, A, \) and \( B \) are symmetric positive semi-definite almost surely.*

**Assumption 2 (Stationarity)** *\( \Omega < +\infty \) and all eigenvalues of \( \Gamma = A + B \) are smaller than 1 in modulus.*

**Assumption 3 (Innovation process)** *The innovation process \( \{z_t\}, t \in \mathbb{Z} \), is i.i.d. and its distribution has finite fourth moments.*

Assumption 1 ensures that the conditional covariance matrix \( H_t \) is symmetric positive semi-definite almost surely (Ding and Engle, 1994). Assumption 2 ensures that the multivariate GARCH process \( \varepsilon_t \) is covariance stationary (Engle and Kroner, 1995). Under Assumptions 1 and 2, the multivariate model is well behaved, with a positive semi-definite covariance matrix at each date \( t \). Assumption 3 implies that the time dependency in the model is fully captured by the covariance matrix \( H_t \), so that the innovation \( z_t \) is an i.i.d. vector.

### 3 Aggregation

This section describes the properties of a portfolio composed of \( N \) risky assets with a strong GARCH(1,1) conditional covariance matrix. A risk-free asset may be included in the portfolio, provided \( \varepsilon_{i,t} \) is defined as the unexpected excess return over the risk-free rate. Portfolio weights are denoted by \( w = (w_1, \ldots, w_N)' \). The weights satisfy \( w_i = O(N^{-1}) \), with \( 0 < w_i < 1, \forall i \), and \( \sum_{i=1}^{N} w_i = 1 \).

---

4Portfolio weights are assumed to be constant over time. This assumption typically corresponds to a strategic allocation problem. We could also consider the case of an equally-weighted portfolio, with \( w_i = 1/N, \forall i \). The case for time-varying weights would raise new issues, in particular if their dynamics depend on the parameters of the individual squared returns. The investigation of this extension is beyond the scope of this paper and left for future research.
\[ \varepsilon^{(w)}_{p,t} = \sum_{i=1}^{N} w_i \varepsilon_{i,t}, \text{ i.e., the cross-section mean of } \varepsilon_{i,t}, \text{ where the exponent } (w) \text{ indicates that the mean is computed over the } N \text{ assets with weights } w. \] The aggregate squared return is then defined as:

\[ X^{(w)}_{p,t} = (\varepsilon^{(w)}_{p,t})^2 = \left( \sum_{i=1}^{N} w_i \varepsilon_{i,t} \right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j X_{ij,t}. \]

In the following, we will omit the exponent \((w)\) from the aggregate variables to save on notations. For a given weight vector, the cross-section moment of order \(k\) for parameter \(\vartheta\) driving the covariance matrix dynamic is denoted by \(\tilde{E}^{(w)}[\vartheta^k] \equiv m_k^{(\vartheta)} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \vartheta_{ij}^k\), where the tilde is used to avoid confusion with the time series average. Under Assumptions 1 to 3, the following proposition holds regarding the dynamic of the aggregate squared returns.

**Proposition 1**

Let \(\{ \varepsilon_t \}, t \in \mathbb{Z}, \) be an \(N\)-dimensional strong GARCH(1,1) process defined by equations (1) – (3), with \(X_t = \varepsilon_t \varepsilon'_t\). Under Assumptions 1 to 3, the aggregate squared return \(X_{p,t}\) satisfies in the limit (as \(T \to \infty\)):

\[ X_{p,t} = h_{p,t} + v_{p,t}, \quad (5) \]

\[ h_{p,t} = \Omega_p + \sum_{k=1}^{\infty} \Psi_k X_{p,t-k} + \sum_{k=1}^{\infty} \Phi_k h_{p,t-k}, \quad (6) \]

where \(h_{p,t}\) is the linear projection of \(X_{p,t}\) on \(\mathcal{F}_t = \{1, X_{p,t-1}, X_{p,t-2}, \cdots\} \). The aggregate innovation, \(v_{p,t} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \nu_{ij,t}\), is a weak white noise with zero mean and constant variance. \(\Omega_p\) is defined in Appendix 1. The aggregate squared return also satisfies the infinite ARMA representation:

\[ X_{p,t} = \Omega_p + \sum_{k=1}^{\infty} \Lambda_k X_{p,t-k} + v_{p,t} - \sum_{k=1}^{\infty} \Phi_k v_{p,t-k}. \quad (7) \]
Parameters $\Psi_k$, $\Phi_k$, and $\Lambda_k$ in equations (6) and (7) denote the following cross-section means:

$$
\Psi_k = \tilde{E}(w)[\psi_k], \quad \psi_1 = \alpha, \quad \psi_k = (\lambda_{k-1} - \Lambda_{k-1}) \alpha, \quad k = 1, 2, \cdots
$$

$$
\Phi_k = \tilde{E}(w)[\phi_k], \quad \phi_1 = \beta, \quad \phi_k = (\lambda_{k-1} - \Lambda_{k-1}) \beta,
$$

$$
\Lambda_k = \tilde{E}(w)[\lambda_k], \quad \lambda_1 = \gamma, \quad \lambda_k = (\lambda_{k-1} - \Lambda_{k-1}) \gamma = \psi_k + \phi_k.
$$

Proof: See Appendix 1.

There are several comments regarding Proposition 1. First, the term $h_{p,t}$ in equation (6) cannot be interpreted as the conditional variance process of the aggregate return (Komunjer, 2001). Therefore, the main equation of interest is the dynamic of the aggregate squared return (equation (7)). As this equation indicates, the aggregate return is driven by a weak GARCH process with an infinite number of lags. In fact, depending on the characteristics of the individual processes, the aggregate ARMA process may be of a low order. For instance, assume that parameters $\{\alpha_{ij}\}$ and $\{\beta_{ij}\}$ are heterogeneous across assets, but that all of the $\{\gamma_{ij}\}$ are equal to a single value, $\gamma_{ij} = \bar{\gamma}$, $\forall i, j$, such that all the assets share the same volatility persistence. In such a model, the aggregate squared return simply writes as an ARMA(1,1) process:

$$
X_{p,t} = \Omega_p + \Lambda_1 X_{p,t-1} + v_{p,t} - \Phi_1 v_{p,t-1},
$$

with $\Lambda_1 = \bar{\gamma}$, $\Phi_1 = \tilde{E}(w)[\beta]$ and $\Lambda_k = \Phi_k = 0$, $\forall k > 1$ in equation (7) (see Appendix 1). A similar ARMA(1,1) process was obtained by Nijman and Sentana (1996, equation (10)) in the case of two independent processes with the same persistence parameter. This example clearly illustrates that the additional lags in equation (7) arise from the heterogeneity in the persistence parameter. Proposition 1 also helps understanding why estimating a weak GARCH(1,1) model for the aggregate process cannot correct for the heterogeneity across the assets. Indeed, it provides consistent estimates of the aggregate parameters ($\Omega_p$, $\Lambda_1$, $\Phi_1$) in equation (7) only if the additional lags $\Lambda_k$ and $\Phi_k$ are all equal to 0, $\forall k > 1$, i.e., if all of the $\{\gamma_{ij}\}$ are equal to a single value.
Second, the dynamic of the aggregate squared return is driven by the properties of the sequences of parameters \( \{\Lambda_k\} \) and \( \{\Phi_k\} \). Contemplating equation (7) reveals that the parameters \( \{\Lambda_k\} \) are related to the moments of the cross-section distribution of the persistence parameters \( \{\gamma_{ij}\} \). Indeed, straightforward computation shows that:

\[
\Lambda_k = \bar{E}(w)[\lambda_k] = \bar{E}(w)[\gamma^k] - \sum_{r=1}^{k-1} \Lambda_r \bar{E}(w)[\gamma^{k-r}], \quad k \geq 1. \tag{9}
\]

As the \( \{\gamma_{ij}\} \) are bounded between 0 and 1 (under Assumption 3), the cross-section moments of order \( k \), 

\[
m_k^{(\gamma)} = \bar{E}(w)[\gamma^k],
\]

are decreasing to 0 as \( k \to \infty \), with \( 0 \leq \cdots \leq m_s^{(\gamma)} \leq m_{s-1}^{(\gamma)} \leq \cdots \leq m_1^{(\gamma)} < 1 \), \( \forall s > 1 \). As equation (9) clearly indicates, \( \Lambda_k \) is a linear combination of terms of the form \( \{m_k^{(\gamma)}, m_{k-1}^{(\gamma)}, \cdots, (m_1^{(\gamma)})^k\} \) with all the terms involving a sum of powers of \( \gamma \) equal to \( k \). For instance, the first \( \Lambda_k \) are:

\[
\begin{align*}
\Lambda_1 &= m_1^{(\gamma)}, \\
\Lambda_2 &= m_2^{(\gamma)} - (m_1^{(\gamma)})^2, \\
\Lambda_3 &= m_3^{(\gamma)} - 2m_1^{(\gamma)}m_2^{(\gamma)} + (m_1^{(\gamma)})^3, \\
\Lambda_4 &= m_4^{(\gamma)} - 2m_1^{(\gamma)}m_3^{(\gamma)} + 3(m_1^{(\gamma)})^2m_2^{(\gamma)} - (m_2^{(\gamma)})^2 - (m_1^{(\gamma)})^4.
\end{align*}
\]

Straightforward, but tedious computation also shows that: (1) the \( \{\Lambda_k\} \) are positive, given that \( \gamma \) is bounded between 0 and 1; (2) the \( \{\Lambda_k\} \) are decreasing to 0 as \( k \) increases to \( \infty \); (3) the limit sum of the \( \{\Lambda_k\} \) parameters is given by:

\[
\sum_{k=1}^{\infty} \Lambda_k = \sum_{k=1}^{\infty} \left[ \bar{E}(w)[\gamma^k] - \sum_{r=1}^{k-1} \Lambda_r \bar{E}(w)[\gamma^{k-r}] \right] = \frac{\sum_{k=1}^{\infty} \bar{E}(w)[\gamma^k]}{1 + \sum_{k=1}^{\infty} \bar{E}(w)[\gamma^k]}, \tag{10}
\]

so that \( 0 \leq \sum_{k=1}^{\infty} \Lambda_k < 1 \). Therefore, there exist an \( \varepsilon > 0 \) and a \( K_\Lambda \in \mathbb{N} \), such that \( 0 \leq \Lambda_k \leq \varepsilon \), \( \forall k > K_\Lambda \). Thus, we can select the number of lags in equation (7) in such a way that the contribution of the non-included terms \( \Lambda_k \), \( \forall k > K_\Lambda \), is made negligible relative to that of the included part.

It is worth noticing that the estimates of parameters \( \{\Lambda_k\} \) from aggregate equation (7) can be used to recover the cross-section moments of the individual persistence parameters
\{\gamma_{ij}\}.^{5} \text{ The first moments of } \{\gamma_{ij}\} \text{ are:}

\begin{align*}
\tilde{E}^{(w)}[\gamma] &= \Lambda_1, \\
\tilde{V}^{(w)}[\gamma] &= \Lambda_2, \\
\tilde{S}^{(w)}[\gamma] &= (\Lambda_3 - \Lambda_1\Lambda_2) / (\Lambda_2)^{3/2}, \\
\tilde{K}^{(w)}[\gamma] &= (\Lambda_4 - 2\Lambda_1\Lambda_3 + \Lambda_1^2\Lambda_2 + \Lambda_2^2) / (\Lambda_2)^2,
\end{align*}

where \(\tilde{V}^{(w)}\), \(\tilde{S}^{(w)}\), and \(\tilde{K}^{(w)}\) denote the cross-section variance, skewness, and kurtosis, respectively. As the knowledge of \(\{\Lambda_k\}_{k=1}^{K_A}\) is equivalent to the knowledge of \(\{m^{(\gamma)}_k\}_{k=1}^{K_A}\), we can deduce the cross-section properties of the individual persistence parameter, even when only aggregate data is available.

Parameters \(\{\Phi_k\}\) are defined in a similar way:

\begin{equation}
\Phi_k = \tilde{E}^{(w)}[\phi_k] = \tilde{E}^{(w)}[\beta\gamma^{k-1}] - \sum_{r=1}^{k-1} \Phi_r \tilde{E}^{(w)}[\gamma^{k-r}], \quad k \geq 1.
\end{equation}

Parameter \(\Phi_1\) corresponds to the cross-section mean of \(\{\beta_{ij}\}\). Subsequent parameters \(\Phi_k\) pertaining to lags \(v_{pt-k}\) correspond to cross-section co-moments between \(\{\beta_{ij}\}\) and powers of \(\{\gamma_{ij}\}\). For instance, \(\Phi_2 = \tilde{E}^{(w)}[\beta\gamma] - \tilde{E}^{(w)}[\beta]\tilde{E}^{(w)}[\gamma]\) is the cross-section covariance between \(\{\beta_{ij}\}\) and \(\{\gamma_{ij}\}\). Consequently, estimates of parameters \(\{\Phi_k\}\) provide information about the cross-section joint distribution of \(\{\beta_{ij}\}\) and \(\{\gamma_{ij}\}\).

As above, we can compute the limit sum of the \(\{\Phi_k\}\) parameters as:

\begin{equation}
\sum_{k=1}^{\infty} \Phi_k = \sum_{k=1}^{\infty} \left[ \tilde{E}^{(w)}[\beta\gamma^{k-1}] - \sum_{r=1}^{k-1} \Phi_r \tilde{E}^{(w)}[\gamma^{k-r}] \right] = \frac{\sum_{k=1}^{\infty} \tilde{E}^{(w)}[\beta\gamma^{k-1}]}{1 + \sum_{k=1}^{\infty} \tilde{E}^{(w)}[\gamma^k]},
\end{equation}

so that \(0 < \sum_{k=1}^{\infty} \Phi_k < \sum_{k=1}^{\infty} \Lambda_k < 1\), as all the \(\{\beta_{ij}\}\) and \(\{\gamma_{ij}\}\) lie in \((0, 1)\), with \(\beta_{ij} \leq \gamma_{ij}\). Again, a number of lags \(K_{\Phi}\) can be selected in equation (7) in such a way that the terms \(\Phi_k, \forall k > K_{\Phi}\), are made negligible.

\(^{5}\)A similar interpretation has been proposed by Robinson (1978) and Granger (1980) for the aggregation of autoregressive processes and by Lewbel (1994) for the aggregation of linear dynamic processes.
With similar computation, we obtain that in the limit the aggregate constant term \( \Omega_p \) is defined as:

\[
\Omega_p = \bar{E}^{(w)}[\omega] + \sum_{k=1}^{\infty} \bar{E}^{(w)}[(\lambda_k - \Lambda_k) \omega] = \frac{\sum_{k=1}^{\infty} \bar{E}^{(w)}[\omega \gamma^{k-1}]}{1 + \sum_{k=1}^{\infty} \bar{E}^{(w)}[\gamma^k]}.
\] (17)

Given these expressions, it is possible to investigate the conditions ensuring the stationarity of the aggregate unexpected return process \( \varepsilon_{p,t} \) as \( (K_\Lambda, K_\Phi) \rightarrow \infty \). Covariance stationarity requires that \( V[\varepsilon_{p,t}] = E[X_{p,t}] < \infty \). The aggregate process (7) can be rewritten as:

\[
\Lambda(L)X_{p,t} = \Omega_p + \Phi(L)v_{p,t},
\] (18)

where \( \Lambda(L) = 1 - \Lambda_1 L - \Lambda_2 L^2 - \cdots \) and \( \Phi(L) = 1 - \Phi_1 L - \Phi_2 L^2 - \cdots \), with \( L \) the lag operator. The condition for invertibility of \( \Lambda(L) \) is that the roots of \( |\Lambda(z)| = 0 \) lie outside the unit circle, which implies \( \Lambda(1) > 0 \) or, equivalently, \( \sum_{k=1}^{\infty} \Lambda_k < 1 \). This condition is guaranteed by equation (10). Therefore, we can rewrite equation (18) as:

\[
X_{p,t} = \Lambda(1)^{-1}\Omega_p + \delta(L)v_{p,t},
\]

with \( \delta(L) = \Lambda(L)^{-1}\Phi(L) \) and \( \delta_0 = 1 \). We obtain that the unconditional variance of the aggregate return is:

\[
E[X_{p,t}] = \frac{\Omega_p}{1 - \sum_{k=1}^{\infty} \Lambda_k} = \sum_{k=1}^{\infty} \bar{E}^{(w)}[\omega \gamma^{k-1}] = \bar{E}^{(w)}\left[\frac{\omega}{1 - \gamma}\right] = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j h_{ij},
\]

where \( h_{ij} = \omega_{ij}/(1 - \gamma_{ij}) \) denotes the unconditional covariance between assets \( i \) and \( j \). Under Assumptions 1 and 2, the \( \{h_{ij}\} \) are finite and so is \( E[X_{p,t}] \). Therefore, \( \varepsilon_{p,t} \) is a covariance stationary process.

### 4 Estimation of the Aggregate Model

In this section, we describe two estimation procedures for the aggregate process in the case of contemporaneous aggregation, when no information is available about the individual assets.
Imposing a strong GARCH(1,1) structure to the aggregate squared return:

\[ X_{p,t} = h_{p,t} + v_{p,t}, \]  
\[ h_{p,t} = \Omega_p + \Psi_1 X_{p,t-1} + \Phi_1 h_{p,t-1}, \]

will yield inconsistent estimators in general. Indeed the parameters \( \Psi_1 \) and \( \Phi_1 \) will correspond to the theoretical values \( \tilde{E}^{(w)}[\alpha] \) and \( \tilde{E}^{(w)}[\beta] \) only if the persistence parameters are in fact homogeneous across assets, i.e. \( \gamma_{ij} = \tilde{\gamma}, \forall i,j \). As a consequence, except under very special circumstances, the aggregate squared return cannot be expected to be a strong GARCH(1,1) process, so that the QMLE is an inconsistent estimator. This result has been already pointed out by Nijman and Sentana (1996) in the case of two independent individual processes: “the QML estimator is approximately consistent in some cases and clearly inconsistent in others.” Komunjer (2001) has established in their context the reasons for the inconsistency of the QMLE, which also apply to this paper.\(^6\)

### 4.1 The Least-Square Estimator (LSE)

A first approach designed for the estimation of a weak GARCH\((p,q)\) process has been proposed by Francq and Zakoian (2000). This estimator, called Least-Square Estimator (LSE), explicitly acknowledges the possible misspecification of the first two moments and therefore does not rely on QMLE. It is based on the minimization of the sum of the squared residuals of the aggregate squared return process. The definition of the resulting Least-Square Aggregation-Corrected Estimator (LS-ACE) is given below.

**Definition 1** The Least-Square Aggregation-Corrected Estimator LS-ACE\((K_\Lambda, K_\Phi)\), denoted by \( \theta_{LS}^{AC} = (\Omega_p, \Lambda_1, \cdots, \Lambda_{K_\Lambda}, \Phi_1, \cdots, \Phi_{K_\Phi})' \), is defined as:

\[ \theta_{LS}^{AC} \in \arg \min_\theta Q_T(\theta) = \frac{1}{T} \sum_{t=1}^T (v_{p,t}(\theta))^2, \]  

\(^6\)Komunjer (2001) also proposes a new QMLE for weak GARCH processes, designed to account for the deficiencies of the standard estimator in this context. Unfortunately, the large-sample properties of this new estimator are still quite disappointing, resulting in a severe underestimation of \( \alpha \), similar to that obtained with the Nijman and Sentana (1996) approach.
where
\[ v_{p,t}(\theta) = X_{p,t} - \Omega_p - \sum_{k=1}^{K_\Lambda} \Lambda_k X_{p,t-k} + \sum_{k=1}^{K_\Phi} \Phi_k v_{p,t-k}(\theta). \] (22)

The expressions for \( \Lambda_k \) and \( \Phi_k \) are given in Proposition 1, the expression for \( \Omega_p \) is in Appendix 1.

The consistency and asymptotic normality of this estimator have been proved in Francq and Zakoian (2000). They also describe how to construct the asymptotic covariance matrix of the estimator, even under situations where \( \varepsilon_t \) is the innovation of a preliminary ARMA filter. In our context, the asymptotic distribution is given by:

\[ \sqrt{T}(\hat{\theta}_{AC}^{LS} - \theta_0) \sim N(0, V^{LS}), \]

with asymptotic covariance matrix \( V^{LS} = J^{-1}IJ^{-1} \), where

\[ I = \lim_{T \to \infty} V \left[ \sqrt{T} \frac{\partial}{\partial \theta} Q_T(\theta_0) \right] \quad \text{and} \quad J = \lim_{T \to \infty} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_T(\theta_0) \right) \]

are evaluated at the true parameter value \( \theta_0 \).

4.2 The Minimum Distance Estimator (MDE)

A second estimator, which does not rely on QMLE, has been proposed by Baillie and Chung (2001) in the context of a GARCH(1,1) process. This minimum distance estimator (MDE) is motivated by the idea of replicating some properties of the data. A typical example of such properties is the autocorrelogram of the aggregate squared returns. The objective of the estimator is then to minimize the distance between the theoretical autocorrelations and their empirical counterparts. It is described more precisely in the following definition.

**Definition 2** The Minimum-Distance Aggregation-Corrected Estimator MD-ACE(\( K_{\Lambda}, K_{\Phi}, K_\rho \)), denoted by \( \theta_{AC}^{MD} = (\Omega_p, \Lambda_1, \cdots, \Lambda_{K_\Lambda}, \Phi_1, \cdots, \Phi_{K_\Phi})' \), is defined as:

\[ \theta_{AC}^{MD} \in \arg\min_{\theta} \ (\rho(\theta) - \hat{\rho})' W (\rho(\theta) - \hat{\rho}), \] (23)
where \( \rho(\theta) = (\rho_1(\theta), \ldots, \rho_K(\theta))^\prime \) and \( \hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_K)^\prime \) denote the first \( K \) theoretical autocorrelations of an ARMA\((K_\Lambda, K_\Phi)\) process and their empirical counterparts, respectively, and \( W \) is a weighting matrix. The constant term is estimated by \( \Omega_p = (1 - \sum_{k=1}^{K_\Lambda} \Lambda_k) E[X_{p,t}] \).

A usual choice for the weighting matrix \( W \) is a consistent estimate of the inverse of the covariance matrix of \( \hat{\rho} \). The theoretical autocorrelations \( \rho(\theta) \) are obtained using the approach described in Brockwell and Davis (1991, section 3.3). Baillie and Chung (2001) give the asymptotic distribution of the MDE:

\[
\sqrt{T}(\hat{\theta}_{AC}^{MD} - \theta_0) \sim N(0, V^{MD}),
\]

with \( V^{MD} = (D' C^{-1} D)^{-1} \), where \( D = \partial \rho(\theta)/\partial \theta \) is evaluated at the true parameter value \( \theta_0 \) and \( C = \{c_{ij}\} \), \( c_{ij} = \sum_{k=1}^{\infty} (\rho_{k+i} - \rho_{k-i} - 2 \rho_i \rho_k) (\rho_{k+j} - \rho_{k-j} - 2 \rho_j \rho_k) \), is the asymptotic covariance matrix of the sample autocorrelations (Bartlett, 1946). As the innovation process \( v_{p,t} \) is not i.i.d., we need to estimate the matrix \( C \) in a consistent way. For this purpose, we follow the approach described by Baillie and Chung (2001), based on the Newey and West (1987) procedure.

### 4.3 Parametrization of the parameter cross-section distribution

As we discussed in Section 3, the parameters \( \{\Lambda_k\} \) and \( \{\Phi_k\} \) decrease to 0 as \( k \) goes to infinity. Therefore, the numbers of lags \( K_\Lambda \) and \( K_\Phi \) can be selected to ensure that the contribution of the non-included part is made negligible relative to that of the included part, yielding consistent estimators of \( \theta_{AC} \). However, solving problems (21) and (23) with a large number of lags and unrestricted parameters would be inefficient, as the sequences of parameters \( \{\Lambda_k\} \) and \( \{\Phi_k\} \) are known to be restricted by their relation to the moments and co-moments of the cross-section distribution of the parameters.\(^7\)

\(^7\)It turns out that, in typical financial applications, the persistence parameters are rather concentrated around values slightly below 1. This implies that the convergence of the \( \{\Lambda_k\} \) parameters toward 0 is fast as \( k \) increases. As the subsequent Monte-Carlo simulations will show, estimation of equation (7) provides essentially unbiased estimates of the aggregate parameters even for a relatively small number of lags.
In principle, these relations can be explicitly incorporated in the estimation through a flexible parametric representation of the cross-section distribution. Assume that the joint distribution of \( \{\beta_{ij}, \gamma_{ij}\} \) is known and depends on a new set of parameters \( \tilde{\theta} \) of smaller dimension. Then the sequences of parameters \( \{\Lambda_k\} \) and \( \{\Phi_k\} \) can be obtained as functions of \( \tilde{\theta} \), so that the innovation in equation (22) can be rewritten as:

\[
v_{p,t}(\tilde{\theta}) = X_{p,t} - \Omega_p - \sum_{k=1}^{K_A} \Lambda_k(\tilde{\theta}) X_{p,t-k} + \sum_{k=1}^{K_A} \Phi_k(\tilde{\theta}) v_{p,t-k}(\tilde{\theta}).
\]

A similar framework has been described by Zaffaroni (2004b) for his goodness-of-fit test in an ARCH(\( \infty \)) process.

We start with the sequence of aggregate parameters \( \{\Lambda_k\} \), which are related to the moments of the persistence parameters \( \{\gamma_{ij}\} \) (equation (9)). As in Robinson (1978), Granger (1980), and Gonçalves and Gouriéroux (1988), we make the following assumption regarding the cross-section distribution of \( \{\gamma_{ij}\} \).

**Assumption 4** The cross-section distribution of the persistence parameter \( \{\gamma_{ij}\} \) is described by a Beta distribution with parameters \( p \) and \( q \):

\[
f_\gamma(\gamma) = \frac{\gamma^{p-1} (1 - \gamma)^{q-1}}{B(p, q)},
\]

where \( p, q \in (0, \infty) \) and \( B(\cdot, \cdot) \) is the Beta function.

This parametric distribution covers the range of values \((0, 1)\), which represents the admissible interval for \( \gamma \). The Beta distribution is able to reproduce a wide range of distribution shapes, like a leftward asymmetric bell shape (for \( 1 < q < p \)) or a continuously increasing distribution (for \( 0 < q < 1 < p \)).\(^9\) The non-central moments of the Beta

\(^8\) \( \Omega_p \) can be estimated directly as a single free parameter.

\(^9\) The case \( q < 1 \) is known in the literature on the aggregation of ARMA processes as the “long memory” case because it generates a sequence of non-central moments that is not absolutely summable, therefore inducing a long-memory pattern in the underlying aggregate series (see Granger, 1980, Abadir and Talmain, 2002, Zaffaroni, 2004a). In the GARCH framework, long memory is ruled out (see Zaffaroni, 2004b, 2007).
distribution with parameters $p$ and $q$ are given by:

$$
\tilde{E}(w) [\gamma^k] = \frac{B(p + k, q)}{B(p, q)} = \frac{\Gamma(p + q) \Gamma(p + k)}{\Gamma(p) \Gamma(p + q + k)},
$$

(25)

where $\Gamma(\cdot)$ is the Gamma function. For given values of $p$ and $q$, the non-central moments of $\{\gamma_{ij}\}$ are obtained by equation (25) and the parameters $\{\Lambda_1, \cdots, \Lambda_{K_\gamma}\}$ are directly deduced from equation (9), giving all the autoregressive parameters of equation (7).

As stated above, the sequence of aggregate parameters $\{\Phi_k\}$ is related to the co-moments of the joint distribution of $\{\gamma_{ij}\}$ and $\{\beta_{ij}\}$ (equation (15)). Parameterizing these co-moments is more problematic, because it requires some assumptions about the kind of dependence between $\{\gamma_{ij}\}$ and $\{\beta_{ij}\}$. To our knowledge, there is no simple way to define a joint Beta distribution, whose dependence structure can be described as a free parameter. As a consequence, to avoid excessive restrictions on their joint distribution, we estimate the cross-section mean $\Phi_0 = \tilde{E}(w) [\beta]$ and the sequence of co-moments $\Phi_k$, $k = 1, \cdots, K_\Phi$, directly as free parameters. The parameter set for the LS-ACE and MD-ACE estimators therefore reduces to $\tilde{\theta} = (\Omega, p, q, \Phi_1, \cdots, \Phi_{K_\Phi})'$.

5 Performance of the Estimators

This section aims at evaluating the finite-sample properties of the estimators. This evaluation is based on Monte-Carlo simulations, which reproduce the properties of a large sample of U.S. equities. We show that the ACE provides unbiased estimates of the parameters driving the aggregate squared returns. Finally, we report some estimates of the aggregate return dynamic using U.S. aggregate data only. The comparison of the competing estimators on real data confirms that the ACE outperforms the usual estimators.

5.1 Calibration Based on U.S. Equities

The calibration of the individual parameters for the Monte-Carlo simulations is based on a sample of 75 U.S. companies between January 1988 and December 2010 for a total of
In the following, we use the generic notation $\vartheta_i \equiv \vartheta_{ii}$ for the parameters. The individual variance parameters $(\omega_i, \alpha_i, \gamma_i)$ of each of the 75 individual stocks and the covariance parameters $(\omega_{ij}, \alpha_{ij}, \gamma_{ij})$ of each of the 2,775 pairs of stocks are estimated using the flexible GARCH approach of Ledoit, Santa-Clara, and Wolf (2003), which ensures the positive semi-definiteness of the covariance matrices. This subsection aims at describing the main properties of the parameter estimates, which will be used for the simulations.

Table 1 reports some summary statistics on the parameter estimates of the individual conditional variances and covariances, based on ML estimation. The cross-section mean estimates of the variance parameters $\alpha_i$, $\beta_i$, and $\gamma_i$ are 0.049, 0.943, and 0.992, respectively (Panel A). The large value obtained for the persistence parameters $\gamma_i$ is consistent with that reported in the empirical literature on stock returns. Regarding the conditional covariance parameters, the mean estimates of $\alpha_{ij}$, $\beta_{ij}$, and $\gamma_{ij}$ are 0.035, 0.942, and 0.977, respectively (Panel B). To further investigate the characteristics of the individual parameters, a Beta distribution is adjusted to each set of parameters ($\alpha_i$, $\alpha_{ij}$, $\beta_i$, $\beta_{ij}$, $\gamma_i$, $\gamma_{ij}$). The table reports the estimates of parameters $p$ and $q$ for each distribution. As it appears clearly, there are some significant differences between the characteristics of the variance parameters ($\alpha_i$, $\beta_i$, $\gamma_i$) and the covariance parameters ($\alpha_{ij}$, $\beta_{ij}$, $\gamma_{ij}$). Given the dominant role played by the covariance terms in the aggregated variance (97.3% of the terms involved), we mostly focus on in the sequel on the covariance parameters. Figure 1 displays the histogram of the parameters $\{\vartheta_{ij}\}$ and the estimated Beta distribution $f_\vartheta(\vartheta)$, for $\vartheta = \alpha$, $\beta$, and $\gamma$. As the figure illustrates, the fit of the actual parameters is very good for all the parameter sets. We notice that the range of values is in fact rather narrow. All the estimates of $\alpha_{ij}$ range between 0.01 and 0.07, all the estimates of $\gamma_{ij}$ are above 0.94.

For simulation purpose, another important property is the dependence between the individual parameters. Clearly, we cannot simulate $\alpha$ and $\beta$ independently from each
other because it could imply values of $\gamma$ larger than 1. Similarly, simulating $\omega$ and $\gamma$ independently from each other could generate extremely erratic values for $h = \omega/(1-\gamma)$, when $\gamma$ is close to 1. To address this issue, we consider the correlation between the individual parameters estimated on U.S. equities (Panel C). Regarding first the parameters $\alpha$, $\beta$, and $\gamma$, the table reveals that $\gamma_{ij}$ is positively and strongly correlated with $\beta_{ij}$ but weakly correlated with $\alpha_{ij}$ (0.88 and −0.205, respectively). Therefore, we proceed as follows in the Monte-Carlo experiments: parameters $\alpha_i$ and $\gamma_i$ are simulated from independent Beta distributions with the parameters $p_i$ and $q_i$ reported in Panel A. Parameters $\alpha_{ij}$ and $\gamma_{ij}$ are drawn from independent Beta distributions with the parameters $p_{ij}$ and $q_{ij}$ reported in Panel B. Then, we define $\beta_i = \gamma_i - \alpha_i$ and $\beta_{ij} = \gamma_{ij} - \alpha_{ij}$.

The calibration of the unconditional variances and covariances also requires additional information about the constant terms $\omega_i$ and $\omega_{ij}$, the unconditional variances $h_i = \omega_i/(1-\alpha_i-\beta_i)$ and covariances $h_{ij} = \omega_{ij}/(1-\alpha_{ij}-\beta_{ij})$. The mean estimates of the unconditional variances $h_i$ and covariances $h_{ij}$ are 0.475 and 0.065, respectively. From Panel C, we also notice that the correlation is highly negative between $\gamma_{ij}$ and $\omega_{ij}$ (−0.743) but close to 0 between $\gamma_{ij}$ and $h_{ij}$ (−0.191). Therefore, the simulations are performed as follows: The unconditional variances $h_i$ are drawn from a symmetric Beta distribution with $p_{hi} = q_{hi} = 3$ in the range of the estimated variances $[h_i, \overline{h_i}]$. The unconditional covariances $h_{ij}$ are drawn from a Beta distribution with $p_{hc} = q_{hc} = 3$ in the range $[h_{ij}, \overline{h_{ij}}]$. We then define the constant terms as $\omega_i = (1-\gamma_i)h_i$ and $\omega_{ij} = (1-\gamma_{ij})h_{ij}$.

### 5.2 Simulation: Baseline Case

For each simulation, samples of variance parameters $(\alpha_i, \gamma_i, h_i)$ and covariance parameters $(\alpha_{ij}, \gamma_{ij}, h_{ij})$ for $i, j = 1, \ldots, N$, are drawn from their respective distribution, as described in the previous subsection. Then $N$ time-series of individual innovations $\{z_{i,t}\}_{t=1,\ldots,T}$ are drawn from a normal $N(0,1)$ distribution and the unexpected returns $\{\varepsilon_{i,t}\}_{t=1,\ldots,T}$ are constructed for $i = 1, \ldots, N$. The portfolio unexpected return $\{\varepsilon_{p,t}\}_{t=1,\ldots,T}$ is obtained by

\[ \varepsilon_{p,t} = \sum_{i=1}^{N} w_{i} \varepsilon_{i,t} \]

\[ \text{More precisely, if } \tilde{h}_i \text{ is drawn from a standard Beta}(p_h,q_h) \text{ distribution, the unconditional variance is defined as from } h_i = \overline{h} + (\overline{h} - \underline{h})\tilde{h}_i, \text{ where } \underline{h} \text{ and } \overline{h} \text{ denote the minimum and maximum estimates of the unconditional variances, respectively. The choice of } p_h = q_h = 3 \text{ ensures that the resulting } h_i \text{ are rather dispersed in the interval } [\underline{h}, \overline{h}]. \]
aggregation, with portfolio weights \( w = (1/N, \cdots, 1/N)' \). Finally, the parameters driving the aggregate squared return \( X_{p,t} = \varepsilon_{p,t}^2 \) are estimated from aggregate data only.

We consider two alternative estimators in order to evaluate the magnitude of the bias induced by imposing parameter homogeneity when deriving the aggregate squared return dynamic. The first one is based on the usual (implicit) assumption of parameter homogeneity, i.e., the QMLE of the strong GARCH(1,1) process:

\[
 h_{p,t} = \Omega_p + \Psi_1 X_{p,t-1} + \Phi_1 h_{p,t-1}. \tag{26}
\]

The second estimator, consistent with parameter heterogeneity, is the Least-Square ACE of the weak GARCH\((K_\Lambda,K_\Phi)\) process:

\[
 X_{p,t} = \Omega_p + \sum_{k=1}^{K_\Lambda} \Lambda_k X_{p,t-k} + v_{p,t} - \sum_{k=1}^{K_\Phi} \Phi_k v_{p,t-k}. \tag{27}
\]

In the baseline case, the number of observations per sample is \( T = 6,000 \) and the number of assets varies from \( N = 20 \) to 40. Each experiment is based on 1,000 replications. It should be noticed that these simulation experiments are not designed to exactly match all the features observed on U.S. equity returns, but rather to mimic some of their main properties.

Table 2 reports summary statistics of parameter estimates for the QMLE and ACE procedures. We begin with the case \( N = 20 \), which is a realistic number of asset classes in a strategic allocation approach. For the ACE, we report the estimates of \( \{\Omega_p, \Psi_1, \Phi, \Lambda_1\} \), for comparability with the QMLE, as well as the estimates of the Beta parameters \( p \) and \( q \). As expected, the QMLE provides biased estimates of the variance parameters. The most striking result is the severe downward bias in the \( \gamma \)-type parameter \((\Lambda_1 = \Psi_1 + \Phi_1)\). The median estimate is 0.835, while the expected value is 0.97. This bias is not due to the estimation of the \( \alpha \)-type parameter \((\Psi_1)\), as its median estimate is equal to 0.041,

---

13The Least-Square ACE estimator is based on \( K_\Lambda = 20 \) and 40 lags and \( K_\Phi = 5 \) lags, so that the first five terms \( \Phi_i, i = 1, \cdots, 5 \), are freely estimated. Results for the Minimum-Distance ACE are very similar and not reported to save space.

14For instance, actual data may be generated by asymmetric GARCH processes and/or fat-tailed innovations, which are features not introduced in this experiment.
which is rather close to the expected value 0.033 with a narrow confidence interval. On the opposite, the median estimate of the $\beta$-type parameter ($\Phi_1$) is far from the expected value (0.794 instead of 0.937) with a large uncertainty across simulations.

Increasing the number of assets does not help in estimating the persistence parameter, as the value of $\Lambda_1$ is still severely underestimated even with $N = 40$ (with a median estimate of 0.777). This result indicates that the QMLE is not able to generate the high persistence found in the simulated aggregate squared returns.

With regard to the ACE, the table reveals that for $N = 20$ the persistence parameter $\Lambda_1$ is correctly and precisely estimated to be 0.977 for both values of $K_\Lambda$ (20 and 40). This result suggests that a moderate number of additional lags is sufficient to correct for the aggregation bias. The parameter $\Psi_1$ is also very well estimated, with median estimates of 0.034 and 0.039, respectively.

Increasing the number of assets in the portfolio does not alter the parameter estimates significantly. This result is important, because it suggests that the ACE is able to reproduce rather closely the properties of the aggregate process, even for a relatively small number of assets.

5.3 Simulation: Robustness Check

To evaluate the robustness of the results presented above, we performed additional simulation experiments based on alternative assumptions regarding the range of the unconditional correlations, the distribution of the innovation process, and the choice of the portfolio weight vector. All simulation results, based on $T = 6,000$ and $N = 40$ assets, are reported in Table 3.\textsuperscript{15}

The first experiment relies on the effect of increasing the correlation between the assets. As outlined by Zaffaroni (2007), dynamic conditional heteroskedasticity of the aggregate process requires a sufficiently strong cross-correlation. While the baseline case was calibrated with a moderate positive correlation using the mean value found on U.S. stocks

\textsuperscript{15}Other experiments essentially left the patterns already described unaltered. In particular, there is no sizeable effect on the parameter estimates when the number of lags in the ACE ($K_\Phi$ and $K_\Lambda$) is increased or when the range of the unconditional variances is widened. The results, not reported in order to save space, are available upon request.
(0.167), this experiment considers the case of highly correlated assets \((\rho_{ij} \in [0.75; 0.9])\). As Panel A reveals, the median estimate of the persistence parameter \(\Lambda_1\) obtained from the QMLE is in the same range as in the baseline case (0.759), and therefore far below the expected value. The estimate of \(\Psi_1\) remains close to the expected value. ACEs turn out to be very robust to changes in the range of correlations across assets. The estimate of \(\Lambda_1\) only slightly decreases towards its expected value.

The second experiment considers a non-normal distribution of the innovation process. Although \(z_{i,t}\) has been assumed to be normally distributed in the simulations so far, it is well known that the empirical distribution of asset returns is often asymmetric and/or fat-tailed. The interaction between the variance dynamics and the distribution properties of returns has been highlighted by Engle (1982) and more recently by He and Teräsvirta (1999). To illustrate the consequences of innovations drawn from distributions with fat tails, Panel B of the table reports the results for a \(t\) distribution with 5 degrees of freedom. As expected, the magnitude of the bias in the QMLE is increased. The median estimates of the parameter \(\Lambda_1\) produced by the QMLE is decreased from 0.835 for the normal innovations to 0.808 for \(t(5)\) innovations. Introducing asymmetry into the innovation distribution through a skewed \(t\) distribution does not further affect these parameter estimates with any significance. Again, the properties of the ACEs are not altered by the change in the conditional distribution regardless of the number of lags \(K_{\Lambda}\). This result is consistent with the fact that the ACEs, which are based on Least-Square estimation, do not rely on any particular distributional assumption (provided the innovation’s fourth moment is finite).

The last experiment evaluates the effect of the portfolio weights on the performance of the estimators. While the previous simulations were based on equal weights, we consider now a portfolio with short sales: weights are randomly drawn between \(-0.2\) and \(0.2\), with the sum of the weights equal to 1. Again, the table reveals that the QMLE underestimates the persistence parameter, although to a lesser extent (Panel C). On the opposite, the ACEs produce parameter estimates that are very close to the expected values. These results suggest that no positivity restrictions on portfolio weights are in fact required to
obtain consistent estimators of the parameters driving aggregate squared returns. The ACEs easily accommodate portfolios with different weights or even with short sales.

5.4 Evidence from the U.S. Market Aggregate Dynamics

The simulation results reported above suggest that, under contemporaneous aggregation, the dynamic of the aggregate squared returns cannot be correctly estimated by the standard strong GARCH(1,1) model. Its main failure is that it produces severely biased estimates of the persistence parameter. It is likely to have dramatic implications on variance forecasting, an issue which we address in this section. We now consider the estimation and forecast of the aggregate squared returns dynamics, assuming that only aggregate data is available, and evaluate the effect of the aggregation bias. The aggregate return is here simply defined as the equally-weighted average of the individual returns on U.S. equities already discussed. Then, the dynamic of the aggregate squared return is modeled with the usual strong GARCH(1,1) process (equation (26)) or with the weak GARCH($K_A,K_\Phi$) process defined in Proposition 3 (equation (27)).

To avoid any over-fitting in the forecast exercise, the sample is divided into two sub-periods. The first 12 years (from 1988 to 1999) are used for the estimation of the model (3,131 observations), and the last eleven years (from January 2000 to December 2010) are used for the out-of-sample forecasts (2,869 observations).

5.4.1 Estimation of the Aggregate Dynamics

Table 4 reports the parameter estimates. For the QMLE, the usual GARCH(1,1) parameters are reported. For the ACEs, we report the parameters $p$ and $q$ of the Beta distribution. We also report the implied estimates of the standard deviation, skewness, and kurtosis of the cross-section of $\{\gamma_{ij}\}$ deduced from equations (11)–(14). This allows a comparison with the cross-section estimates reported in Table 1.

As the table reveals, the estimates of the persistence parameters display large differences between the QMLE and ACEs, in line with the simulation evidence reported in Section 5.2. The QML estimate of $\Lambda_1$ is low (0.907), reflecting a large downward bias. In contrast, both ACEs produce a high persistence of aggregate squared returns (0.98 and
These values can be compared with the actual average of the persistence parameters over the complete sample (0.977, as reported in Table 1). It should be mentioned that the parameter $\Psi_1$ is over-estimated by the QMLE (0.127), whereas the value found with the ACEs (0.05) is almost in line with the actual average (0.035).

Although both ACEs provide very similar parameter estimates, the table indicates that the ACE(40, 5) also more accurately reproduces the moments of the persistence parameter distribution. Indeed, the estimates of the cross-section variance, skewness, and kurtosis of $\{\gamma_{ij}\}$ implied by the ACE(40, 5) are closer to their empirical counterparts (reported in Table 1) than the estimates implied by the ACE(20, 5).

To gain further insight on the relative performance of these estimators in fitting aggregate squared returns, Figure 2 compares the empirical ACF of the squared returns $X_{p,t}$ with the ACF implied by the QMLE and the ACE(40, 5). As is clearly evident from the figure, the empirical ACF of the aggregate squared returns displays a slow decay. The first-order autocorrelation of $X_{p,t}$ is about 0.1. At order 50, it decreases to about 0.04, and at order 100 it is still about 0.02. The QMLE fails to produce a slow decay of the ACF, although it correctly estimates the first autocorrelations. The implied autocorrelation is indistinguishable from 0 at order 50. In contrast, this pattern is correctly reproduced by the ACE. The figure shows that the ACE also captures the slow decay observed beyond.

### 5.4.2 Comparison of the Variance Forecasts

As the individual components are in fact available in our dataset, it is possible to define a disaggregate estimator (henceforth labeled “DISE”), which uses all the information available on individual returns. This estimator consists in estimating the joint dynamics of all the individual stocks (using the approach of Ledoit, Santa-Clara, and Wolf, 2003), forecasting all individual variances and covariances, and then aggregating these forecasts to produce a forecast of the aggregate variance. Although this approach is rather time consuming, it is feasible in our context with a moderate number of assets. It is formally defined as:

$$\hat{h}_{DISE,t} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \hat{h}_{ij,t},$$
where \( \hat{h}_{ij,t} \) denotes the forecasts of the conditional (co)variance at date \( t \) deduced from the multivariate GARCH(1,1) model. The effect of the aggregation bias can be evaluated statistically by comparing the relative performances of the disaggregate and aggregate estimation techniques for forecasting the conditional variance of aggregate returns.

One difficulty is that the aggregate variance at date \( t \) is not observable, but measured with noise by squared unexpected returns \( \hat{X}_{p,t} \). As shown by Patton (2011), most standard loss functions used to rank competing variance forecasts are not robust to the use of noisy proxies and therefore result in incorrect ranking of the competing forecasts. The variance forecasts are thus compared using robust loss functions, which are selected among the family described by Patton (2011):

\[
L_{1,t}(\hat{X}_{p,t}, \hat{h}_t) = (\hat{X}_{p,t} - \hat{h}_t)^2,
\]
\[
L_{2,t}(\hat{X}_{p,t}, \hat{h}_t) = \hat{X}_{p,t}/\hat{h}_t - \log(\hat{X}_{p,t}/\hat{h}_t) - 1,
\]
\[
L_{3,t}(\hat{X}_{p,t}, \hat{h}_t) = (\hat{X}_{p,t}^3 - \hat{h}_t^3)/6 - \hat{h}_t^2(\hat{X}_{p,t} - \hat{h}_t)/2,
\]

where \( \hat{h}_t \) 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 two loss functions is based on the test developed by Diebold and Mariano (1995) and West (1996). For instance, the loss difference between the forecasts based on the DISE and the ACE is defined as:

\[
d_{k,t} = L_{k,t}(\hat{X}_{p,t}, \hat{h}_{\text{DISE},t}) - L_{k,t}(\hat{X}_{p,t}, \hat{h}_{\text{ACE},t}), \quad \text{for } k = 1, 2, 3,
\]

where \( \hat{h}_{\text{DISE},t} \) and \( \hat{h}_{\text{ACE},t} \) denote the variance forecast based on the DISE and the ACE, respectively. The Diebold-Mariano/West test is based on the t-stat associated with the loss difference \( DM_k = \bar{d}_k/\bar{\sigma}_k \), 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 \( N(0, 1) \).\(^{16}\)

\(^{16}\)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 \( \tau \) is the number of observations over the out-of-sample period.
Table 5 reports the Diebold-Mariano/West statistics $DM_k$ for the test that the estimation techniques have the same forecasting ability over the out-of-sample period, for various subperiods. Several results are of importance. First, most test statistics are positive, meaning that, in most cases, the aggregate techniques produce more accurate variance forecasts than the disaggregate estimator. Although individual squared returns provide useful information for forecasting the aggregate variance, the disaggregate estimator is plagued by the estimation error surrounding the individual parameter estimates. The table reveals that, at least for the present data, variance forecasts based on aggregate data are more accurate than those based on disaggregate data. The only exception occurs for the QMLE over the 2004-07 subsample, which corresponds to a period of very low volatility (except at the very end of the period). Under such circumstances, the disaggregate estimator outperforms the QMLE, although insignificantly.

Second, the ACE consistently dominates the disaggregate estimator for all loss functions. Most of the test statistics are positive and significant at the 1% significance level, except over the 2004-07 subsample. In particular, it performs very well for the loss function $L_{3,t}$ focusing on over-predictions. In contrast, the QMLE(1,1) fails in providing better forecasts than the disaggregate estimator under the loss function $L_{3,t}$.

To sum up, the ACE is able to significantly reduce under- as well as over-predictions relative to the disaggregate estimator, while the usual QMLE barely out-performs the disaggregate estimator. Presumably, this is due to the ability of the ACE to correctly reproduce the persistence of the actual squared returns.

6 Conclusion

This paper describes the dynamics of aggregate squared returns in the presence of a large number of heterogeneous assets, when the covariance matrix is driven by a strong multivariate GARCH(1,1) process. In the limit, the aggregate squared return is an ARMA process with an infinite number of lags. This result establishes a relationship between the parameters of the ARMA process and the moments of the cross-section distribution of the persistence parameter. The proposed estimation procedures explicitly acknowledge
this relationship to parameterize the sequence of parameters of the ARMA process, and therefore to reduce significantly the computational burden in the estimation of the aggregate process. In contrast to the usual QMLE, the estimation procedure provides unbiased estimates of the parameters driving aggregate squared returns and performs very well in finite samples.

The paper also evaluates the effect of the aggregation bias on the estimation of aggregate squared returns over a large sample of U.S. equities. Once the aggregation bias is adequately corrected, the aggregate squared return process is more persistent than suggested by the usual QMLE. In addition, the ACE produces more accurate variance forecasts than disaggregated or QML estimation techniques.

Some issues currently remain unresolved. 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 individual variance asymmetries are lost when only aggregate returns are considered. Another interesting extension is the case of time-varying weights. We have investigated the case of constant weights, which includes the equally-weighted portfolio. For sector indices or asset classes, however, the weights of the various components may vary over time, for instance in correlation with the relative market capitalization of the individual assets. In this case, the parameters of the aggregate squared return process are themselves time-varying as they reflect the cross-section average of the individual parameters based on time-dependent weights. In such instances, the composition effects must be explicitly taken into account. These extensions are left for further research.
7 Appendix: Proof of Proposition 1

It should be noticed that the matrix of covariance innovations can be rewritten as \( v_t = X_t - H_t = \varepsilon_t' - H_t = \left(H_t^{1/2}\right)(z_t' - \text{I}_N)(H_t^{1/2})' \). As the central term has a zero mean and is uncorrelated with the external terms, we obtain that the time-series expectation is \( E[v_t] = 0 \). \( v_t \) is also serially uncorrelated, but it is not an i.i.d. sequence, because its conditional covariance matrix varies over time. The properties of \( v_t \) have been investigated by Hafner (2003) in a similar context. The aggregate squared return innovation is \( v_{p,t} = E^{(w)}[v_t] = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j v_{ij,t} \). We can deduce from above that \( v_{p,t} \) is also serially uncorrelated, with time-series expectation \( E[v_{p,t}] = 0 \), with a time-varying conditional variance \( \sigma_{v,t}^2 = V[v_{p,t}|\mathcal{F}_t] \), where \( \mathcal{F}_t \) is the aggregate information set available at time \( t \). As the \( v_{ij,t} \) are in general cross-correlated, the aggregate variance \( \sigma_{v,t}^2 \) does not necessarily converge to 0 even as \( N \) increases.

The aggregate squared return process \( X_{p,t} \) is obtained as follows (we omit the exponent \((w)\) for ease of notation):

\[
X_{p,t} = \left( \sum_{i=1}^{N} w_i \varepsilon_{i,t} \right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j X_{ij,t} \\
= \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \left( \omega_{ij} + \gamma_{ij} X_{ij,t-1} + v_{ij,t} - \beta_{ij} v_{ij,t-1} \right) \\
= \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \omega_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \gamma_{ij} X_{ij,t-1} + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \left( v_{ij,t} - \beta_{ij} v_{ij,t-1} \right) \tag{28}
\]

The second term in equation (28) is simply given by:

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \gamma_{ij} X_{ij,t-1} = \Lambda_1 \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j X_{ij,t-1} + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j (\gamma_{ij} - \Lambda_1) X_{ij,t-1} = \Lambda_1 X_{p,t-1} + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j (\gamma_{ij} - \Lambda_1) X_{ij,t-1},
\]
where $\Lambda_1 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \gamma_{ij} = \widehat{E}^{(w)}[\gamma]$ denotes the cross-section mean of $\{\gamma_{ij}\}$. The third term in equation (28) is:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j (v_{ij,t} - \beta_{ij} v_{ij,t-1}) = v_{p,t} - \widehat{E}^{(w)}[\beta v_{t-1}].$$

The term $\widehat{E}^{(w)}[\beta v_{t-1}] = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \beta_{ij} v_{ij,t}$ can be written as

$$\widehat{E}^{(w)}[\beta v_{t-1}] = \widehat{E}^{(w)}[\beta] \widehat{E}^{(w)}[v_{t-1}] + E^{(w)}[(\beta - \widehat{E}^{(w)}[\beta])v_{t-1}] = \Phi_1 v_{p,t-1} + E^{(w)}[(\beta - \Phi_1) v_{t-1}],$$

with $\Phi_1 = \widehat{E}^{(w)}[\beta]$. Equation (28) therefore rewrites as:

$$X_{p,t} = \widehat{E}^{(w)}[\omega] + \Lambda_1 X_{p,t-1} + v_{p,t} - \Phi_1 v_{p,t-1} - E^{(w)}[(\beta - \Phi_1) v_{t-1}] + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j (\gamma_{ij} - \Lambda_1) X_{ij,t-1}.$$

Using again $X_{ij,t-1} = \omega_{ij} + \gamma_{ij} X_{ij,t-2} + v_{ij,t-1} - \beta_{ij} v_{ij,t-2}$ in this expression, we have:

$$X_{p,t} = \widehat{E}^{(w)}[\omega] + \widehat{E}^{(w)}[(\gamma - \Lambda_1) \omega] + \Lambda_1 X_{p,t-1} + \Lambda_2 X_{p,t-2} + v_{p,t} - \Phi_1 v_{p,t-1} - E^{(w)}[(\beta - \Phi_1) v_{t-1}] + \Phi_2 v_{p,t-2} - \widehat{E}^{(w)}[\phi_2 - \Phi_2] v_{t-2}$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \left\{ (\gamma_{ij} - \Lambda_1) \gamma_{ij} - \widehat{E}^{(w)}[(\gamma - \Lambda_1) \gamma] \right\} X_{ij,t-2},$$

where $\Lambda_2 = \widehat{E}^{(w)}[(\gamma - \Lambda_1) \gamma] = \widehat{V}^{(w)}[\gamma]$ and $\Phi_2 = \widehat{E}^{(w)}[\phi_2]$, with $\phi_2 = (\gamma - \Lambda_1) \beta$.

Proceeding iteratively, we obtain in the limit, as $T \rightarrow \infty$:

$$X_{p,t} = \Omega_p + \sum_{k=1}^{\infty} \Lambda_k X_{p,t-k} + v_{p,t} - \sum_{k=1}^{\infty} \Phi_k v_{p,t-k} + \sum_{k=1}^{\infty} \widehat{E}^{(w)}[(\psi_k - \Psi_k) v_{t-k}$$

$$+ \lim_{K \rightarrow \infty} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \left\{ (\lambda_{ij,K-1} - \Lambda_{K-1}) \gamma_{ij} - \widehat{E}^{(w)}[(\lambda_{K-1} - \Lambda_{K-1}) \gamma] \right\} X_{ij,t-K}, (29)$$

where $\Lambda_k = \widehat{E}^{(w)}[\lambda_k]$ with $\lambda_1 = \gamma$ and $\lambda_k = (\lambda_{k-1} - \Lambda_{k-1}) \gamma$, $\Phi_k = \widehat{E}^{(w)}[\phi_k]$ with $\phi_1 = \beta$ and $\psi_k = (\lambda_{k-1} - \Lambda_{k-1}) \beta$, and $\Psi_k = \widehat{E}^{(w)}[\psi_k]$ with $\psi_1 = \alpha$ and $\psi_k = (\lambda_{k-1} - \Lambda_{k-1}) \alpha$. The
constant term is defined as:

\[
\Omega_p = \tilde{E}^{(w)}[\omega] + \sum_{k=1}^{\infty} \tilde{E}^{(w)}[\lambda_k - \Lambda_k] \omega .
\]

In expression (29), the last term is negligible and the cross-section means \(\tilde{E}^{(w)}[(\psi_k - \Psi_k)\nu_{t-k}]\) are all equal to 0 as \(N\) is large enough, because the innovation is asymptotically orthogonal to the parameter set. We therefore obtain equation (7) in Proposition 1:

\[
X_{p,t} = \Omega_p + \sum_{k=1}^{\infty} \Lambda_k X_{p,t-k} + v_{p,t} - \sum_{k=1}^{\infty} \Phi_k v_{p,t-k}.
\] (30)

From equation (30), we deduce, by rearranging terms, equation (6) in Proposition 1:

\[
h_{p,t} = \Omega_p + \sum_{k=1}^{\infty} \Psi_k X_{p,t-k} + \sum_{k=1}^{\infty} \Phi_k h_{p,t-k},
\]

with \(\Psi_k = \Lambda_k - \Phi_k\).

Equation (30) also clearly simplifies when \(\gamma_{ij} = \bar{\gamma}, \forall i, j\). We obtain an ARMA(1,1) process, similar to the one proposed by Nijman and Sentana (1996):

\[
X_{p,t} = \Omega_p + \Lambda_1 X_{p,t-1} + v_{p,t} - \Phi_1 v_{p,t-1},
\] (31)

with \(\Omega_p = \tilde{E}^{(w)}[\omega], \Lambda_1 = \bar{\gamma}, \Phi_1 = \tilde{E}^{(w)}[\beta]\), and \(\Lambda_k = \Phi_k = 0, \forall k > 1\).
References


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

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_i$</th>
<th>$\beta_i$</th>
<th>$\gamma_i$</th>
<th>$\omega_i(\times100)$</th>
<th>$h_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Conditional variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.049</td>
<td>0.943</td>
<td>0.992</td>
<td>0.329</td>
<td>0.475</td>
</tr>
<tr>
<td>Std dev.</td>
<td>0.020</td>
<td>0.024</td>
<td>0.007</td>
<td>0.352</td>
<td>0.307</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.642</td>
<td>-1.647</td>
<td>-2.409</td>
<td>2.482</td>
<td>1.948</td>
</tr>
<tr>
<td>$p$</td>
<td>6.945</td>
<td>104.625</td>
<td>138.028</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>std err.</td>
<td>(0.173)</td>
<td>(2.721)</td>
<td>(3.925)</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{V}(w)^{1/2}$ ($\times100$)</td>
<td>1.817</td>
<td>2.197</td>
<td>0.753</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{S}(w)$</td>
<td>0.694</td>
<td>-0.713</td>
<td>-1.853</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{K}(w)$</td>
<td>3.676</td>
<td>3.704</td>
<td>8.072</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\alpha_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Conditional covariances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.035</td>
<td>0.942</td>
<td>0.977</td>
<td>0.138</td>
<td>0.065</td>
</tr>
<tr>
<td>Std dev.</td>
<td>0.009</td>
<td>0.017</td>
<td>0.014</td>
<td>0.089</td>
<td>0.033</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.774</td>
<td>-1.142</td>
<td>-1.561</td>
<td>1.492</td>
<td>2.524</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.703</td>
<td>4.625</td>
<td>5.754</td>
<td>5.846</td>
<td>18.500</td>
</tr>
<tr>
<td>$p$</td>
<td>16.715</td>
<td>190.958</td>
<td>145.961</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>std err.</td>
<td>(0.418)</td>
<td>(5.015)</td>
<td>(3.846)</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{V}(w)^{1/2}$ ($\times100$)</td>
<td>0.834</td>
<td>1.632</td>
<td>1.218</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{S}(w)$</td>
<td>0.462</td>
<td>-0.530</td>
<td>-1.035</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{K}(w)$</td>
<td>3.307</td>
<td>3.390</td>
<td>4.556</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>$\alpha_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Correlation matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The parameters of the conditional variance and covariance processes are estimated for U.S. equity returns over the 1988-2010 sample. Parameters are ($\alpha_i$, $\beta_i$, $\gamma_i$, $\omega_i$, $h_i$) for conditional variances (Panel A) and ($\alpha_{ij}$, $\beta_{ij}$, $\gamma_{ij}$, $\omega_{ij}$, $h_{ij}$) for conditional covariances (Panel B). Summary statistics are the mean, standard deviation, skewness, and kurtosis of the empirical distribution, the ML estimates of the parameters $p$ and $q$ of the corresponding Beta distribution (with the standard error in parentheses), and the standard deviation, skewness, and kurtosis implied by the estimated Beta distribution. Panel C provides the cross-correlations between the parameter estimates ($\alpha_{ij}$, $\beta_{ij}$, $\gamma_{ij}$, $\omega_{ij}$, $h_{ij}$) of the conditional covariance processes.
<table>
<thead>
<tr>
<th></th>
<th>Expected value</th>
<th>QMLE</th>
<th>ACE(20, 5)</th>
<th>ACE(40, 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: N = 20, T = 6,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Omega_p$</td>
<td>0.052</td>
<td>0.178</td>
<td>0.026</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.235)</td>
<td>(0.070)</td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>$\Psi_1$</td>
<td>0.033</td>
<td>0.041</td>
<td>0.034</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>$\Phi_1$</td>
<td>0.937</td>
<td>0.794</td>
<td>0.943</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>(0.204)</td>
<td>(0.069)</td>
<td>(0.060)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>$\Lambda_1$</td>
<td>0.970</td>
<td>0.835</td>
<td>0.977</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>(0.208)</td>
<td>(0.060)</td>
<td>(0.050)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>$p$</td>
<td>–</td>
<td>–</td>
<td>145.96</td>
<td>145.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.013)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>–</td>
<td>3.498</td>
<td>3.502</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.921)</td>
<td>(1.122)</td>
</tr>
<tr>
<td><strong>Panel B: N = 40, T = 6,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Omega_p$</td>
<td>0.052</td>
<td>0.183</td>
<td>0.020</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.220)</td>
<td>(0.074)</td>
<td>(0.068)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>$\Psi_1$</td>
<td>0.033</td>
<td>0.025</td>
<td>0.036</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>$\Phi_1$</td>
<td>0.937</td>
<td>0.752</td>
<td>0.941</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>(0.256)</td>
<td>(0.098)</td>
<td>(0.093)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>$\Lambda_1$</td>
<td>0.970</td>
<td>0.777</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>(0.260)</td>
<td>(0.089)</td>
<td>(0.081)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>$p$</td>
<td>–</td>
<td>–</td>
<td>145.96</td>
<td>145.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>–</td>
<td>3.593</td>
<td>3.599</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.158)</td>
<td>(0.164)</td>
</tr>
</tbody>
</table>

Note: The table provides the median of the GARCH parameter estimates for the QMLE and some ACEs. The ACEs are based on $K_\Lambda = 20$ and 40 lags and $K_\Phi = 5$ lags. The median of the absolute deviations from the median is reported in parentheses.
Table 3: Simulation experiments: Estimates of the aggregate GARCH parameters

<table>
<thead>
<tr>
<th></th>
<th>Expected value</th>
<th>QMLE</th>
<th>ACE(20,5)</th>
<th>ACE(40,5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: High correlations</strong> ($\rho_{ij} \in [0.75; 0.9]$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Omega_p$</td>
<td>0.065</td>
<td>0.207</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.219)</td>
<td>(0.100)</td>
<td>(0.099)</td>
<td></td>
</tr>
<tr>
<td>$\Psi_1$</td>
<td>0.033</td>
<td>0.030</td>
<td>0.035</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>$\Phi_1$</td>
<td>0.937</td>
<td>0.728</td>
<td>0.941</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>(0.236)</td>
<td>(0.115)</td>
<td>(0.113)</td>
<td></td>
</tr>
<tr>
<td>$\Lambda_1$</td>
<td>0.970</td>
<td>0.759</td>
<td>0.975</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>(0.241)</td>
<td>(0.114)</td>
<td>(0.110)</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>–</td>
<td>–</td>
<td>145.95</td>
<td>145.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.039)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>–</td>
<td>3.748</td>
<td>3.824</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.853)</td>
<td>(2.434)</td>
</tr>
<tr>
<td><strong>Panel B: $t$ distribution</strong> ($\nu = 5$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Omega_p$</td>
<td>0.052</td>
<td>0.154</td>
<td>0.019</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td>(0.063)</td>
<td>(0.050)</td>
<td></td>
</tr>
<tr>
<td>$\Psi_1$</td>
<td>0.033</td>
<td>0.032</td>
<td>0.036</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.032)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>$\Phi_1$</td>
<td>0.937</td>
<td>0.777</td>
<td>0.940</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>(0.234)</td>
<td>(0.089)</td>
<td>(0.077)</td>
<td></td>
</tr>
<tr>
<td>$\Lambda_1$</td>
<td>0.970</td>
<td>0.808</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>(0.237)</td>
<td>(0.076)</td>
<td>(0.061)</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>–</td>
<td>–</td>
<td>145.93</td>
<td>145.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>–</td>
<td>3.571</td>
<td>3.577</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.136)</td>
<td>(0.143)</td>
</tr>
<tr>
<td><strong>Panel C: Alternative weight vector with short sales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Omega_p$</td>
<td>0.052</td>
<td>0.228</td>
<td>0.039</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.276)</td>
<td>(0.097)</td>
<td>(0.096)</td>
<td></td>
</tr>
<tr>
<td>$\Psi_1$</td>
<td>0.033</td>
<td>0.037</td>
<td>0.038</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>$\Phi_1$</td>
<td>0.937</td>
<td>0.826</td>
<td>0.939</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>(0.169)</td>
<td>(0.060)</td>
<td>(0.057)</td>
<td></td>
</tr>
<tr>
<td>$\Lambda_1$</td>
<td>0.970</td>
<td>0.863</td>
<td>0.977</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>(0.172)</td>
<td>(0.059)</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>–</td>
<td>–</td>
<td>145.93</td>
<td>145.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.039)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>–</td>
<td>3.748</td>
<td>3.447</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.860)</td>
<td>(2.054)</td>
</tr>
</tbody>
</table>

Note: The table provides the median of the GARCH parameter estimates for the QMLE and some ACEs under various changes in the baseline experiment. The ACEs are based on $K_{\Lambda} = 20$ and 40 lags and $K_{\Phi} = 5$ lags. The median of the absolute deviations from the median is reported in parentheses.
Table 4: U.S. stocks: Estimates of the aggregate GARCH parameters

<table>
<thead>
<tr>
<th></th>
<th>QMLE</th>
<th>ACE(20,5)</th>
<th>ACE(40,5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_p$</td>
<td>0.008</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\Psi_1$</td>
<td>0.127</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.051)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>$\Phi_1$</td>
<td>0.780</td>
<td>0.930</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.051)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>$\Lambda_1$</td>
<td>0.907</td>
<td>0.980</td>
<td>0.979</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.049)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>$p$</td>
<td>–</td>
<td>431.24</td>
<td>433.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.732)</td>
<td>(18.594)</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>9.013</td>
<td>9.497</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.822)</td>
<td>(0.274)</td>
</tr>
<tr>
<td>$\tilde{V}^{(w)[\gamma]}[1/2} \times 100$</td>
<td>–</td>
<td>0.440</td>
<td>0.688</td>
</tr>
<tr>
<td>$\tilde{S}^{(w)[\gamma]}$</td>
<td>–</td>
<td>-0.430</td>
<td>-0.626</td>
</tr>
<tr>
<td>$\tilde{K}^{(w)[\gamma]}$</td>
<td>–</td>
<td>3.271</td>
<td>3.573</td>
</tr>
</tbody>
</table>

Note: The parameters of the aggregate GARCH model are estimated by QMLE and some ACEs for the U.S. equity portfolio over the 1988-1999 sample. The ACEs are based on $K_A = 20$ and 40 lags and $K_\Phi = 5$ lags. For the ACEs, the table also reports the estimates of the standard deviation, skewness, and kurtosis of the cross-section distributions of $\{\gamma_{ij}\}$ implied by the parameter estimates. The standard error of the parameter estimates is reported in parentheses.
Table 5: Test of forecasting ability of aggregate estimators against the disaggregate estimator

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: QMLE vs. DISE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_1$</td>
<td>2.401</td>
<td></td>
<td>-0.557</td>
<td>2.194</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_2$</td>
<td>3.737</td>
<td>2.562</td>
<td>-0.947</td>
<td>3.179</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_3$</td>
<td>1.768</td>
<td>1.733</td>
<td>-0.886</td>
<td>1.706</td>
</tr>
<tr>
<td><strong>Panel B: ACE(20,5) vs. DISE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_1$</td>
<td>2.903</td>
<td>2.653</td>
<td>2.221</td>
<td>2.732</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_2$</td>
<td>5.769</td>
<td>3.730</td>
<td>1.137</td>
<td>4.748</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_3$</td>
<td>2.391</td>
<td>2.266</td>
<td>1.851</td>
<td>2.361</td>
</tr>
<tr>
<td><strong>Panel C: ACE(40,5) vs. DISE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_1$</td>
<td>2.903</td>
<td>2.653</td>
<td>2.221</td>
<td>2.731</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_2$</td>
<td>5.765</td>
<td>3.728</td>
<td>1.135</td>
<td>4.746</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DM_3$</td>
<td>2.391</td>
<td>2.268</td>
<td>1.851</td>
<td>2.360</td>
</tr>
</tbody>
</table>

Note: The table provides statistics on the forecasting ability of the QMLE and some ACEs relative to the disaggregate estimator (DISE, defined in the text) for the U.S. equity returns over various subperiods. $DM_1$, $DM_2$, and $DM_3$ are the Diebold-Mariano/West test statistics associated with the loss functions $L_1$, $L_2$, and $L_3$ described in the text. Under the null hypothesis that the aggregate estimators have no superior forecasting ability relative to the disaggregate estimator, 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.
Figure 1: Empirical distributions and fitted Beta distributions of individual GARCH parameters

Note: This figure compares the empirical histograms of the individual parameters \( \{\alpha_{ij}\} \), \( \{\beta_{ij}\} \), and \( \{\gamma_{ij}\} \) to the estimated Beta distributions \( f_{\alpha}(\cdot) \), \( f_{\beta}(\cdot) \), and \( f_{\gamma}(\cdot) \), over U.S. equities for the 1988-2010 sample.
Figure 2: Empirical and estimated ACF of aggregate squared returns

Note: The empirical ACF of aggregate squared returns is compared to the estimated ACF obtained from the QMLE and ACE(40,5) for U.S. equity portfolio over the 1988-1999 sample.