Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model

Xiangdong LONG
Judge Business School, University of Cambridge, Cambridge CB2 1AG, U.K. (xl256@cam.ac.uk)

Liangjun SU
School of Economics, Singapore Management University, Singapore 178903 (ljsu@smu.edu.sg)

Aman ULLAH
Department of Economics, University of California, Riverside, CA 92521-0427 (aman.ullah@ucr.edu)

We propose a semiparametric conditional covariance (SCC) estimator that combines the first-stage parametric conditional covariance (PCC) estimator with the second-stage nonparametric correction estimator in a multiplicative way. We prove the asymptotic normality of our SCC estimator, propose a nonparametric test for the correct specification of PCC models, and study its asymptotic properties. We evaluate the finite sample performance of our test and SCC estimator and compare the latter with that of the PCC estimator, purely nonparametric estimator, and Hafner, Dijk, and Franses's (2006) estimator in terms of mean squared error and Value-at-Risk losses via simulations and real data analyses.

KEY WORDS: Conditional covariance matrix; Multivariate GARCH; Portfolio; Semiparametric estimator; Specification test.

1. INTRODUCTION

Since the seminal work of Engle (1982), there has developed a huge literature on modeling the time-varying volatility of economic data in univariate case. Nevertheless, for asset allocation, risk management, hedging, and asset pricing multivariate generalized autoregressive conditional heteroscedasticity (MGARCH) models are of more importance both theoretically and practically because they model the volatility and covolatility of multiple financial assets jointly. Many recent works were done in the area of MGARCH models, such as the VECH model of Bollerslev, Engle, and Wooldridge (1988), the BEKK model of Baba et al. (1991) and Engle and Kroner (1995), the dynamic conditional correlation (DCC) model of Engle (2002) and Engle and Sheppard (2001), the Factor GARCH model of Engle, Ng, and Rothschild (1990), to name just a few. However, all these existing MGARCH models share two common features: the normality assumption on the error's distribution and the linearity of dynamic conditional covariance matrix. The exceptions include the regime switching dynamic conditional correlation model of Pelletier (2006), the smooth transition conditional correlation (STCC) model by Silvennoinen and Teräsvirta (2005), and the asymmetric dynamic conditional correlation model by Cappiello, Engle, and Sheppard (2003), where parametric nonlinear conditional correlation models are used with Gaussian errors, and the copula-based MGARCH model by Lee and Long (2009), where copula was used to construct non-Gaussian errors. The normality assumption was rejected by Fama and French (1993), Richardson and Smith (1993), Longin and Solnik (2001), Ang and Chen (2002), and Mashal and Zeevi (2002), and so on. The linear dynamic assumption excludes possible nonlinearity. Once we diverge from linearity, there is too much freedom to specify nonlinearity.

In this article, we propose a semiparametric conditional covariance (SCC) model, which combines parametric and nonparametric estimators of conditional covariance matrix in a multiplicative way. We first model the conditional covariance matrix parametrically, just as we do for the conventional parametric MGARCH models. Then we model the conditional covariance of the standardized residuals nonparametrically. The estimate of the latter will serve as a nonparametric correction factor for the parametric conditional covariance (PCC) estimator. Such combined estimation was done by Olkin and Speigelman (1987) in density function, by Glad (1998) in conditional mean estimation, and by Mishra, Su, and Ullah (2010) in conditional variance estimation. Nevertheless, to our knowledge, there is no such combined estimator for conditional covariance matrix.

We provide asymptotic theory for our semiparametric estimator. It possesses several advantages over both pure parametric and nonparametric estimators. First, our SCC model avoids the common shortcomings of parametric MGARCH models on potential misspecifications of functional form and density function. It does not rely on either the distributional assumption on the error term or the parametric functional form of the conditional covariance matrix. Second, when the parametric model is misspecified, the parametric estimator of the conditional covariance is generally inconsistent despite the fact that the finite dimensional parameter in the parametric model may converge to some pseudo-true parameter (see White 1994). In contrast, our semiparametric estimator can still be consistent with the true

Journal of Business & Economic Statistics
January 2011, Vol. 29, No. 1
DOI: 10.1198/jbes.2009.07057

conditional covariance matrix under certain conditions. Third, when the parametric model is correctly specified, as expected, our semiparametric estimator is less efficient than the parametric estimator, but it can achieve the parametric convergence rate with a fixed bandwidth.

The original contribution to the literature lies in three aspects. First, we are among the first to consider the combined estimators of conditional covariance matrix. Our SCC estimators can be regarded as an extension of Mishra, Su, and Ullah (2010) from the conditional variance (one-dimension) case to the conditional covariance (multi-dimension) case. For notational simplicity, we focus on local constant (Nadaraya–Watson) estimation instead of local polynomial estimation. Our new findings suggest that the proposed SCC estimator has the same asymptotic variance as the one-step nonparametric conditional covariance (NCC) estimator, but different asymptotic biases. Second, based on the estimator of the nonparametric correction factor, we propose a formal test for the correct specification of PCC models, which was not addressed in earlier literature on combined estimation. Third, our theoretical results are validated via Monte Carlo simulations and real data analyses.

We report a small set of Monte Carlo simulation results to evaluate the finite sample performance of our nonparametric test and SCC estimator and compare the latter with that of the PCC estimator, the NCC estimator, and Hafner, Dijk, and Franses's (2006, HDF hereafter) semiparametric estimator. The data generating processes (DGP's) used in our simulations are motivated by the nonlinear and nonnormal stylized facts widely observed in financial data, for instance, conditional correlation tends to be high during the crisis period and low during the tranquil period. Simulations suggest that our nonparametric test for the correct specification of PCC models performs reasonably well in finite samples. For comparison across different estimators, we use both mean squared error (MSE) and 1% Value-at-Risk (VaR) losses. To evaluate a portfolio's VaR loss, we consider two portfolio-weighting mechanisms, namely equal weight (EW) and minimum variance weight (MVW). We find that our semiparametric estimators tend to outperform their parametric counterparts and the NCC and HDF's estimators.

In empirical analysis, we carry out in-sample (IS) estimation and out-of-sample (OoS) forecasting for the conditional covariance matrix of paired market indexes in three datasets. Our nonparametric tests reject all commonly used PCC models for all three datasets at the 1% significance level. This is in favor of the use of a semiparametric or nonparametric estimator for the conditional covariance. When we fit the datasets by our SCC model, the PCC model, the NCC model, and the HDF model, we find that our SCC model can always reduce the IS losses of the start-up PCC model regardless of portfolio weights, generally reduces the OoS losses over the PCC models, and tends to perform best across different models.

The rest of the article is organized as follows. We briefly review some PCC models in Section 2. In Section 3 we present our SCC model and estimator, propose a nonparametric test for the correct specification of PCC models, and study their asymptotic properties under the null hypothesis and a sequence of local alternatives. In Section 4 we provide a small set of Monte Carlo experiments to evaluate the finite sample performance of our SCC estimators and nonparametric test, and apply all conditional covariance models on three paired stock indices. All proofs are relegated to the Appendix.

To proceed, we define some notation that will be used throughout the article. Let $\mathbf{I}_k$ denote a $k \times k$ identity matrix. Let $\mathbf{z} = (z_1, \ldots, z_k)'$ be a $k \times 1$ vector and $\mathbf{Z}$ be a symmetric $k \times k$ matrix with (i, j)th element z_{ij}. The Euclidean norm of $\mathbf{z}$ or $\mathbf{Z}$ is denoted as $\|\mathbf{z}\|$ or $\|\mathbf{Z}\|$. We define the following operators: $\operatorname{diag}(\mathbf{Z})$ denotes the diagonal matrix with z_i in the (i, i)th place; $\mathbf{Z}^*$ denotes a diagonal matrix with the square roots of the diagonal elements of $\mathbf{Z}$ on its diagonal when $\mathbf{Z}$ is positive definite; $\operatorname{vec}(\mathbf{Z})$ stacks the columns of $\mathbf{Z}$ into a $k^2 \times 1$ vector; $\operatorname{vech}(\mathbf{Z})$ stacks the lower triangular part of $\mathbf{Z}$ (including the diagonal elements) into a $k(k+1)/2 \times 1$ vector. Further, we use D_k to denote the $k^2 \times (k(k+1)/2)$ unique duplication matrix and D_k^+ to denote its generalized inverse, which is of size $(k(k+1)/2) \times k^2$. That is, $\operatorname{vec}(\mathbf{Z}) = D_k \operatorname{vech}(\mathbf{Z})$, $\operatorname{vech}(\mathbf{Z}) = D_k^+ \operatorname{vec}(\mathbf{Z})$, $D_k^+ = (D_k' D_k)^{-1} D_k'$, and $D_k^+ D_k = \mathbf{I}_{k(k+1)/2}$. Here we use the fact that $D_k' D_k$ is nonsingular. Let $N_k \equiv D_k D_k^+$. We will use the following properties of N_k: N_k is symmetric, $N_k D_k = D_k$, $N_k D_k^{+\prime} = D_k^{+\prime}$, and $N_k(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A}) N_k$, where $\mathbf{A}$ is a $k \times k$ matrix. For more details, see Magnus and Neudecker (1999, pp. 48–50).

2. PARAMETRIC CONDITIONAL COVARIANCE MODELS

Suppose the return series $\{\mathbf{r}_t\}_{t=1}^T$ of the interested financial data follows the stochastic process:

$$\mathbf{r}_t | \mathcal{F}_{t-1} \sim \mathbf{P}(\boldsymbol{\mu}_t, \mathbf{H}_t; \theta), \qquad t = 1, \ldots, T, \tag{2.1}$$

where $\mathbf{r}_t \equiv (r_{1t}, \ldots, r_{kt})'$ is an $k \times 1$ vector, $\mathcal{F}_{t-1}$ is the information set (σ-field) at time $t-1$, $E(\mathbf{r}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t$, $E(\mathbf{r}_t \mathbf{r}_t' | \mathcal{F}_{t-1}) = \mathbf{H}_t$, $\mathbf{H}_t$ is the conditional covariance matrix, and $\mathbf{P}$ is the joint cdf of $\mathbf{r}_t$, and θ represents the parameters in the distribution. Like Engle (2002), for simplicity we assume the conditional mean $\boldsymbol{\mu}_t$ is zero. If not, necessary standardization should be applied on the data. Thus we can write the model for $\mathbf{r}_t$ as

$$\mathbf{r}_t = \mathbf{H}_t^{1/2} \mathbf{e}_t, \tag{2.2}$$

where $\mathbf{e}_t \equiv \mathbf{H}_t^{-1/2} \mathbf{r}_t$ is the standardized error with $E(\mathbf{e}_t | \mathcal{F}_{t-1}) = \mathbf{0}$ and $E(\mathbf{e}_t \mathbf{e}_t' | \mathcal{F}_{t-1}) = \mathbf{I}_k$. $\mathbf{e}_t$ is typically assumed to follow the standard normal distribution: $\mathbf{e}_t \sim$ iid $\mathrm{N}(0, \mathbf{I}_k)$. We are interested in estimating the conditional covariance matrix $\mathbf{H}_t$ of $\mathbf{r}_t$ without such a distributional assumption.

The conditional covariance matrix $\mathbf{H}_t$ can be decomposed as

$$\mathbf{H}_t = \mathbf{D}_t(\theta) \mathbf{R}_t(\theta) \mathbf{D}_t(\theta), \tag{2.3}$$

where $\mathbf{R}_t(\theta)$ is the conditional correlation matrix with the (i, j)th element denoted as $\rho_{ij,t}(\theta)$, which stands for the conditional correlation between r_{it} and r_{jt} and can be time-varying; $\mathbf{D}_t(\theta) = \operatorname{diag}(\sqrt{h_{1,t}}, \ldots, \sqrt{h_{k,t}})$ is a diagonal matrix with the square root of the conditional variances $h_{i,t}$, parameterized by the vector θ, on the diagonal. It is well known (e.g., Engle 2002) that the conditional correlation matrix $\mathbf{R}_t(\boldsymbol{\theta})$ is also the conditional covariance matrix of the standardized returns $\boldsymbol{\varepsilon}_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{kt})' = \mathbf{D}_t^{-1}(\theta) \mathbf{r}_t$, i.e., $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}) = \mathbf{R}_t(\theta)$.

Now we review some popular parametric models for the conditional covariance matrix $\mathbf{H}_t$, which will be used in Section 4.

These models stem from two different modeling methodologies. First, the BEKK model specifies the elements of $\mathbf{H}_t$ directly:

$$\mathbf{H}_t = \delta\delta' + \sum_{i=1}^{p} \overline{\mathbf{A}}_i \mathbf{H}_{t-i} \overline{\mathbf{A}}_i' + \sum_{j=1}^{q} \overline{\mathbf{B}}_j (\mathbf{r}_{t-j}\mathbf{r}_{t-j}') \overline{\mathbf{B}}_j', \tag{2.4}$$

where $\boldsymbol{\delta}$ is a $k \times k$ low-triangle matrix, and different matrix properties of $\overline{\mathbf{A}}_i$ and $\overline{\mathbf{B}}_j$ lead to three types of BEKK models: the matrices $\overline{\mathbf{A}}_i$ and $\overline{\mathbf{B}}_j$ in the full, diagonal, and scalar BEKK models are full matrices, diagonal matrices, and scalars, respectively. Second, instead of modeling the conditional covariance matrix directly, observing $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$ in Equation (2.3), one can model $\mathbf{H}_t$ indirectly through modeling $\mathbf{D}_t$ and $\mathbf{R}_t$ separately. The resulting models include the CCC model by Bollerslev (1990), the VC model by Tse and Tsui (2002), the DCC model by Engle (2002), among others. The CCC model assumes that $\mathbf{R}_t = \mathbf{R}$, a constant matrix, and hence the time-varying feature of conditional covariance can only be attributed to the time-varying conditional variances. The VC model by Tse and Tsui (2002) specifies univariate GARCH(p, q) models for individual returns and GARCH-type dynamic evolutions for the conditional correlation process $\{\mathbf{R}_t\}$:

$$\mathbf{R}_t = \left(1 - \sum_{i=1}^{p} \gamma_i - \sum_{j=1}^{q} \beta_j\right)\bar{\mathbf{R}} + \sum_{i=1}^{p} \gamma_i \mathbf{R}_{t-i} + \sum_{j=1}^{q} \beta_j \hat{\mathbf{R}}_{t-j}, \tag{2.5}$$

where $\bar{\mathbf{R}}$, $\mathbf{R}_t$, and $\hat{\mathbf{R}}_t$ are the unconditional, conditional, and sample correlation matrices at time t with unit diagonal elements. Similar to the CCC and VC models, the DCC model also uses the two-stage modeling strategy. In the first stage, one models the conditional variance processes with the usual univariate GARCH models and then obtains the standardized residual $\hat{\boldsymbol{\varepsilon}}_t$. In the second stage, one models the conditional covariance $\mathbf{Q}_t$ of $\boldsymbol{\varepsilon}_t$ as

$$\mathbf{Q}_t = \left(1 - \sum_{i=1}^{p} \gamma_i - \sum_{j=1}^{q} \beta_j\right)\bar{\mathbf{Q}} + \sum_{i=1}^{p} \gamma_i \mathbf{Q}_{t-i} + \sum_{j=1}^{q} \beta_j (\hat{\boldsymbol{\varepsilon}}_{t-j}\hat{\boldsymbol{\varepsilon}}_{t-j}'), \tag{2.6}$$

where $\bar{\mathbf{Q}}$ is the sample covariance matrix for $\hat{\boldsymbol{\varepsilon}}_t$. The basic properties of the correlation matrix, such as positive-definiteness and unit diagonal element, are ensured by using the transformation

$$\mathbf{R}_t = \mathbf{Q}_t^{*-1}\mathbf{Q}_t\mathbf{Q}_t^{*-1}, \tag{2.7}$$

where $\mathbf{Q}_t^*$ is a diagonal matrix with the square roots of the diagonal elements of $\mathbf{Q}_t$ on its diagonal.

In all the above models, the functional form of conditional covariance matrix is assumed to be known and the maximum likelihood estimation is done under the assumption of normality. These assumptions will not be required for the semiparametric estimators introduced below.

3. AN ALTERNATIVE SEMIPARAMETRIC CONDITIONAL COVARIANCE ESTIMATOR

In this section we first review HDF's semiparametric estimator and propose an alternative semiparametric estimator for the conditional covariance matrix. We then study the asymptotic properties of our SCC estimator and propose a nonparametric test for the correct specification of PCC models.

3.1 HDF's Semiparametric Estimator

Motivated by the idea that the conditional correlations depend on exogenous factors such as the market return or volatility, HDF proposed the following semiparametric model for $\mathbf{r}_t$:

$$\begin{aligned} &\mathbf{r}_t = \mathbf{D}_t(\theta)\boldsymbol{\varepsilon}_t, \\ &E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = 0, \qquad E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}) = \mathbf{R}(\mathbf{x}_t), \end{aligned} \tag{3.1}$$

where $\mathbf{D}_t(\theta)$ is as defined before [after Equation (2.3)], and $\mathbf{x}_t$ is observable at time $t-1$ and $\mathbf{x}_t \in \mathcal{F}_{t-1}$. Assuming that θ can be estimated by $\hat{\theta}$ at the parametric $\sqrt{T}$-rate, they define standardized residuals by $\widetilde{\boldsymbol{\varepsilon}}_t \equiv \boldsymbol{\varepsilon}_t(\hat{\theta}) = \mathbf{D}_t(\hat{\theta})^{-1}\mathbf{r}_t$. Then they regress $\widetilde{\boldsymbol{\varepsilon}}_t\widetilde{\boldsymbol{\varepsilon}}_t'$ on $\mathbf{x}_t$ nonparametrically to obtain $\widetilde{\mathbf{Q}}(\mathbf{x})$, the Nadaraya–Watson kernel estimator of $E(\widetilde{\boldsymbol{\varepsilon}}_t\widetilde{\boldsymbol{\varepsilon}}_t'|\mathbf{x}_t = \mathbf{x})$. Their semiparametric conditional correlation matrix estimator is defined by

$$\widetilde{\mathbf{R}}(\mathbf{x}) = (\widetilde{\mathbf{Q}}^*(\mathbf{x}))^{-1}\widetilde{\mathbf{Q}}(\mathbf{x})(\widetilde{\mathbf{Q}}^*(\mathbf{x}))^{-1}, \tag{3.2}$$

where $\widetilde{\mathbf{Q}}^*(\mathbf{x})$ is a diagonal matrix with the square roots of the diagonal elements of $\widetilde{\mathbf{Q}}(\mathbf{x})$ on its diagonal. Their semiparametric estimator of $\mathbf{H}_t$ can be written as follows

$$\widetilde{\mathbf{H}}_t = \mathbf{D}_t(\hat{\theta})\widetilde{\mathbf{R}}(\mathbf{x}_t)\mathbf{D}_t(\hat{\theta}). \tag{3.3}$$

Clearly, the HDF's estimators require correct specification of the conditional variance process in order to obtain a final consistent conditional correlation or covariance estimator. This is unsatisfactory since it is extremely hard to know a priori the correct form of the conditional variance process. Below we propose an alternative SCC estimator that can be consistent even if the conditional variance process may be misspecified in the first stage and it requires similar assumption to that in Equation (3.1).

3.2 An Alternative Semiparametric Estimator

Motivated by Glad (1998) and Mishra, Su, and Ullah (2010), we propose an alternative SCC estimator, which combines in a multiplicative way the *parametric* conditional covariance estimator from the first stage with the *nonparametric* conditional covariance estimator from the second stage. Essentially, this estimator nonparametrically adjusts the initial PCC estimator.

Let $\mathbf{H}_t = E(\mathbf{r}_t\mathbf{r}_t'|\mathcal{F}_{t-1})$ be the true time-varying conditional covariance process:

$$\mathbf{r}_t = \mathbf{H}_t^{1/2}\mathbf{e}_t, \qquad E(\mathbf{e}_t|\mathcal{F}_{t-1}) = 0, \qquad E(\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}) = \mathbf{I}_k, \tag{3.4}$$

where $\mathbf{H}_t^{1/2}$ is the symmetric square root matrix of $\mathbf{H}_t$. Let $\{\mathbf{H}_{p,t}(\theta)\}$ be a parametrically-specified time-varying conditional covariance process for $\mathbf{r}_t$, where $\theta \in \Theta \subset \mathbb{R}^p$ and

$\mathbf{H}_{p,t}(\theta) \in \mathcal{F}_{t-1}$. Analogous to Mishra, Su, and Ullah (2010), our estimation strategy builds on the simple identity

$$\mathbf{H}_t = \mathbf{H}_{p,t}(\theta)^{1/2} E[\mathbf{e}_t(\theta)\mathbf{e}_t(\theta)'|\mathcal{F}_{t-1}]\mathbf{H}_{p,t}(\theta)^{1/2}, \tag{3.5}$$

where $\mathbf{H}_{p,t}(\theta)^{1/2}$ is the symmetric square root matrix of $\mathbf{H}_{p,t}(\theta)$, and $\mathbf{e}_t(\theta) = \mathbf{H}_{p,t}(\theta)^{-1/2}\mathbf{r}_t$ is the standardized error from the parametric model. When $\theta = \theta_*$, some pseudo-true parameter value, we write $\mathbf{H}_{p,t} = \mathbf{H}_{p,t}(\theta_*)$ and $\mathbf{e}_t = \mathbf{e}_t(\theta_*)$. It is clear that the parametric component $\mathbf{H}_{p,t}(\theta)$ in Equation (3.5) can be any PCC model reviewed in Section 2 and estimated by some standard parametric method. To propose a reasonable estimator for the nonparametric component $E[\mathbf{e}_t(\theta)\mathbf{e}_t(\theta)'|\mathcal{F}_{t-1}]$, we follow the HDF's idea and assume that the conditional expectation of $\mathbf{e}_t\mathbf{e}_t'$ depends on the current information set $\mathcal{F}_{t-1}$ only through a $q \times 1$ observable vector $\mathbf{x}_t = (x_{1t}, \ldots, x_{qt})'$. That is,

$$E[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}] = \mathbf{G}_{np}(\mathbf{x}_t), \tag{3.6}$$

where $\mathbf{x}_t \in \mathcal{F}_{t-1}$. There is a fundamental difference between Equation (3.6) and the last expression in Equation (3.1). In order for $\mathbf{R}(\mathbf{x}_t)$ in Equation (3.1) to be a conditional correlation matrix, the conditional variance matrix or equivalently $\{\mathbf{D}_t(\theta)\}$ has to be specified correctly. Fortunately, there is no such a requirement for our definition of $\mathbf{G}_{np}(\mathbf{x}_t)$.

Let $\mathbf{G}_{np,t} = \mathbf{G}_{np}(\mathbf{x}_t)$. Equation (3.5) then reduces to

$$\mathbf{H}_t = \mathbf{H}_{p,t}^{1/2}\mathbf{G}_{np,t}\mathbf{H}_{p,t}^{1/2}. \tag{3.7}$$

Based upon Equations (3.5)–(3.7), we can estimate $\mathbf{H}_t$ in two stages:

Stage 1: Estimate the parameter θ by $\hat{\theta}$ in the parametric specification $\{\mathbf{H}_{p,t}(\theta)\}$ for the conditional covariance process. Define the standardized residuals by $\hat{\mathbf{e}}_t = \hat{\mathbf{H}}_{p,t}^{-1/2}\mathbf{r}_t$, where $\hat{\mathbf{H}}_{p,t} = \mathbf{H}_{p,t}(\hat{\theta})$.

Stage 2: Estimate $E[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}, \mathbf{x}_t = \mathbf{x}]$ nonparametrically by

$$\hat{\mathbf{G}}_{np}(\mathbf{x}) = \frac{\sum_{s=1}^T \hat{\mathbf{e}}_s\hat{\mathbf{e}}_s' K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})}{\sum_{s=1}^T K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})}, \tag{3.8}$$

where $K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}) = \prod_{l=1}^q h_l^{-1} k((x_{ls} - x_l)/h_l)$, $\mathbf{h} = (h_1, \ldots, h_q)$, $h_l = h_l(T)$, $l = 1, \ldots, q$, are bandwidth parameters, and k is a kernel function. Let $\hat{\mathbf{G}}_{np,t} = \hat{\mathbf{G}}_{np}(\mathbf{x}_t)$. Then our SCC estimator of $\mathbf{H}_t$ is obtained as

$$\hat{\mathbf{H}}_{sp,t} = \hat{\mathbf{H}}_{p,t}^{1/2}\hat{\mathbf{G}}_{np,t}\hat{\mathbf{H}}_{p,t}^{1/2}. \tag{3.9}$$

Correspondingly, the estimator of conditional correlation matrix from our SCC model is

$$\hat{\mathbf{R}}_{sp,t} = (\hat{\mathbf{H}}_{sp,t}^*)^{-1}\hat{\mathbf{H}}_{sp,t}(\hat{\mathbf{H}}_{sp,t}^*)^{-1}, \tag{3.10}$$

where $\hat{\mathbf{H}}_{sp,t}^*$ is a diagonal matrix with the square roots of the diagonal elements of $\hat{\mathbf{H}}_{sp,t}$ on its diagonal.

To proceed, we make a few remarks.

Remark 1. When $k = 1$, $\hat{\mathbf{H}}_{sp,t}$ reduces to the semiparametric estimator of conditional variance in the spirit of Mishra, Su, and Ullah (2010) who used the local polynomial estimation technique instead. In the above analysis, we assume $\mathbf{x}_t$ is observable. It turns out this is not necessary. In fact, we can allow $\mathbf{x}_t$ to be estimated from the data at a certain rate.

Remark 2. When the parametric model $\mathbf{H}_{p,t}$ is correctly specified, i.e., $\mathbf{H}_{p,t}(\theta_0) = \mathbf{H}_t$ a.s. for some $\theta_0 \in \Theta$ and $\theta_0 = \theta_*$, we have:

$$\mathbf{G}_{np}(\mathbf{x}_t) = E[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}] = \mathbf{I}_k. \tag{3.11}$$

In this case, $\hat{\mathbf{G}}_{np,t}$ is estimating the $k \times k$ identity matrix. On the other hand, if the parametric model $\mathbf{H}_{p,t}$ is misspecified, $\mathbf{G}_{np}(\mathbf{x}_t)$ will not be an identity matrix, and $\hat{\mathbf{G}}_{np,t}$ will serve as a nonparametric correction factor, which nonparametrically adjusts the initial PCC estimator. In Section 3.4 we will propose a test for the correct specification of PCC models based on Equation (3.11).

Remark 3. Our SCC estimator is quite different from that of HDF. In the special case where $\hat{\mathbf{H}}_{p,t}^{1/2} = \mathbf{D}_t(\hat{\theta})$, then $\hat{\mathbf{G}}_{np,t}$ is the same as $\widetilde{\mathbf{Q}}(\mathbf{x}_t)$ obtained by HDF. So

$$\hat{\mathbf{H}}_{sp,t} = \mathbf{D}_t(\hat{\theta})\widetilde{\mathbf{Q}}(\mathbf{x}_t)\mathbf{D}_t(\hat{\theta}).$$

We can show that $\hat{\mathbf{H}}_{sp,t}$ is asymptotically equivalent to $\widetilde{\mathbf{H}}_t = \mathbf{D}_t(\hat{\theta})(\widetilde{\mathbf{Q}}^*(\mathbf{x}_t))^{-1}\widetilde{\mathbf{Q}}(\mathbf{x}_t)(\widetilde{\mathbf{Q}}^*(\mathbf{x}_t))^{-1}\mathbf{D}_t(\hat{\theta})$. In a general case where $\hat{\mathbf{H}}_{p,t}^{1/2} \neq \mathbf{D}_t(\hat{\theta})$, $\hat{\mathbf{G}}_{np,t}$ is not equal to $\widetilde{\mathbf{Q}}(\mathbf{x}_t)$ and $\hat{\mathbf{H}}_{sp,t}$ and $\widetilde{\mathbf{H}}_t$ may have quite different properties in both large and small samples. If the parametric models [$\mathbf{H}_{p,t}(\theta)$ in our case and $\mathbf{D}_t(\theta)$ in HDF's case] are misspecified, our estimator for the conditional covariance matrix is still consistent under weak conditions while that of HDF is generally inconsistent.

3.3 Asymptotic Property of Our SCC Estimator

To study the asymptotic property of our SCC estimator, we make the following set of assumptions.

Assumptions.

A1. The strictly stationary process $\{\mathbf{r}_t, \mathbf{x}_t\}$ is α-mixing with mixing coefficients $\alpha(j)$ satisfying $\sum_{j=1}^{\infty} j^a\alpha(j)^{\delta/(\delta+2)} < \infty$ for some $\delta > 0$ and $a > \delta/(\delta+2)$. Also, $E(\|\mathbf{r}_t\|^{2(2+\delta)}) < \infty$ and $E(\|\mathbf{x}_t\|^{2+\delta}) < \infty$.

A2. The pseudo-true parameter $\theta_* \in \Theta \subset \mathbb{R}^p$ governing the PCC process $\{\mathbf{H}_{p,t}(\theta)\}$ exists uniquely and lies in the interior of a compact set Θ.

A3. $\hat{\theta} - \theta_* = O_P(T^{-1/2})$.

A4. $\mathbf{H}_{p,t} \equiv \mathbf{H}_{p,t}(\theta_*)$ is symmetric, finite, and positive-definite for each t. The process $\{\mathbf{e}_t = \mathbf{H}_{p,t}^{-1/2}\mathbf{r}_t\}$ is strictly stationary and α-mixing with mixing coefficients $\alpha(j)$. $\mathbf{x}_t$ has a q-dimension continuous density $f(\mathbf{x})$ that is bounded away from zero at $\mathbf{x}$.

A5. Let $\mathbf{H}_{p,t}(\theta)$ has continuous derivatives in the neighborhood of θ_*. $\mathbf{G}_{np}(\mathbf{x})$ have second order continuous partial derivatives in the neighborhood of $\mathbf{x}$. For some $\epsilon > 0$, $\sup_{\{\theta:\|\theta-\theta_*\|\leq\epsilon\}} \|\boldsymbol{\xi}_t(\theta)\| \leq \overline{D}_t$, where $\boldsymbol{\xi}_t(\theta) = \partial\mathbf{e}_t(\theta)/\partial\theta'$ and $E(\overline{D}_t^2) < \infty$.

A6. Let $\mu_{ij} = \int u^i k(u)^j\, du$. The kernel $k(\cdot)$ is a symmetric bounded density function such that $\mu_{21} < \infty$ and $|uk(u)| \to 0$ as $|u| \to \infty$.

A7. As $T \to \infty$, $h_j \to 0$, $T\mathbf{h}! \to \infty$, and $T\|\mathbf{h}\|^4\mathbf{h}! \to c \in [0, \infty)$, where $\mathbf{h}! = \prod_{j=1}^q h_j$.

Assumption A1 is a high-level assumption. When the individual return series follows a GARCH(1, 1) process, HDF shows that the α-mixing of $\{\mathbf{r}_t\}$ can be satisfied under weak conditions. Assumptions A2–A3 do not require the correct specification for modeling the parametric component. For example, whether the parametric model is true or not, under some regularity conditions for the quasimaximum likelihood estimation QMLE, the pseudo true parameter θ_* exists uniquely (White 1994, chapter 2) and can be estimated consistently at the regular $\sqrt{T}$ rate (White 1994, chapter 6). Assumptions A4 and A5 impose some regularity conditions on the $\{\mathbf{H}_{p,t}(\theta)\}$ process. Assumptions A6 and A7 are standard in the nonparametric kernel estimation literature.

The following theorem establishes the asymptotic property of $\hat{\mathbf{G}}_{np}(\mathbf{x})$.

Theorem 3.1. Under Assumptions A1–A7,

$$\sqrt{T\mathbf{h}!}\left\{\mathrm{vech}(\hat{\mathbf{G}}_{np}(\mathbf{x})) - \mathrm{vech}(\mathbf{G}_{np}(\mathbf{x})) - \mathrm{vech}(\mathbf{B}(\mathbf{x}))\right\} \xrightarrow{d} \mathrm{N}(0, \mu_{02}^{q} f(\mathbf{x})^{-1} D_k^{+} \mathbf{\Omega}(\mathbf{x}) D_k^{+\prime}), \quad (3.12)$$

where $\mathbf{\Omega}(\mathbf{x}) = (\omega_{ij,lm}(\mathbf{x}))$ is a $k^2 \times k^2$ matrix with typical elements

$$\omega_{ij,lm}(\mathbf{x}) = \mathrm{Cov}(\varrho_{ij,t}, \varrho_{lm,t} | \mathbf{x}_t = \mathbf{x}) \quad \text{with } \varrho_{ij,t} = e_{it} e_{jt},$$

$\mathbf{B}(\mathbf{x}) = (\mathbf{B}_{ij}(\mathbf{x}))$ is a $k \times k$ matrix with typical elements

$$\mathbf{B}_{ij}(\mathbf{x}) = \frac{\mu_{21}}{2f(\mathbf{x})} \sum_{l=1}^{q} \left[2 \frac{\partial f(\mathbf{x})}{\partial x_l} \frac{\partial \mathbf{G}_{np,ij}(\mathbf{x})}{\partial x_l} + f(\mathbf{x}) \frac{\partial^2 \mathbf{G}_{np,ij}(\mathbf{x})}{\partial x_l \, \partial x_l} \right] h_l^2,$$

where e_{it} is the ith element of $\mathbf{e}_t$ and $\mathbf{G}_{np,ij}(\mathbf{x})$ is the (i, j)th element of $\mathbf{G}_{np}(\mathbf{x})$.

Remark 4. Theorem 3.1 implies that we can estimate $\mathbf{G}_{np}(\mathbf{x})$ consistently by $\hat{\mathbf{G}}_{np}(\mathbf{x})$, which has the usual asymptotic bias and variance structure as typical local constant estimators. Let $\boldsymbol{\eta}_t = \mathrm{vech}(\mathbf{e}_t \mathbf{e}_t')$. We can get an alternative expression for $D_k^{+} \mathbf{\Omega}(\mathbf{x}) D_k^{+\prime}$:

$$D_k^{+} \mathbf{\Omega}(\mathbf{x}) D_k^{+\prime} = \mathrm{Var}(\boldsymbol{\eta}_t | \mathbf{x}_t = \mathbf{x}).$$

When the start-up PCC model is correctly specified, i.e., $\mathbf{H}_t = \mathbf{H}_{p,t}(\theta_*)$, then $\mathbf{G}_{np}(\mathbf{x}) = \mathbf{I}_k$, and the asymptotic bias term in Equation (3.12) vanishes $[\mathbf{B}(\mathbf{x}) = 0]$.

The asymptotic property of our semiparametric estimator for the conditional covariance matrix $\mathbf{H}_t$ is stated in the following corollary.

Corollary 3.2. (i) For any $\mathbf{x}_t$ such that $f(\mathbf{x}_t)$ is bounded away from 0, $\hat{\mathbf{H}}_{sp,t}$ and $\hat{\mathbf{R}}_{sp,t}$ are consistent for $\mathbf{H}_t$ and $\mathbf{R}_t$, respectively. That is,

$$\hat{\mathbf{H}}_{sp,t} = \hat{\mathbf{H}}_{p,t}^{1/2} \hat{\mathbf{G}}_{np,t} \hat{\mathbf{H}}_{p,t}^{1/2} \xrightarrow{p} \mathbf{H}_t, \qquad \text{and}$$

$$\hat{\mathbf{R}}_{sp,t} = (\hat{\mathbf{H}}_{sp,t}^{*})^{-1} \hat{\mathbf{H}}_{sp,t} (\hat{\mathbf{H}}_{sp,t}^{*})^{-1} \xrightarrow{p} \mathbf{R}_t.$$

(ii) $\sqrt{T\mathbf{h}!}\{\mathrm{vech}(\hat{\mathbf{H}}_{sp,t}) - \mathrm{vech}(\mathbf{H}_t) - \overline{\mathbf{B}}_t(\mathbf{x}_t)\} \xrightarrow{d} \mathrm{MN}(0, \mu_{02}^{q} \times f(\mathbf{x}_t)^{-1} D_k^{+} \overline{\mathbf{\Omega}}_t(\mathbf{x}_t) D_k^{+\prime})$, where $\overline{\mathbf{B}}_t(\mathbf{x}) = \mathrm{vech}(\mathbf{H}_{p,t}^{1/2} \mathbf{B}(\mathbf{x}) \mathbf{H}_{p,t}^{1/2})$ and $\overline{\mathbf{\Omega}}_t(\mathbf{x}) = (\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}) \mathbf{\Omega}(\mathbf{x}) (\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2})$. That is, conditional on $\mathbf{H}_{p,t}$ and $\mathbf{x}_t$, $\sqrt{T\mathbf{h}!}\{\mathrm{vech}(\hat{\mathbf{H}}_{sp,t}) - \mathrm{vech}(\mathbf{H}_t) - \overline{\mathbf{B}}_t(\mathbf{x}_t)\}$ is asymptotically normal with mean zero and variance $\mu_{02}^{q} \times f(\mathbf{x}_t)^{-1} D_k^{+} \overline{\mathbf{\Omega}}_t(\mathbf{x}_t) D_k^{+\prime}$.

Remark 5. Corollary 3.2(i) says that we can obtain a consistent estimator for the conditional covariance and correlation matrix. Corollary 3.2(ii) essentially says that $\hat{\mathbf{H}}_{sp,t}$ is also asymptotically normally distributed conditional on $\mathbf{H}_{p,t}$ and $\mathbf{x}_t$, and it inherits the asymptotic bias and variance structure of $\hat{\mathbf{G}}_{np}(\mathbf{x}_t)$. By the delta method, one can also show that the semiparametric estimator for conditional correlation matrix is also asymptotically distributed with the nonparametric convergence rate $\sqrt{T\mathbf{h}!}$.

Remark 6. To compare our estimator with the parametric estimator of conditional covariance, first note that when the parametric component is correctly specified, as expected, our estimator is less efficient than the parametric one since our estimator has a slower convergence rate than the parametric estimator as $\|\mathbf{h}\| \to 0$. Nevertheless, when $\mathbf{h}$ is kept fixed, a careful examination of the proof of Theorem 3.1 and Corollary 3.2 indicates that our semiparametric estimator is consistent with the true conditional covariance with the regular parametric $\sqrt{T}$-rate of convergence. In this sense, we say that our estimator is almost as good as the parametric estimator in terms of convergence rate when $\mathbf{h}$ is kept fixed. Next, in the case of misspecification, the PCC estimator is usually inconsistent (even though $\hat{\theta}$ is consistent for some pseudo-true parameter θ_*), while our semiparametric conditional covariance estimator is still consistent. Similar remarks hold true for the estimators of conditional correlation matrix.

Remark 7. Like Mishra, Su, and Ullah (2010), we can also compare our semiparametric estimator of conditional covariance with the one-step nonparametric kernel estimator. For the ease of comparison, we consider the simplest case where both $\mathbf{H}_{p,t}$ and $\mathbf{H}_t$ depend on the information set $\mathcal{F}_{t-1}$ only through $\mathbf{x}_t$. In this case, we can write $\mathbf{H}_{p,t} = \mathbf{H}_p(\mathbf{x}_t)$ and $\mathbf{H}_t = \mathbf{H}(\mathbf{x}_t)$, and the one-step nonparametric kernel estimator of $\mathbf{H}_t = \mathbf{H}(\mathbf{x}_t)$ is given by

$$\hat{\mathbf{H}}_{np,t} = \frac{\sum_{s=1}^{T} \mathbf{r}_s \mathbf{r}_s' K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t)}{\sum_{s=1}^{T} K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t)}.$$

In the sequel, we refer to $\hat{\mathbf{H}}_{np,t}$ as the nonparametric conditional covariance (NCC) estimator. Standard nonparametric regression theory reveals that

$$\sqrt{T\mathbf{h}!}\left\{\mathrm{vech}(\hat{\mathbf{H}}_{np,t}) - \mathrm{vech}(\mathbf{H}_t) - \mathrm{vech}(\mathbf{B}_{np}(\mathbf{x}_t))\right\} \xrightarrow{d} \mathrm{MN}(0, \mu_{02}^{q} f(\mathbf{x}_t)^{-1} D_k^{+} \mathbf{\Omega}_{np}(\mathbf{x}_t) D_k^{+\prime}),$$

where $\mathbf{\Omega}_{np}(\mathbf{x}) = (\omega_{ij,lm}^{(np)}(\mathbf{x}))$ is a $k^2 \times k^2$ matrix with typical elements $\omega_{ij,lm}^{(np)}(\mathbf{x}) = \mathrm{Cov}(r_{it} r_{jt}, r_{lt} r_{mt} | \mathbf{x}_t = \mathbf{x})$, and $\mathbf{B}_{np}(\mathbf{x}) = (\mathbf{B}_{np,ij}(\mathbf{x}))$ is a $k \times k$ matrix with typical elements

$$\mathbf{B}_{np,ij}(\mathbf{x}) = \frac{\mu_{21}}{2f(\mathbf{x})} \sum_{l=1}^{q} \left[2 \frac{\partial f(\mathbf{x})}{\partial x_l} \frac{\partial \mathbf{H}_{ij}(\mathbf{x})}{\partial x_l} + f(\mathbf{x}) \frac{\partial^2 \mathbf{H}_{ij}(\mathbf{x})}{\partial x_l \, \partial x_l} \right] h_l^2, \quad (3.13)$$

where $\mathbf{H}_{ij}(\mathbf{x})$ denotes the (i, j)th element of $\mathbf{H}(\mathbf{x})$, and r_{it} is the ith element of $\mathbf{r}_t$.

On the other hand, when both $\mathbf{H}_{p,t}$ and $\mathbf{H}_t$ depend on the information set $\mathcal{F}_{t-1}$ only through $\mathbf{x}_t$, it is easy to verify that

$$\begin{aligned}\overline{\boldsymbol{\Omega}}_t(\mathbf{x}_t) &= \big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)\boldsymbol{\Omega}(\mathbf{x}_t)\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big) \\ &= \big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)E\big(\mathrm{vec}(\mathbf{e}_t\mathbf{e}_t')[\mathrm{vec}(\mathbf{e}_t\mathbf{e}_t')]'|\mathbf{x}_t\big) \\ &\quad \times \big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big) \\ &= E\big(\mathrm{vec}\big(\mathbf{H}_{p,t}^{1/2}\mathbf{e}_t\mathbf{e}_t'\mathbf{H}_{p,t}^{1/2}\big)\big[\mathrm{vec}\big(\mathbf{H}_{p,t}^{1/2}\mathbf{e}_t\mathbf{e}_t'\mathbf{H}_{p,t}^{1/2}\big)\big]'|\mathbf{x}_t\big) \\ &= E\big(\mathrm{vec}(\mathbf{r}_t\mathbf{r}_t')[\mathrm{vec}(\mathbf{r}_t\mathbf{r}_t')]'|\mathbf{x}_t\big) = \boldsymbol{\Omega}_{np}(\mathbf{x}_t),\end{aligned}$$

by the fact that $(\mathbf{A} \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{e}_t\mathbf{e}_t') = \mathrm{vec}(\mathbf{A}\mathbf{e}_t\mathbf{e}_t'\mathbf{A})$ for any $k \times k$ matrix $\mathbf{A}$. This implies that our SCC estimator shares the same asymptotic variance–covariance matrix as the NCC estimator. So we are left to compare the asymptotic bias of our SCC estimator with that of the NCC estimator, i.e., to compare $\overline{\mathbf{B}}_t(\mathbf{x}_t) = \mathrm{vech}(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x}_t)\mathbf{H}_{p,t}^{1/2})$ with $\mathrm{vech}(\mathbf{B}_{np}(\mathbf{x}_t))$.

A typical element of $\overline{\mathbf{B}}_t(\mathbf{x}_t)$ is given by

$$\begin{aligned}\overline{\mathbf{B}}_{t,ij}(\mathbf{x}_t) &= \frac{\mu_{21}}{2f(\mathbf{x})}\sum_{l=1}^{k}\sum_{m=1}^{k}\mathbf{H}_{p,il}^{1/2}(\mathbf{x}_t) \\ &\quad \times \sum_{s=1}^{q}\left[2\frac{\partial f(\mathbf{x}_t)}{\partial x_s}\frac{\partial \mathbf{G}_{np,lm}(\mathbf{x}_t)}{\partial x_s} + f(\mathbf{x})\frac{\partial^2 \mathbf{G}_{np,lm}(\mathbf{x}_t)}{\partial x_s\,\partial x_s}\right] \\ &\quad \times h_s^2\mathbf{H}_{p,mj}^{1/2}(\mathbf{x}_t), \qquad (3.14)\end{aligned}$$

where $\mathbf{H}_{p,il}^{1/2}(\mathbf{x})$ denotes the (i,l)th element of $\mathbf{H}_p^{1/2}(\mathbf{x})$ and $\mathbf{G}_{np,lm}(\mathbf{x})$ is similarly defined. Unfortunately, the above expression appears too complicated to compare with $\mathbf{B}_{np,ij}(\mathbf{x}_t)$ defined by Equation (3.13). Only in the special case where $k = 1$ and $q = 1$ and where the local constant method is replaced by the local linear method can we follow Mishra, Su, and Ullah (2010) and show that $\overline{\mathbf{B}}_{t,ij}(\mathbf{x}_t)$ is smaller than $\mathbf{B}_{np,ij}(\mathbf{x}_t)$ in absolute value under weak conditions.

3.4 Test for the Correct Specification of PCC Models

In this section we propose a test of correct specification of parametric conditional covariance models based on Equation (3.11). The null hypothesis is

$$H_0 : \mathbf{G}_{np}(\mathbf{x}_t) = \mathbf{I}_k \quad \text{a.s.}, \qquad (3.15)$$

and the alternative hypothesis is

$$H_1 : \Pr(\mathbf{G}_{np}(\mathbf{x}_t) = \mathbf{I}_k) < 1. \qquad (3.16)$$

Let $\sigma_{ij}(\mathbf{x})$ denote the (i,j) element of $\mathbf{G}_{np}(\mathbf{x}), i,j = 1,\dots,k$. That is, $\sigma_{ij}(\mathbf{x}_t) = E[e_{it}e_{jt}|\mathcal{F}_{t-1}]$, where, recall, e_{it} denotes the ith element of $\mathbf{e}_t$. We can rewrite the null hypothesis as

$$H_0 : P(\sigma_{ij}(\mathbf{x}_t) = \delta_{ij}) = 1 \quad \text{for all } i,j = 1,\dots,k, \qquad (3.17)$$

and the alternative hypothesis as

$$H_1 : \Pr(\sigma_{ij}(\mathbf{x}_t) = \delta_{ij}) < 1 \quad \text{for some } i,j = 1,\dots,k, \qquad (3.18)$$

where δ_{ij} is Kronecker's delta, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Recall that $f(\mathbf{x})$ denotes the density function of $\mathbf{x}_t$. When the null and alternative hypotheses are written in the form of Equations (3.17) and (3.18), we can construct consistent tests of H_0 versus H_1 using various distance measures. A convenient choice is to use the measure

$$\Gamma = \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int(\sigma_{ij}(\mathbf{x}) - \delta_{ij})^2 f^2(\mathbf{x})\,d\mathbf{x} \geq 0, \qquad (3.19)$$

and $\Gamma = 0$, if and only if H_0 given by Equation (3.17) holds. Note that the use of density weight in the definition of Γ will help us avoid the random denominator issue. We will propose a test statistic based upon a kernel estimator of Γ.

To construct the sample analog of Γ, we first obtain estimators of $\sigma_{ij}(\mathbf{x})$ and $f(\mathbf{x})$, which are given by

$$\hat{\sigma}_{ij}(\mathbf{x}) = \frac{T^{-1}\sum_{s=1}^{T}\hat{e}_{it}\hat{e}_{jt}K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})}{\hat{f}(\mathbf{x})} \quad \text{and} \quad \hat{f}(\mathbf{x}) = T^{-1}\sum_{s=1}^{T}K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}), \qquad (3.20)$$

where $\hat{e}_{it}$ is the ith element of $\hat{\mathbf{e}}_t$. Note that $\hat{\sigma}_{ij}(\mathbf{x})$ is the (i,j) element of $\hat{\mathbf{G}}_{np}(\mathbf{x}_t)$. We then estimate Γ by the following functional:

$$\begin{aligned}\hat{\Gamma}_1 &= \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int(\hat{\sigma}_{ij}(\mathbf{x}) - \delta_{ij})^2\hat{f}^2(\mathbf{x})\,d\mathbf{x} \\ &= \frac{1}{T^2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t=1}^{T}(\hat{e}_{is}\hat{e}_{js} - \delta_{ij}) \\ &\quad \times (\hat{e}_{it}\hat{e}_{jt} - \delta_{ij})\overline{K}_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t), \qquad (3.21)\end{aligned}$$

where $\overline{K}_{\mathbf{h}}(\mathbf{u}) = \prod_{l=1}^{q}h_l^{-1}\overline{k}(u_l/h_l)$, $\mathbf{u} = (u_1,\dots,u_q)$, and $\overline{k}(u) = \int k(v)k(u-v)\,dv$ is the convolution kernel derived from k. For example, if $k(u) = \exp(-u^2/2)/\sqrt{2\pi}$, then $\overline{k}(u) = \exp(-u^2/4)/\sqrt{4\pi}$, a normal density with zero mean and variance 2.

The above statistic is simple to compute and offers a natural way to test H_0 in Equation (3.17). Nevertheless, we propose a bias-adjusted test statistic, namely,

$$\hat{\Gamma} = \frac{1}{T^2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t \neq s}^{T}(\hat{e}_{is}\hat{e}_{js} - \delta_{ij})(\hat{e}_{it}\hat{e}_{jt} - \delta_{ij})\overline{K}_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t). \qquad (3.22)$$

In effect, $\hat{\Gamma}$ removes the "diagonal" $(s = t)$ terms from $\hat{\Gamma}_1$ in Equation (3.21), thus reducing the bias of the statistic. A similar idea was used in Lavergne and Vuong (2000), Su and White (2007), and Su and Ullah (2009). We will show that after being appropriately scaled, $\hat{\Gamma}$ is asymptotically normally distributed under suitable assumptions.

To derive the asymptotic properties of the test statistic $\hat{\Gamma}$, we add the following assumptions.

Assumptions.

A8. Let $\varepsilon_{ijt} \equiv e_{it}e_{jt} - \delta_{ij}$. For $i,j = 1,\dots,k$, $E(|\varepsilon_{ijt}|^{4(1+\delta)}) \leq C$ and $E|\varepsilon_{ijt_1}^{r_1}\ \varepsilon_{ijt_2}^{r_2}\ \cdots\ \varepsilon_{ijt_l}^{r_l}|^{1+\delta} \leq C$ for some $C < \infty$, where $2 \leq l \leq 4$, $0 \leq r_s \leq 4$, and $\sum_{s=1}^{l} r_s \leq 8$.

A9. (i) Let $\mu_{ij2}(\mathbf{x}) \equiv E(\varepsilon_{ijt}^2|\mathbf{x}_t = \mathbf{x})$ and $\mu_{ij4}(\mathbf{x}) = E(\varepsilon_{ijt}^4|\mathbf{x}_t = \mathbf{x})$. Both $\mu_{ij2}(\mathbf{x})$ and $\mu_{ij4}(\mathbf{x})$ satisfy the Lipschitz condition: for $i, j = 1, \ldots, k$ and $l = 2, 4$, $|\mu_{ijl}(\mathbf{x}+\mathbf{x}^*) - \mu_{ijl}(\mathbf{x})| \le d_{ijl}(\mathbf{x})\|\mathbf{x}^*\|$, where $\|\cdot\|$ denotes the Euclidean norm and $\int d_{ijl}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} < C < \infty$. (ii) The joint density $f_{t_1,\ldots,t_l}(\cdot)$ of $(\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_l})$ $(1 \le l \le 4)$ exists and satisfies the Lipschitz condition: $|f_{t_1,\ldots,t_l}(\mathbf{x}^{(1)} + \mathbf{v}^{(1)}, \ldots, \mathbf{x}^{(l)} + \mathbf{v}^{(l)}) - f_{t_1,\ldots,t_l}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(l)})| \le D_{t_1,\ldots,t_l}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(l)})\|\mathbf{v}\|$, where $\mathbf{v}' = (\mathbf{v}^{(1)'}, \ldots, \mathbf{v}^{(l)'})$, $\int D_{t_1,\ldots,t_l}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(l)})\|\mathbf{v}\|^{2(1+\delta)}\, d\mathbf{v} \le C$ and $\int D_{t_1,\ldots,t_l}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(l)}) f_{t_1,\ldots,t_l}(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(l)})\, d\mathbf{v} \le C$ for some $C < \infty$.

Assumptions A8 and A9 are common in nonparametric estimation with strong mixing data (see Gao and King 2003). They are mainly used in the proof of Theorem 3.3 below.

Define

$$\sigma_0^2 \equiv 2 \int \overline{K}^2(\mathbf{u})\, d\mathbf{u} \sum_{i_1=1}^{k-1} \sum_{j_1=i_1}^{k} \sum_{i_2=1}^{k-1} \sum_{j_2=i_2}^{k} E\big[b_{i_1 j_1 i_2 j_2}^2(\mathbf{x}_t) f(\mathbf{x}_t)\big],$$

where $b_{i_1 j_1 i_2 j_2}(\mathbf{x}) = E[(e_{i_1 t} e_{j_1 t} - \delta_{i_1 j_1})(e_{i_2 t} e_{j_2 t} - \delta_{i_2 j_2})|\mathbf{x}_t = \mathbf{x}]$, and $\overline{K}(\mathbf{u}) = \prod_{l=1}^{q} \bar{k}(u_l)$. The asymptotic null distribution of $\hat{\Gamma}$ is established in the next theorem.

Theorem 3.3. Under Assumptions A1–A9 and under H_0, $T(\mathbf{h}!)^{1/2}\hat{\Gamma} \xrightarrow{d} \mathrm{N}(0, \sigma_0^2)$.

The proof is tedious and is relegated to the Appendix. From the proof we know that $T(\mathbf{h}!)^{1/2}\hat{\Gamma} = T(\mathbf{h}!)^{1/2}\overline{\Gamma} + o_P(1)$, where

$$\overline{\Gamma} = \frac{1}{T^2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \ne s}^{T} (e_{is} e_{js} - \delta_{ij})(e_{it} e_{jt} - \delta_{ij}) \overline{K}_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t).$$

This means that the first stage parametric estimation of the conditional covariance matrix does not affect the first order asymptotic properties of the test. To implement the test, we require a consistent estimate of the variance σ_0^2. Define

$$\hat{\sigma}^2 \equiv 2T^{-2}\mathbf{h}! \sum_{s=1}^{T} \sum_{t \ne s}^{T} \left[\sum_{i=1}^{k-1} \sum_{j=i}^{k} (\hat{e}_{it}\hat{e}_{jt} - \delta_{ij})(\hat{e}_{is}\hat{e}_{js} - \delta_{ij}) \right]^2 \times \overline{K}_{\mathbf{h}}^2(\mathbf{x}_t - \mathbf{x}_s). \quad (3.23)$$

It is easy to show that $\hat{\sigma}^2$ is consistent for σ_0^2 under H_0. We then compare

$$\hat{T} \equiv T(\mathbf{h}!)^{1/2}\hat{\Gamma}/\sqrt{\hat{\sigma}^2}, \quad (3.24)$$

with the one-sided critical value z_α from the standard normal distribution, and reject the null when $\hat{T} > z_\alpha$.

To examine the asymptotic local power of our test, we consider the following local alternatives:

$$H_1(\gamma_T): \sigma_{ij}(\mathbf{x}) = \delta_{ij} + \gamma_T \Delta_{ij}(\mathbf{x}), \qquad i, j = 1, \ldots, k, \quad (3.25)$$

where $\Delta_{ij}(\mathbf{x})$ satisfies $E|\Delta_{ij}(\mathbf{x}_t)|^{2+\delta} < \infty$ and $\gamma_T \to 0$ as $T \to \infty$. Define

$$\Delta_0 \equiv \int \sum_{i=1}^{k-1} \sum_{j=i}^{k} \Delta_{ij}^2(\mathbf{x}) f^2(\mathbf{x})\, d\mathbf{x}. \quad (3.26)$$

The following theorem establishes the local power property of our test.

Theorem 3.4. Under Assumptions A1–A9, suppose that $\gamma_T = T^{-1/2}(\mathbf{h}!)^{-1/4}$ in $H_1(\gamma_T)$. Then, the power of the test satisfies $P(\hat{T} \ge z_\alpha | H_1(\gamma_T)) \to 1 - \Phi(z_\alpha - \Delta_0/\sigma_0)$, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal.

Theorem 3.4 implies that the test has nontrivial asymptotic power against alternatives for which $\Delta_0 > 0$. The power increases with the magnitude of Δ_0/σ_0. Furthermore, by taking a large bandwidth we can make the alternative magnitude against which the test has nontrivial power, i.e., γ_T, arbitrarily close to the parametric rate $T^{-1/2}$.

4. SIMULATIONS AND EMPIRICAL ANALYSES

4.1 Monte Carlo Simulations

In this section, we conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our test and to compare our SCC estimators with several existing estimators of conditional covariance in terms of MSE and VaR losses.

4.1.1 Data Generating Processes. We generate data according to six data generating processes (DGP's), among which DGP's 1 and 2 will be used for the level study of our test and DGP's 3–6 are for power study and for the study of finite sample performance of various estimators of conditional covariance.

DGP 1 adopts the BEKK specification. We generate $\mathbf{e}_t \sim$ iid $\mathrm{N}(\mathbf{0}, \mathbf{I}_2)$ and set $\mathbf{r}_t \equiv \mathbf{H}_t^{1/2}\mathbf{e}_t$, where $\mathbf{H}_t = \boldsymbol{\delta}\boldsymbol{\delta}' + 0.05\mathbf{r}_{t-1} \times \mathbf{r}_{t-1}' + 0.9\mathbf{H}_{t-1}$, and

$$\boldsymbol{\delta} = \begin{pmatrix} 0.3509 & 0 \\ -0.0682 & 0.5726 \end{pmatrix}.$$

DGP 2 adopts the CCC specification. At time t, we first generate the correlation matrix $\mathbf{R}_t$ with the constant off-diagonal element 0.4, and the diagonal matrix $\mathbf{D} = \mathrm{diag}(\sqrt{h_{1,t}}, \sqrt{h_{2,t}})$, where

$$h_{1,t} = 0.5 + 0.05r_{1,t-1}^2 + 0.9h_{1,t-1}, \qquad \text{and}$$

$$h_{2,t} = 0.5 + 0.05r_{2,t-1}^2 + 0.7h_{2,t-1}.$$

Then we generate $\mathbf{e}_t \sim$ iid $\mathrm{N}(\mathbf{0}, \mathbf{I}_2)$ and set $\mathbf{r}_t = \mathbf{H}_t^{1/2}\mathbf{e}_t$, where $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$.

For the next two DGP's, we consider *nonlinear* specification for the time-varying conditional variance and correlation functions. DGP 3 specifies a bivariate GARCH-X process:

$$r_{i,t} = \sqrt{h_{i,t}}\varepsilon_{it}, \qquad i = 1, 2,$$

$$h_{1,t} = 0.5 + 0.05r_{1,t-1}^2 + 0.9h_{1,t-1} + 0.6x_{1t}^2,$$

$$h_{2,t} = 0.5 + 0.1r_{2,t-1}^2 + 0.6h_{2,t-1} + 0.9x_{2t}^2,$$

$$\boldsymbol{\varepsilon}_t \equiv (\varepsilon_{1t}, \varepsilon_{2t}) \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_t),$$

where $x_{it}, i = 1, 2$, are each iid $U(0, 1)$ and mutually independent, and

$$\mathbf{R}_t = \sigma^2(\mathbf{x}_t) \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix},$$

with $\mathbf{x}_t = (x_{1t}, x_{2t})'$, $\sigma^2(\mathbf{x}_t) = 0.25 + x_{1t}^2 + x_{2t}^2$, and $\rho_t = 0.5 + 0.4\cos(\pi t/10)$. The nonlinear characteristics of $h_{i,t}$ and ρ_t can be traced back to the simulation designs in Su and Ullah (2009) and Engle (2002), respectively. DGP 4 distinguishes itself from

DGP 3 by its specification on ρ_t in $\mathbf{R}_t$. In DGP 4, we set $\rho_t = 0.99 - 1.98/\{1+\exp(0.5\max(x_{1t}^2, x_{2t}^2))\}$, which is motivated by the stylized fact in financial markets that conditional correlation in crisis periods is higher than that in tranquil periods.

The last two DGP's, namely DGP's 5 and 6, consider *non-Gaussian* errors. They are identical to DGP 4 except the generation of $\mathbf{e}_t$. In DGP 5, e_{it}, $i = 1, 2$, are iid, uniformly distributed (U) on $[-\sqrt{3}, \sqrt{3}]$, and mutually independent; and in DGP 6, $e_{1t} \sim$ iid $U(-\sqrt{3}, \sqrt{3})$, $e_{2t} \sim$ iid N(0, 1), and they are mutually independent.

4.1.2 Test Results. For the level study, the correct parametric MGARCH model, namely the BEKK model for DGP 1 and the CCC model for DGP 2, is applied to fit the simulated data from DGP's 1 and 2. For the power study, we fit the data generated from DGP's 3–6 with CCC model to obtain the PCC estimator where GARCH(1, 1) model is considered for conditional variance. After fitting the parametric estimator $\hat{\mathbf{H}}_{pt}$, we obtain the standardized residuals $\hat{\mathbf{e}}_t = \hat{\mathbf{H}}_t^{-1/2}\mathbf{r}_t$ and then conduct our nonparametric test based on the residuals and $\mathbf{x}_t$. We choose $\mathbf{x}_t = \mathbf{r}_{t-1}$ in DGP's 1 and 2, and set $\mathbf{x}_t$ as given in the definition of DGP's 3–6.

To implement our test, we need to choose the kernel and bandwidth. We choose the Gaussian kernel: $k(u) = \exp(-u^2/2)/\sqrt{2\pi}$ and select the bandwidth following the lead of Horowitz and Spokoiny (2001) and Su and Ullah (2009). Specifically, we use a geometric grid consisting of N points $\mathbf{h}^{(s)}$, where $\mathbf{h}^{(s)} = (h_1^{(s)}, h_2^{(s)})$, $h_i^{(s)} = \omega^s s_i h_{\min}$, $i = 1, 2$, $s = 0, 1, \ldots, N-1$, s_i is the sample standard deviation of $\{x_{it}\}_{t=1}^T$, $N = [\log T] + 1$, $[\cdot]$ is the integer part of $\cdot$, $h_{\min} = T^{-4/(3q)}$, $h_{\max} = 0.5T^{-1/1000}$, and $\omega = (h_{\max}/h_{\min})^{1/(N-1)}$. For each $\mathbf{h}^{(s)}$, we calculate the test statistic in Equation (3.24) and denote it as $\hat{T}(\mathbf{h}^{(s)})$. Define

$$\operatorname{Sup} T \equiv \max_{0 \le s \le \mathcal{N}-1} \hat{T}\big(\mathbf{h}^{(s)}\big). \tag{4.1}$$

Even though $\hat{T}(\mathbf{h}^{(s)})$ is asymptotically distributed as N(0, 1) under the null for each s, the distribution of $\operatorname{Sup} T$ is unknown. Fortunately, we can use the wild bootstrap approximation to obtain the critical values.

We obtain the bootstrap residuals by $e_{it}^* = \hat{e}_{it} v_{it}$, $i = 1, 2$, $t = 1, \ldots, T$, where $\hat{e}_{it}$ are the standardized residuals from the first stage parametric estimation; $\{v_{it}\}$ are mutually independent iid sequences with mean 0, variance 1, and a finite fourth moment; and they are independent of the process $\{\mathbf{r}_t\}$. In our simulation, we draw v_{it} independently from a distribution with probability masses $p = (1+\sqrt{5})/(2\sqrt{5})$ and $1-p$ at the points $(1-\sqrt{5})/2$ and $(1+\sqrt{5})/2$, respectively. Based upon the bootstrap resampling data $\{e_{it}^*, i = 1, 2\}_{t=1}^T$ and $\{\mathbf{x}_t\}_{t=1}^T$, we construct the bootstrap version $\operatorname{Sup} T_n^*$ of the test statistic $\operatorname{Sup} T_n$. We repeat this procedure B times and obtain the sequence $\{\operatorname{Sup} T_{n,b}^*\}_{b=1}^B$. We reject the null when $p^* = B^{-1}\sum_{b=1}^B \mathbf{1}(\operatorname{Sup} T_n \le \operatorname{Sup} T_{n,b}^*)$ is smaller than the given level of significance, where $\mathbf{1}(\cdot)$ is the usual indicator function.

Table 1 reports the simulation results for DGP's 1–6. The number of replications M is 1000 and 500 for DGP's 1 and 2, and DGP's 3–6, respectively. In each case, we use $B = 200$ bootstrap resamples in each replication to obtain the p-value for our test. From the table, we see that our test is undersized for small to moderate sample sizes like $T = 250$ or 500. Despite this, the test exhibits reasonably good power behavior. In particular, as the sample size doubles, the power increases quickly. In addition, as expected, the rejection frequencies for DGP's 5 and 6, which exhibit both nonlinearity and non-Gaussianity, are much higher than those for DGP 4 with the presence of only nonlinearity.

Table 1. Finite sample rejection frequency for DGP's 1–6

	$T = 250$			$T = 500$		
DGP\Level	1%	5%	10%	1%	5%	10%
1	0.001	0.016	0.047	0.003	0.024	0.052
2	0.001	0.013	0.051	0.004	0.018	0.048
3	0.032	0.246	0.472	0.526	0.900	0.972
4	0.076	0.402	0.640	0.762	0.972	0.996
5	0.602	0.898	0.962	1.000	1.000	1.000
6	0.618	0.902	0.960	0.996	1.000	1.000

4.1.3 Evaluation of the SCC Estimates. To study the finite sample properties of our SCC estimates, we simulate data according to DGP's 3–6. For each DGP, we simulate 500 observations on $\mathbf{r}_t = (r_{1t}, r_{2t})'$, which represents roughly two-year daily data. The number of replications for each case is $M = 200$. We consider four parametric models for estimating the conditional correlation of $\mathbf{r}_t$, namely the CCC, VC, scalar BEKK, and DCC models reviewed in Section 2. In each case, we obtain our SCC estimators by choosing the conditioning variable as $\mathbf{x}_t$. To obtain our SCC estimators, we need to choose both the kernel and the bandwidth. It is well known that the choice of kernel function $k(\cdot)$ is not important in nonparametric or semiparametric estimation. We simply use the Gaussian kernel: $k(u) = \exp(-u^2/2)/\sqrt{2\pi}$. For the bandwidth, we follow the idea of grid-searching and set $h_i = c_j\hat{\sigma}_i n^{-1/6}$, $i = 1, 2$, where $\hat{\sigma}_i$ is the sample standard deviation of r_{it}, and the optimal c_j is chosen from $0.5, 0.6, \ldots, 5$ by minimizing the loss function of the corresponding semiparametric model.

We consider two loss functions for evaluation. The first is the MSE loss (cf. Engle 2002):

$$\operatorname{MSE}(\hat{\rho}_t) = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T}\big(\hat{\rho}_t^{(m)} - \rho_t^{(m)}\big)^2, \tag{4.2}$$

where $\rho_t^{(m)}$ and $\hat{\rho}_t^{(m)}$ are the true conditional correlation and its estimates at time t in the mth replication, respectively, and M is the number of replications. The second is based on the portfolios' VaR. The Basel Committee on Banking Supervision uses VaR to estimate the risk exposure of financial institutes for a 10-day holding period and 99% coverage ($\alpha = 1\%$). Denote the VaR of the weighted portfolio with tail probability α from model j within our framework as

$$\mathrm{VaR}_t^{\alpha,j} = \Phi_\alpha^j\sqrt{\boldsymbol{\omega}_t'\mathbf{H}_t^j\boldsymbol{\omega}_t}, \tag{4.3}$$

where Φ_α^j is the quantile of the cumulative distribution function of a weighted portfolio at tail probability $\alpha \in (0, 1)$ from model j. Apart from adopting the quantiles of standard normal distribution, Bauwens and Laurent (2005) used a Monte Carlo simulation and HDF (2006) employed the quantiles of the standardized portfolio returns. We adopt the method of HDF (2006)

Table 2. Mean square error (MSE) comparison for DGP's 3–6

Estimate\DGP	3	4	5	6
NCC	0.128	0.012	0.012	0.013
(%)	−0.157	22.981	24.528	24.699
CCC	0.128	0.016	0.016	0.017
CCC-NW	0.128	0.012	0.012	0.012
(%)	−0.392	22.981	25.786	25.904
VC	0.367	0.049	0.020	0.038
VC-NW	0.232	0.023	0.014	0.016
(%)	36.802	53.455	32.338	58.005
HDF	0.128	0.012	0.012	0.012
(%)	−0.392	22.981	25.786	25.904
BEKK	0.119	0.023	0.023	0.024
BEKK-NW	0.119	0.019	0.019	0.019
(%)	0.084	15.721	17.333	17.447
DCC	0.122	0.019	0.019	0.020
DCC-NW	0.122	0.015	0.015	0.015
(%)	0.082	19.474	22.460	21.538

to compute Φ_α^j. The VaR loss function for model j is

$$Q^{\alpha,j} = \frac{1}{T}\sum_{t=1}^{T}[\alpha - \mathbf{1}(y_t < \mathrm{VaR}_t^{\alpha,j})](y_t - \mathrm{VaR}_t^{\alpha,j}), \qquad (4.4)$$

where $\alpha = 1\%$. EW and MVW take $\boldsymbol{\omega}_t = k^{-1}\iota$ and $\boldsymbol{\omega}_t = \mathbf{H}_t^{-1}\boldsymbol{\iota}/(\boldsymbol{\iota}'\mathbf{H}_t^{-1}\boldsymbol{\iota})$, respectively, where $\boldsymbol{\iota}$ is a k-vector of ones.

Tables 2 and 3 compare the finite sample performance of various conditional covariance estimators. In addition to the absolute loss values for these estimators, the relative improvement ratios (%) are also reported. For each of our SCC estimates, the improvement ratio of the SCC estimates (CCC-NW, VC-NW, BEEK-NW, DCC-NW) over their PCC counterparts (CCC, VC, BEEK, DCC) is defined as

$$\mathrm{ratio} = 100\{\mathrm{Loss(PCC)} - \mathrm{Loss(SCC)}\}/\,\mathrm{Loss(PCC)}, \qquad (4.5)$$

where Loss(SCC) and Loss(PCC) are the mean square error (MSE) or VaR loss for the SCC estimate and the start-up PCC estimate, respectively. Since the NCC models have no start-up parametric model, we compare them with the parametric CCC estimate. For the HDF estimator, we take the parametric CCC model as the start-up model. A positive value of the improvement ratio means better performance of SCC estimators than their start-up PCC estimators or NCC/HDF estimators than the parametric CCC estimators. We summarize some interesting findings below. First, in terms of MSE, our SCC estimates usually beat the start-up PCC estimates except for the CCC estimate in DGP 3. Second, regardless of portfolio weighting methods, our SCC estimates always demonstrate better performance than the corresponding PCC estimates in terms of VaR loss. Third, we observe higher VaR loss improvement ratio of the SCC estimate over its start-up PCC estimate in the MVW portfolio than in the EW portfolio across nearly all DGP's. The only exception is the improvement ratio of the VC-NW estimate over the parametric VC estimate in DGP 3: the ratio for the VaR loss of EW portfolio is 76.03%, higher than that of MVW portfolio, 71.70%. Fourth, regarding MSE, the superiority ranking of semiparametric estimators is not always the same as that of parametric estimators. In DGP 5, for instance, the performance of the parametric estimators in the CCC, DCC, BEKK, and VC models deteriorates in order, while the deteriorating order of their semiparametric counterparts is the CCC-NW, VC-NW, DCC-NW, and BEKK-NW models.

4.2 Empirical Analysis

We examine three sets of financial daily time series data, the Dow Jones Industrial Average Index and Standard & Poor's 500 Index (DJIA&SPX) from January 2, 2003 to December 31, 2007 ($T = 1258$ observations); Cotation Assistée en Continu 40 and Financial Times Stock Exchange 100 Index (CAC&FTSE) from January 2, 2003 to December 31, 2007 ($T = 1281$ observations); and Hang Seng Index and Straits Times Index (HSI&STI) from January 2, 2003 to December 31, 2007 ($T = 1260$ observations). All datasets are downloaded from Yahoo

Table 3. Value-at-Risk (VaR) loss comparison for DGP's 3–6

	EW				MVW			
Estimate\DGP	3	4	5	6	3	4	5	6
NCC	0.082	0.075	0.069	0.061	0.033	0.034	0.033	0.035
(%)	2.038	0.399	−0.146	6.769	29.892	27.015	23.148	24.086
HDF	0.083	0.075	0.069	0.065	0.044	0.042	0.037	0.042
(%)	0.000	−0.133	0.000	0.615	5.806	9.368	15.509	9.462
CCC	0.083	0.075	0.069	0.065	0.047	0.046	0.043	0.047
CCC-NW	0.081	0.074	0.068	0.060	0.031	0.033	0.032	0.034
(%)	2.638	1.332	0.729	8.154	33.118	29.194	25.000	27.527
VC	0.328	0.085	0.069	0.067	0.297	0.060	0.043	0.050
VC-NW	0.079	0.074	0.068	0.060	0.084	0.036	0.032	0.034
(%)	76.031	13.130	0.729	10.912	71.698	40.168	25.463	33.135
BEKK	0.083	0.076	0.069	0.066	0.047	0.046	0.043	0.047
BEKK-NW	0.081	0.074	0.069	0.060	0.031	0.033	0.032	0.034
(%)	2.521	2.243	0.291	8.092	32.976	28.913	24.651	27.350
DCC	0.083	0.075	0.068	0.065	0.046	0.046	0.043	0.046
DCC-NW	0.081	0.074	0.068	0.060	0.031	0.032	0.032	0.034
(%)	2.292	1.198	0.731	8.308	33.045	29.258	24.942	27.586

Finance. For ease of interpretation, we compute the percentage returns ($\mathbf{r}_t$) as log returns multiplied by 100 and then demeaned. We split the whole samples at day R, the last day of 2006, use samples from 2003 to 2006 for in-sample (IS hereafter) estimation, and apply the "fixed scheme" to do one-day-ahead conditional covariance matrix forecast throughout end of year 2007. The IS standardized residuals are sorted to compute the p-value for VaR calculation later. "Fixed scheme" means in the whole forecasting period we keep using the same parameters, whose estimation is based on information set $\mathcal{F}_R$. For the out-of-sample (OoS hereafter) forecasting, the forecast length is 251, 255, and 250 days for the three datasets, respectively.

For each series, we assume the conditional mean is zero based on efficient market hypothesis. When implementing our nonparametric test for the correct specification of the PCC model based on the standardized residuals from the IS estimation, we choose the kernel and bandwidth as in Section 4.1.2. We choose the conditioning variable $\mathbf{x}_t$ as the one-day lagged percentage return, i.e., $\mathbf{x}_t = \mathbf{r}_{t-1}$. We conduct our nonparametric test for the three datasets and reject the null of correct specification of all the four PCC models under investigation at the 1% level. In view of this evidence we apply our SCC models to capture the remaining information in the standardized residuals of various PCC models.

When applying SCC models to these empirical datasets, we choose the kernel function and bandwidth as in Section 4.1.3. The conditioning variable is set as the one-day lagged percentage return, i.e., $\mathbf{x}_t = \mathbf{r}_{t-1}$. To judge the relative fitting and predictive ability of various conditional covariance models, we modify the two types of criterion functions used in Section 4.1.3. The MSE criterion in Equation (4.2) cannot be used here because the true conditional covariance matrix is not observable. Zangari (1997) addressed the advantage of focusing on the volatility h_t^y of the aggregate portfolio $y_t \equiv \omega' \mathbf{r}_t$ instead of the conditional covariance matrix $\mathbf{H}_t$, where $h_t^y = \boldsymbol{\omega}'\mathbf{H}_t\boldsymbol{\omega}$ and $\boldsymbol{\omega}$ is a weight vector. When comparing the predictability of univariate GARCH models, Awartani and Corradi (2005) substituted the unobservable volatility by the squared observed returns because of the rank-preserving property of this substitution under the MSE loss. They concluded that both squared returns and realized volatility are good proxies of the unobservable volatility for the purposes of model comparisons. Because intra-day returns are not available, Pelletier (2006) suggested using the cross-product of daily returns instead of the cumulative cross-product of intra-day returns over the forecast horizon. Following these authors, we compare various models by calculating the predictive measures, $\mathrm{MSE}_{\mathrm{OoS}}^j$ for model j, as

$$\mathrm{MSE}_{\mathrm{OoS}}^j = \frac{1}{(T-R)} \sum_{t=R}^{T-1} (\boldsymbol{\omega}_{t+1}' \hat{\mathbf{H}}_{t+1}^j \boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_{t+1}' \mathbf{r}_{t+1} \mathbf{r}_{t+1}' \boldsymbol{\omega}_{t+1})^2, \quad (4.6)$$

where $\hat{\mathbf{H}}_{t+1}^j$ is the one-step-ahead forecaster of $\mathbf{H}_{t+1}$ at time t from model j. The second loss is modified from VaR loss in Equation (4.3) in simulations:

$$Q_{\mathrm{OoS}}^{\alpha,j} = \frac{1}{(T-R)} \sum_{t=R}^{T-1} [\alpha - \mathbf{1}(y_{t+1} < \mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j})] \times (y_{t+1} - \mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j}), \quad (4.7)$$

where $\mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j} = \Phi_\alpha^j \sqrt{\omega_{t+1}' \hat{\mathbf{H}}_{t+1}^j \boldsymbol{\omega}_{t+1}}$, Φ_α^j is the quantiles of the standardized IS portfolio returns, and $\alpha = 1\%$. The in-sample (IS) losses are similarly defined.

The IS and OoS performance measures of different conditional covariance models over these empirical datasets are presented in Tables 4 and 5. For each pair of the parametric start-up PCC model and the corresponding SCC model, the improvement ratio is reported in percentage as before. For NCC and HDF models, we report the absolute loss values and the

Table 4. MSE loss for equal weight and minimum variance weight portfolios

	EW						MVW					
	DJIA&SPX		CAC&FTSE		HSI&STI		DJIA&SPX		CAC&FTSE		HSI&STI	
	IS	OoS	IS	OoS	IS	OoS	IS	OoS	IS	OoS	IS	OoS
NCC	1.44	3.05	1.65	3.77	1.48	14.60	0.66	2.45	0.86	5.09	1.14	13.78
%	15.30	−5.43	55.76	−7.51	14.71	−13.84	37.74	−7.30	73.19	−49.22	32.11	−26.25
HDF	1.31	2.87	1.85	3.41	1.41	12.60	0.63	2.27	1.12	3.36	1.07	10.31
%	23.16	0.73	50.62	2.92	18.49	1.80	39.87	0.55	65.09	1.55	35.97	5.62
CCC	1.70	2.89	3.74	3.50	1.74	12.83	1.05	2.28	3.21	3.41	1.68	10.93
CCC-NW	1.31	2.87	1.93	3.39	1.41	12.58	0.63	2.27	1.26	3.36	1.05	10.56
%	23.05	0.77	48.30	3.01	18.80	1.91	40.29	0.53	60.70	1.45	37.12	3.36
VC	1.70	2.89	3.75	3.50	1.74	12.74	1.05	2.30	3.18	3.42	1.68	10.74
VC-NW	1.31	2.87	1.85	3.39	1.41	12.52	0.62	2.28	1.13	3.36	1.05	10.18
%	23.02	0.74	50.69	2.90	18.65	1.74	41.18	0.87	64.38	1.88	37.41	5.20
BEKK	1.71	2.89	3.82	3.54	1.74	12.76	1.07	2.29	2.92	3.83	1.68	10.81
BEKK-NW	1.31	2.87	1.67	3.43	1.42	12.56	0.62	2.26	0.93	3.62	1.05	10.76
%	23.02	0.45	56.23	2.95	18.34	1.61	42.46	1.20	68.42	5.96	37.16	0.47
DCC	1.70	2.89	3.74	3.50	1.74	12.72	1.03	2.30	3.14	3.45	1.69	11.22
DCC-NW	1.31	2.86	1.67	3.40	1.42	12.46	0.60	2.28	0.93	3.35	1.06	10.02
%	23.07	0.73	55.52	2.89	18.46	2.03	41.79	0.97	70.79	2.74	37.29	10.72

Table 5. VaR loss for equal weight and minimum variance portfolios

	EW						MVW					
	DJIA&SPX		CAC&FTSE		HSI&STI		DJIA&SPX		CAC&FTSE		HSI&STI	
	IS	OoS	IS	OoS	IS	OoS	IS	OoS	IS	OoS	IS	OoS
NCC	0.030	0.071	0.034	0.056	0.031	0.117	0.021	0.032	0.025	0.046	0.025	0.112
%	−3.44	−25.53	−5.538	−12.85	4.012	−70.36	12.03	19.30	2.703	−4.56	20.38	−94.12
HDF	0.025	0.057	0.031	0.050	0.029	0.063	0.021	0.039	0.023	0.042	0.024	0.058
%	13.40	0.176	6.154	0.402	10.49	8.029	12.03	3.258	10.81	4.100	22.93	−1.038
CCC	0.029	0.057	0.033	0.050	0.032	0.069	0.024	0.040	0.026	0.044	0.031	0.058
CCC-NW	0.025	0.057	0.031	0.050	0.029	0.063	0.021	0.039	0.023	0.042	0.024	0.058
%	12.72	0.000	5.846	0.402	10.19	7.737	12.03	3.008	10.81	3.872	24.52	−0.519
VC	0.029	0.057	0.033	0.049	0.032	0.062	0.025	0.041	0.027	0.038	0.031	0.054
VC-NW	0.025	0.056	0.031	0.049	0.029	0.058	0.021	0.040	0.024	0.037	0.024	0.056
%	13.61	0.704	6.422	0.406	8.438	6.431	15.10	1.478	12.22	1.583	24.20	−4.664
BEKK	0.029	0.054	0.035	0.051	0.032	0.059	0.025	0.038	0.028	0.036	0.031	0.060
BEKK-NW	0.025	0.054	0.031	0.051	0.030	0.057	0.021	0.038	0.024	0.035	0.024	0.059
%	14.04	0.370	10.951	0.000	4.444	2.381	17.06	0.262	12.33	2.493	24.04	1.656
DCC	0.029	0.057	0.033	0.049	0.032	0.062	0.025	0.041	0.027	0.037	0.032	0.054
DCC-NW	0.025	0.056	0.030	0.049	0.029	0.059	0.021	0.041	0.024	0.036	0.024	0.054
%	13.36	0.885	7.034	0.000	9.091	4.693	14.23	0.735	11.28	3.784	25.32	0.924

improvement ratio relative to the CCC model. We summarize some interesting findings below. First, for both loss functions, our semiparametric model can always reduce the IS loss values of the start-up parametric model no matter which weight is used. Second, in terms of MSE, the improvement ratio of our SCC model against the start-up PCC model is always positive for both IS and OoS evaluations, both EW and MVW portfolios, and all datasets under examination. The same MSE superior pattern is observed in HDF, which produces a positive improvement ratio over CCC model across the datasets and sample period. But this supporting evidence is not found for the NCC model. But the relative out-of-sample gains of our and HDF's semiparametric estimators over the parametric estimators are generally much smaller than the relative in-sample gains. We conjecture that one of the reasons for this is the use of fixed-scheme (instead of rolling-window) forecast. Third, for DJIA&SPX and CAC&FTSE MVW portfolios, our SCC model can always reduce the VaR losses no matter which sample period (IS or OoS) or which start-up parametric model we choose. We do not observe the same phenomena for HSI&STI data, which might be explained by their emerging market properties. Fourth, there exists no semiparametric model that is universally the best across different datasets, weighting schemes, or loss functions. While the SBEKK-NW model has the smallest OoS VaR loss across the weighting methods for DJIA&SPX portfolio, its OoS MSE is bigger than that of the CCC-NW, VC-NW, and DCC-NW models for the equal weight DJIA&SPX portfolio. Last, for the same conditional covariance model, the MVW portfolio always outperforms the EW portfolio in terms of IS losses and generally outperforms the latter in terms of OoS losses.

APPENDIX: PROOF OF THE MAIN RESULTS

We use C to signify a generic constant whose exact value may vary from case to case, and a' to denote the transpose of a. Let $\hat{f}(\mathbf{x}) = T^{-1}\sum_{s=1}^{T} K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})$, and

$$\widetilde{\mathbf{G}}_{np}(\mathbf{x}) = T^{-1}\sum_{s=1}^{T} \mathbf{e}_s\mathbf{e}_s' K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})/\hat{f}(\mathbf{x}).$$

The following two lemmas are needed for the proof of Theorem 3.1.

Lemma A.1. Under Assumptions A1–A7,

$$\sqrt{T\mathbf{h}!}\left\{\mathrm{vech}(\widetilde{\mathbf{G}}_{np}(\mathbf{x})) - \mathrm{vech}(\mathbf{G}_{np}(\mathbf{x})) - \mathrm{vech}(\mathbf{B}(x))\right\} \xrightarrow{d} \mathrm{N}(0, \mu_{02}^{q} f(\mathbf{x})^{-1} D_k^{+}\boldsymbol{\Omega}(\mathbf{x}) D_k^{+\prime}),$$

where $\mathbf{h}! = h_1, \ldots, h_q$, and $\boldsymbol{\Omega}(\mathbf{x})$ and $\mathbf{B}(x)$ are defined in Theorem 3.1.

Proof. Let $W_{Tijs} = K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})e_{is}e_{js}$ and $W_{Tij} = T^{-1} \times \sum_{s=1}^{T} W_{Tijs}$, where e_{is} is the ith element of $\mathbf{e}_s$. Define two $k(k+1)/2$-vectors:

$$W_{Ts} = (W_{T11s}, W_{T21s}, \ldots, W_{Tk1s}, W_{T22s}, \ldots, W_{Tk2s}, \ldots, W_{Tkks})',$$

$$W_T = (W_{T11}, W_{T21}, \ldots, W_{Tk1}, W_{T22}, \ldots, W_{Tk2}, \ldots, W_{Tkk})'.$$

Clearly, $W_T = T^{-1}\sum_{s=1}^{T} W_{Ts}$. The statistic $W_{Tij}/\hat{f}(\mathbf{x})$ estimates the (i, j)th element of $\mathbf{G}_{np}(\mathbf{x})$ by using the pseudo-data $\{\mathbf{e}_t, \mathbf{x}_t\}$. Let $Z_{Ts} = (\mathbf{h}!/T)^{1/2}(W_{Ts} - E(W_{Ts}))$ and $Z_T = \sum_{s=1}^{T} Z_{Ts}$. Write

$$W_T = T^{-1}\sum_{s=1}^{T} E(W_{Ts}) + T^{-1}\sum_{s=1}^{T}(W_{Ts} - E(W_{Ts})) = T^{-1}\sum_{s=1}^{T} E(W_{Ts}) + (T\mathbf{h}!)^{-1/2}\sum_{s=1}^{T} Z_{Ts}.$$

The first term contributes to the bias of $\widetilde{\mathbf{G}}_{np}(\mathbf{x})$, whereas the second term contributes to the variance of $\widetilde{\mathbf{G}}_{np}(\mathbf{x})$. The proof will be completed by proving the following claims:

$$\hat{f}(\mathbf{x}) \xrightarrow{p} f(\mathbf{x}), \tag{A.1}$$

$$T^{-1}\sum_{s=1}^{T} E(W_{Ts}) = f(\mathbf{x})\,\mathrm{vech}(\mathbf{G}_{np}(\mathbf{x})) + f(\mathbf{x})\,\mathrm{vech}(\mathbf{B}(\mathbf{x})) + o_P(\|\mathbf{h}\|^2), \tag{A.2}$$

and

$$Z_T = \sum_{s=1}^{T} Z_{Ts} \xrightarrow{d} \mathrm{N}(0, \mu_{02}^q f(\mathbf{x}) D_k^+ \boldsymbol{\Omega}(\mathbf{x}) D_k^{+\prime}). \tag{A.3}$$

Assumption (A.1) follows from standard results in the kernel density estimation. Using standard arguments for analyzing the bias of the Nadaraya–Watson estimator, we have

$$\begin{aligned} E(W_{Tijs}) &= E[K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}) e_{it} e_{jt}] \\ &= f(\mathbf{x})[\mathbf{G}_{np,ij}(\mathbf{x}) + \mathbf{B}_{ij}(\mathbf{x})] + o_P(\|\mathbf{h}\|^2), \end{aligned}$$

where

$$\mathbf{B}_{ij}(\mathbf{x}) = \frac{\mu_{21}}{2f(\mathbf{x})} \sum_{l=1}^{q} \left[2\frac{\partial f(\mathbf{x})}{\partial x_l} \frac{\partial \mathbf{G}_{np,ij}(\mathbf{x})}{\partial x_l} + f(\mathbf{x}) \frac{\partial^2 \mathbf{G}_{np,ij}(\mathbf{x})}{\partial x_l \, \partial x_l} \right] h_l^2.$$

Thus Assumption (A.2) follows by the stationarity assumption. To show Assumption (A.3), let $\mathbf{c} = (c_{11}, c_{21}, \ldots, c_{k1}, c_{22}, \ldots, c_{k2}, \ldots, c_{kk})'$ denote a $k(k+1)/2$-vector of bounded constants such that $\|\mathbf{c}\| = 1$. By the Cramér–Wold device, it suffices to show

$$\mathbf{c}' Z_T = \sum_{s=1}^{T} \mathbf{c}' Z_{Ts} \xrightarrow{d} \mathrm{N}(0, \mu_{02}^q f(\mathbf{x}) \mathbf{c}' D_k^+ \boldsymbol{\Omega}(\mathbf{x}) D_k^{+\prime} \mathbf{c}). \tag{A.4}$$

By construction, $E(Z_T) = 0$, and

$$\begin{aligned} \mathrm{Var}(\mathbf{c}' Z_T) &= T^{-1}\mathbf{h}! \sum_{t=1}^{T} \mathrm{Var}(\mathbf{c}' W_{Tt}) \\ &\quad + 2T^{-1}\mathbf{h}! \sum\sum_{1\le s<t\le T} \mathrm{Cov}(\mathbf{c}' W_{Ts}, \mathbf{c}' W_{Tt}) \\ &\equiv A_1 + A_2. \end{aligned} \tag{A.5}$$

We calculate A_1 and A_2 in turn.

$$\begin{aligned} A_1 &= T^{-1}\mathbf{h}! \sum_{t=1}^{T} \mathrm{Var}(\mathbf{c}' W_{Tt}) \\ &= \sum\sum_{1\le j\le i\le k} \sum\sum_{1\le m\le l\le k} c_{ij} c_{lm} \\ &\quad \times \left[T^{-1}\mathbf{h}! \sum_{t=1}^{T} E[K_{\mathbf{h}}^2(\mathbf{x}_t - \mathbf{x}) \,\mathrm{Cov}(\varrho_{ij,t}, \varrho_{lm,t} | \mathbf{x}_t = \mathbf{x})] \right] \\ &= \mu_{02}^q f(\mathbf{x}) \sum\sum_{1\le j\le i\le k} \sum\sum_{1\le m\le l\le k} c_{ij} c_{lm} \omega_{ij,lm}(\mathbf{x}) + O(\|\mathbf{h}\|) \\ &= \mu_{02}^q f(\mathbf{x}) \mathbf{c}' D_k^+ \boldsymbol{\Omega}(\mathbf{x}) D_k^{+\prime} \mathbf{c} + O(\|\mathbf{h}\|), \end{aligned} \tag{A.6}$$

where $\varrho_{ij,t} = e_{it} e_{jt}$ and $\omega_{ij,lm}(\mathbf{x}) = \mathrm{Cov}(\varrho_{ij,t}, \varrho_{lm,t} | \mathbf{x}_t = \mathbf{x})$. To calculate A_2, write

$$\begin{aligned} A_2 &= 2T^{-1}\mathbf{h}! \sum\sum_{1\le s<t\le T} \sum\sum_{1\le j\le i\le k} \sum\sum_{1\le m\le l\le k} c_{ij} c_{lm} \,\mathrm{Cov}(W_{Tijs}, W_{Tlmt}) \\ &= 2\mathbf{h}! \sum_{t=2}^{T} \left(1 - \frac{j}{T}\right) \sum\sum_{1\le j\le i\le k} \sum\sum_{1\le m\le l\le k} c_{ij} c_{lm} \\ &\quad \times \mathrm{Cov}(W_{Tij1}, W_{Tlmt}). \end{aligned} \tag{A.7}$$

Noting that even though $\{\mathbf{v}_t\}$ is an mds, this does not ensure that $\mathrm{Cov}(W_{Tijs}, W_{Tlmt}) = 0$ for $s \neq t$. To bound the right-hand side of Assumption (A.7), we split it into two terms as follows

$$\begin{aligned} &\sum_{t=2}^{T} |\,\mathrm{Cov}(W_{Tij1}, W_{Tlmt})| \\ &\quad = \sum_{t=2}^{d_T} |\,\mathrm{Cov}(W_{Tij1}, W_{Tlmt})| + \sum_{t=d_T+1}^{T} |\,\mathrm{Cov}(W_{Tij1}, W_{Tlmt})| \\ &\quad \equiv J_1 + J_2, \end{aligned} \tag{A.8}$$

where d_T is a sequence of positive integers such that $d_T \mathbf{h}! \to 0$ as $T \to \infty$. Since for any $t > 1$, $|E(W_{Tij1} W_{Tlmt})| = O(1)$,

$$J_1 = O(d_n). \tag{A.9}$$

For J_2, by the Davydov's inequality (e.g., Bosq 1996, p. 19), we have

$$\begin{aligned} &|\,\mathrm{Cov}(W_{Tij1} W_{Tlmt})| \\ &\quad \le C[\alpha(t-1)]^{\delta/(2+\delta)} \sup_{i,j} (E|W_{Tij1}|^{2+\delta})^{2/(2+\delta)} \\ &\quad \le C(\mathbf{h}!)^{-(2+2\delta)/(2+\delta)} [\alpha(t-1)]^{\delta/(2+\delta)}. \end{aligned}$$

So by Assumption A1,

$$\begin{aligned} J_2 &\le C(\mathbf{h}!)^{-(2+2\delta)/(2+\delta)} \sum_{t=d_T+1}^{T} [\alpha(t-1)]^{\delta/(2+\delta)} \\ &\le C(\mathbf{h}!)^{-(2+2\delta)/(2+\delta)} d_T^{-a} \sum_{t=d_T}^{\infty} t^a [\alpha(t)]^{\delta/(2+\delta)} \\ &= o((\mathbf{h}!)^{-1}), \end{aligned} \tag{A.10}$$

by choosing d_T such that $d_T^a (\mathbf{h}!)^{\delta/(2+\delta)} \to \infty$. The last condition can be simultaneously met with $d_T \mathbf{h}! \to 0$ for a well-chosen sequence $\{d_T\}$ because $a > \delta/(2+\delta)$ by Assumptions A1 and A7. Assumptions (A.7)–(A.10) imply that

$$A_2 = O(d_n \mathbf{h}!) + o(1) = o(1).$$

Hence,

$$\mathrm{Var}(\mathbf{c}' Z_T) = \mu_{02}^q f(\mathbf{x}) \mathbf{c}' D_k^+ \boldsymbol{\Omega}(\mathbf{x}) D_k^{+\prime} \mathbf{c} + o(1).$$

Using the standard Doob's small-block and large-block technique, we can finish the rest of the normality proof of Assumption (A.4) by following the arguments of Cai, Fan, and Yao (2000, pp. 954–955).

Lemma A.2. Under Assumptions A1–A7,

$$\mathrm{vech}(\hat{\mathbf{G}}_{np}(\mathbf{x})) - \mathrm{vech}(\widetilde{\mathbf{G}}_{np}(\mathbf{x})) = o_P\big((T\mathbf{h}!)^{-1/2}\big).$$

Proof. Let $\Delta(\mathbf{x}) = [\text{vec}(\hat{\mathbf{G}}_{np}(\mathbf{x})) - \text{vec}(\widetilde{\mathbf{G}}_{np}(\mathbf{x}))]\hat{f}(\mathbf{x})$. Noting that $\hat{f}(\mathbf{x}) \xrightarrow{p} f(\mathbf{x}) > 0$ and $\text{vech}(\mathbf{A}) = D_k^+ \text{vec}(\mathbf{A})$ for any symmetric $k \times k$ matrix $\mathbf{A}$, it suffices to show that $\Delta(\mathbf{x}) = o_P((T\mathbf{h}!)^{-1/2})$. By the first order Taylor expansion,

$$\hat{\mathbf{e}}_t = \mathbf{e}_t(\hat{\theta}) = \mathbf{H}_{p,t}^{-1/2}(\hat{\theta})\mathbf{r}_t = \mathbf{e}_t + \boldsymbol{\xi}_t(\overline{\theta})(\hat{\theta} - \theta_*), \tag{A.11}$$

where recall that $\boldsymbol{\xi}_t(\theta) = \partial \mathbf{e}_t(\theta)/\partial\theta'$, and $\overline{\theta}$ lies between $\hat{\theta}$ and θ_*. By Assumptions A2 and A3, $\overline{\theta} \xrightarrow{p} \theta_*$. So

$$\begin{aligned}
\Delta(\mathbf{x}) &= \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\,\text{vec}[\hat{\mathbf{e}}_t\hat{\mathbf{e}}_t' - \mathbf{e}_t\mathbf{e}_t'] \\
&= \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\,\text{vec}[\boldsymbol{\xi}_t(\overline{\theta})(\hat{\theta} - \theta_*)(\hat{\theta} - \theta_*)'\xi_t(\overline{\theta})'] \\
&\quad + \frac{2}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\,\text{vec}[\mathbf{e}_t(\hat{\theta} - \theta_*)'\xi_t(\overline{\theta})'] \\
&= \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})(\boldsymbol{\xi}_t(\overline{\theta}) \otimes \boldsymbol{\xi}_t(\overline{\theta})) \\
&\quad \times \text{vec}[(\hat{\theta} - \theta_*)(\hat{\theta} - \theta_*)'] \\
&\quad + \frac{2}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})(\boldsymbol{\xi}_t(\overline{\theta}) \otimes \mathbf{e}_t)(\hat{\theta} - \theta_*) \\
&\equiv \Delta_1(\mathbf{x}) + 2\Delta_2(\mathbf{x}).
\end{aligned}$$

By the triangle inequality, Markov inequality, and Assumptions A4–A7,

$$\begin{aligned}
\|\Delta_1(\mathbf{x})\| &\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}) \\
&\quad \times \left\|(\boldsymbol{\xi}_t(\overline{\theta}) \otimes \boldsymbol{\xi}_t(\overline{\theta}))\,\text{vec}[(\hat{\theta} - \theta_*)(\hat{\theta} - \theta_*)']\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\|\boldsymbol{\xi}_t(\overline{\theta})\|^2\|\hat{\theta} - \theta_*\|^2 \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\overline{D}_t^2\|\hat{\theta} - \theta_*\|^2 = O_P\left(\frac{1}{T\mathbf{h}!}\right)
\end{aligned}$$

and

$$\begin{aligned}
\|\Delta_2(\mathbf{x})\| &\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left\|(\boldsymbol{\xi}_t(\overline{\theta}) \otimes \mathbf{e}_t)(\hat{\theta} - \theta_*)\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\|\boldsymbol{\xi}_t(\overline{\theta})\|\|\mathbf{e}_t\|\|\hat{\theta} - \theta_*\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\overline{D}_t\|\mathbf{e}_t\|\|\hat{\theta} - \theta_*\| = O_P\big(T^{-1/2}\big).
\end{aligned}$$

Consequently, $\Delta(\mathbf{x}) = O_P((T\mathbf{h}!)^{-1} + T^{-1/2}) = o_P((T\mathbf{h}!)^{-1/2})$.

Proof of Theorem 3.1

The result follows from Lemmas A.1–A.2.

Proof of Corollary 3.2

By Assumptions A3–A5, $\hat{\mathbf{H}}_{p,t} = \mathbf{H}_{p,t}^{1/2}(\hat{\theta}) = \mathbf{H}_{p,t}^{1/2} + o_P(1)$. By Theorem 3.1, $\hat{\mathbf{G}}_{np,t} = \hat{\mathbf{G}}_{np}(\mathbf{x}_t) = \mathbf{G}_{np,t} + o_P(1)$. It follows from the Slutsky theorem that

$$\begin{aligned}
\hat{\mathbf{H}}_{sp,t} &= \hat{\mathbf{H}}_{p,t}^{1/2}\hat{\mathbf{G}}_{np,t}\hat{\mathbf{H}}_{p,t}^{1/2} = \mathbf{H}_{p,t}^{1/2}\mathbf{G}_{np,t}\mathbf{H}_{p,t}^{1/2} + o_P(1) \\
&= \mathbf{H}_t + o_P(1),
\end{aligned}$$

and $\hat{\mathbf{H}}_{sp,t}^* = \mathbf{H}_t^* + o_P(1)$, where $\mathbf{H}_t^*$ is a diagonal matrix with the square roots of the diagonal elements of $\mathbf{H}_t$ on its diagonal. Hence $\hat{\mathbf{R}}_{sp,t} = (\hat{\mathbf{H}}_{sp,t}^*)^{-1}\hat{\mathbf{H}}_{sp,t}(\hat{\mathbf{H}}_{sp,t}^*)^{-1} \xrightarrow{p} (\mathbf{H}_t^*)^{-1}\mathbf{H}_t(\mathbf{H}_t^*)^{-1} = \mathbf{R}_t$.

To show (ii), noting that by Assumptions A3–A5,

$$\begin{aligned}
\hat{\mathbf{H}}_{sp,t} - \mathbf{H}_t &= \hat{\mathbf{H}}_{p,t}^{1/2}\hat{\mathbf{G}}_{np,t}\hat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\mathbf{G}_{np}(\mathbf{x}_t)\mathbf{H}_{p,t}^{1/2} \\
&= \mathbf{H}_{p,t}^{1/2}(\hat{\mathbf{G}}_{np}(\mathbf{x}_t) - \mathbf{G}_{np}(\mathbf{x}_t))\mathbf{H}_{p,t}^{1/2} \\
&\quad + \big\{\big(\hat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\big)\hat{\mathbf{G}}_{np,t}\big(\hat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\big) \\
&\quad + \big(\hat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\big)\hat{\mathbf{G}}_{np,t}\mathbf{H}_{p,t}^{1/2} \\
&\quad + \mathbf{H}_{p,t}^{1/2}\hat{\mathbf{G}}_{np,t}\big(\hat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\big)\big\} \\
&= \mathbf{H}_{p,t}^{1/2}(\hat{\mathbf{G}}_{np}(\mathbf{x}_t) - \mathbf{G}_{np}(\mathbf{x}_t))\mathbf{H}_{p,t}^{1/2} + O_p\big(T^{-1/2}\big),
\end{aligned}$$

we have

$$\begin{aligned}
&\sqrt{T\mathbf{h}!}[\text{vech}(\hat{\mathbf{H}}_{sp,t}) - \text{vech}(\mathbf{H}_t)] \\
&= \sqrt{T\mathbf{h}!}D_k^+[\text{vec}(\hat{\mathbf{H}}_{sp,t}) - \text{vec}(\mathbf{H}_t)] \\
&= \sqrt{T\mathbf{h}!}D_k^+\,\text{vec}\big(\mathbf{H}_{p,t}^{1/2}(\hat{\mathbf{G}}_{np}(\mathbf{x}_t) - \mathbf{G}_{np}(\mathbf{x}_t))\mathbf{H}_{p,t}^{1/2}\big) + o_P(1) \\
&= \sqrt{T\mathbf{h}!}D_k^+\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)\,\text{vec}(\hat{\mathbf{G}}_{np}(\mathbf{x}_t) - \mathbf{G}_{np}(\mathbf{x}_t)) + o_P(1) \\
&= \sqrt{T\mathbf{h}!}D_k^+\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)D_k\,\text{vech}(\hat{\mathbf{G}}_{np}(\mathbf{x}_t) - \mathbf{G}_{np}(\mathbf{x}_t)) \\
&\quad + o_P(1).
\end{aligned}$$

Then by Theorem 3.1,

$$\begin{aligned}
&\sqrt{T\mathbf{h}!}[\text{vech}(\hat{\mathbf{H}}_{sp,t}) - \text{vech}(\mathbf{H}_t) - \overline{\mathbf{B}}_t(\mathbf{x}_t)] \\
&\qquad \xrightarrow{d} \text{MN}(0, \mu_{02}^q f(\mathbf{x}_t)^{-1}\overline{\overline{\boldsymbol{\Omega}}}_t(\mathbf{x})),
\end{aligned}$$

where

$$\begin{aligned}
\overline{\mathbf{B}}_t(\mathbf{x}) &= D_k^+\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)D_k\,\text{vech}(\mathbf{B}(\mathbf{x})) \\
&= D_k^+\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)\,\text{vec}(\mathbf{B}(\mathbf{x})) \\
&= D_k^+\,\text{vec}\big(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x})\mathbf{H}_{p,t}^{1/2}\big) = \text{vech}\big(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x})\mathbf{H}_{p,t}^{1/2}\big),
\end{aligned}$$

by the definitions of vech, vec, D_k, and D_k^+ and the fact that $(\mathbf{A} \otimes \mathbf{A})\,\text{vec}(\mathbf{B}(\mathbf{x})) = \text{vec}(\mathbf{A}\mathbf{B}(\mathbf{x})\mathbf{A})$ for any $k \times k$ matrix $\mathbf{A}$, and

$$\begin{aligned}
\overline{\overline{\boldsymbol{\Omega}}}_t(\mathbf{x}) &= D_k^+\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)D_kD_k^+\boldsymbol{\Omega}(\mathbf{x})D_k^{+\prime}D_k' \\
&\quad \times \big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)(D_k^+)' \\
&= D_k^+N_k\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)\boldsymbol{\Omega}(\mathbf{x})\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)N_k(D_k^+)' \\
&= D_k^+\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)\boldsymbol{\Omega}(\mathbf{x})\big(\mathbf{H}_{p,t}^{1/2} \otimes \mathbf{H}_{p,t}^{1/2}\big)(D_k^+)' \\
&= D_k^+\overline{\boldsymbol{\Omega}}_t(\mathbf{x})(D_k^+)',
\end{aligned}$$

by the fact that $N_k \equiv D_k D_k^+$ is symmetric, $N_k D_k = D_k$, $N_k D_k^{+\prime} = D_k^{+\prime}$, and $N_k(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A}) N_k$ for any $k \times k$ matrix $\mathbf{A}$.

Proof of Theorem 3.3

Let $\boldsymbol{\iota}_i$ denote a $k \times 1$ vector that has 1 in the ith row and 0 elsewhere. Then

$$\hat{e}_{it} = \boldsymbol{\iota}_i' \hat{\mathbf{e}}_t = \boldsymbol{\iota}_i' \hat{\mathbf{H}}_{p,t}^{-1/2} \mathbf{r}_t = \boldsymbol{\iota}_i' \mathbf{H}_{p,t}^{-1/2} \mathbf{r}_t + \boldsymbol{\iota}_i' \big(\hat{\mathbf{H}}_{p,t}^{-1/2} - \mathbf{H}_{p,t}^{-1/2}\big) \mathbf{r}_t = e_{it} + \nu_{it},$$

where $\nu_{it} = \boldsymbol{\iota}_i'(\hat{\mathbf{H}}_{p,t}^{-1/2} - \mathbf{H}_{p,t}^{-1/2}) \mathbf{r}_t$. Note that, for notational simplicity, we have suppressed the dependence of ν_{it} on the sample size T. It follows that

$$\frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{it}\hat{e}_{jt} - \delta_{ij}) K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} [(e_{it} + \nu_{jt})(e_{it} + \nu_{jt}) - \delta_{ij}] K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}) = \sum_{l=1}^{4} A_{ij,l}(\mathbf{x}),$$

and

$$\frac{1}{T^2} \sum_{t=1}^{T} (\hat{e}_{it}\hat{e}_{jt} - \delta_{ij})^2 \overline{K}_{\mathbf{h}}(\mathbf{0}) = \frac{1}{T} \sum_{t=1}^{T} [(e_{it} + \nu_{it})(e_{jt} + \nu_{jt}) - \delta_{ij}]^2 \overline{K}_{\mathbf{h}}(\mathbf{0}) = \sum_{l=1}^{4} B_{ij,l},$$

where

$$A_{ij,1}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} (e_{it}e_{jt} - \delta_{ij}) K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}),$$

$$B_{ij,1} = \frac{1}{T} \sum_{t=1}^{T} (e_{it}e_{jt} - \delta_{ij})^2 \overline{K}_{\mathbf{h}}(\mathbf{0}),$$

$$A_{ij,2}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} \nu_{it}\nu_{jt} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}),$$

$$B_{ij,2} = \frac{1}{T} \sum_{t=1}^{T} \nu_{it}^2\nu_{jt}^2 \overline{K}_{\mathbf{h}}(\mathbf{0}),$$

$$A_{ij,3}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} e_{it}\nu_{jt} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}),$$

$$B_{ij,3} = \frac{1}{T} \sum_{t=1}^{T} e_{it}^2\nu_{jt}^2 \overline{K}_{\mathbf{h}}(\mathbf{0}),$$

$$A_{ij,4}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} e_{jt}\nu_{it} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}),$$

$$B_{ij,4} = \frac{1}{T} \sum_{t=1}^{T} e_{jt}^2\nu_{it}^2 \overline{K}_{\mathbf{h}}(\mathbf{0}).$$

Consequently,

$$\begin{aligned}
\hat{\Gamma} &= \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int \left[\frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{it}\hat{e}_{jt} - \delta_{ij}) K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}) \right]^2 d\mathbf{x} \\
&\quad - \frac{1}{T^2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{t=1}^{T} (\hat{e}_{it}\hat{e}_{jt} - \delta_{ij})^2 \overline{K}_{\mathbf{h}}(\mathbf{0}) \\
&= \sum_{i=1}^{k-1} \sum_{j=i}^{k} \Bigg\{ \Bigg[\int \sum_{l=1}^{4} A_{ij,l}^2(\mathbf{x}) + 2A_{ij,1}(\mathbf{x})A_{ij,2}(\mathbf{x}) \\
&\quad + 2A_{ij,1}(\mathbf{x})A_{ij,3}(\mathbf{x}) + 2A_{ij,1}(\mathbf{x})A_{ij,4}(\mathbf{x}) + 2A_{ij,2}(\mathbf{x})A_{ij,3}(\mathbf{x}) \\
&\quad + 2A_{ij,2}(\mathbf{x})A_{ij,4}(\mathbf{x}) + 2A_{ij,3}(\mathbf{x})A_{ij,4}(\mathbf{x}) \Bigg] d\mathbf{x} - \sum_{l=1}^{4} B_{ij,l} \Bigg\}.
\end{aligned}$$

Then we can write $T(\mathbf{h}!)^{1/2}\hat{\Gamma} = \sum_{l=1}^{10} C_{lT}$, where

$$C_{lT} = T(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \left\{ \int A_{ij,l}^2(\mathbf{x})\, d\mathbf{x} - B_{ij,l} \right\} \quad \text{for } l = 1, 2, 3, 4,$$

$$C_{lT} = T(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int A_{ij,1}(\mathbf{x}) A_{ij,l-3}(\mathbf{x})\, d\mathbf{x} \quad \text{for } l = 5, 6, 7,$$

$$C_{lT} = T(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int A_{ij,2}(\mathbf{x}) A_{ij,l-5}(\mathbf{x})\, d\mathbf{x} \quad \text{for } l = 8, 9,$$

and

$$C_{10T} = T(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int A_{ij,3}(\mathbf{x}) A_{ij,4}(\mathbf{x})\, d\mathbf{x}.$$

The proof will be completed if we can show $C_{1T} \xrightarrow{d} \mathrm{N}(0, \sigma_0^2)$, and $C_{lT} = o_P(1)$ for $l = 2, 3, \ldots, 10$. We only prove $C_{1T} \xrightarrow{d} \mathrm{N}(0, \sigma_0^2)$ and $C_{lT} = o_P(1)$ for $l = 2, 3, 5$ since the other cases are similar.

We first show that $C_{1T} \xrightarrow{d} \mathrm{N}(0, \sigma_0^2)$. Let $\boldsymbol{\varsigma}_t = (\mathbf{x}_t', \mathbf{e}_t')'$ and $\phi(\boldsymbol{\varsigma}_t, \boldsymbol{\varsigma}_s) = (\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} (e_{it}e_{jt} - \delta_{ij})(e_{is}e_{js} - \delta_{ij}) \overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s)$. We can write $C_{1T} = 2T^{-1} \sum_{1 \le t < s \le T} \phi(\boldsymbol{\varsigma}_t, \boldsymbol{\varsigma}_s)$, which is a second order U-statistic and is degenerate under the null. Under Assumptions A1, A4, and A6–A9 and the null hypothesis, one can verify the conditions of lemma B.1 in Gao and King (2003) are satisfied so that a central limit theorem applies to C_{1T}. The asymptotic variance is given by $\lim_{n\to\infty} 2E[\phi(\overline{\boldsymbol{\varsigma}}_t, \boldsymbol{\varsigma}_t)^2] = \sigma_0^2$, where $\overline{\boldsymbol{\varsigma}}_t$ is an independent copy of $\boldsymbol{\varsigma}_t$.

To show $C_{2T} = o_P(1)$, write

$$C_{2T} = T^{-1}(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \ne s}^{T} \nu_{is}\nu_{js}\nu_{it}\nu_{jt} \overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s).$$

By Assumptions (A.11) and A5, $|\nu_{it}| = |\boldsymbol{\iota}_i'(\hat{\mathbf{e}}_t - \mathbf{e}_t)| = |\boldsymbol{\iota}_i' \times \boldsymbol{\xi}_t(\overline{\theta})(\hat{\theta} - \theta_*)| \le \overline{D}_t \|\hat{\theta} - \theta_*\|$, where $\boldsymbol{\xi}_t(\theta) = \partial \mathbf{e}_t(\theta)/\partial\theta'$ and $\overline{\theta}$

lies between $\hat\theta$ and θ_*. By Assumptions A5 and A3,

$$|C_{2T}| \le \frac{k(k+1)}{2} T^{-1}(\mathbf{h}!)^{1/2} \times \sum_{s=1}^{T}\sum_{t\neq s}^{T} \overline{D}_t^2\overline{D}_s^2\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)\|\hat\theta-\theta_*\|^4 = O_p\big(T(\mathbf{h}!)^{1/2}\big)O_p(T^{-2}) = o_P(1),$$

where the second line follows from a simple application of the Markov inequality, and the fact that for $t\neq s$

$$E[\overline{D}_t^2\overline{D}_s^2\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)] \le \big\{E[\overline{D}_t^4\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)]\big\}^{1/2}\big\{E[\overline{D}_s^4\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)]\big\}^{1/2} = O(1).$$

Similarly, noting that

$$C_{3T} = T^{-1}(\mathbf{h}!)^{1/2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T} e_{it}v_{jt}e_{is}v_{js}\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s),$$

we have,

$$|C_{3T}| \le \frac{k(k+1)}{2}T^{-1}(\mathbf{h}!)^{1/2} \times \left|\sum_{s=1}^{T}\sum_{t=1}^{T}\|\mathbf{e}_t\|\|\mathbf{e}_s\|\overline{D}_t\overline{D}_s\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)\right|\|\hat\theta-\theta_*\|^2 = O_p\big(T(\mathbf{h}!)^{1/2}\big)O_p(T^{-1}) = o_P(1).$$

Noting that $C_{5T} = T^{-1}(\mathbf{h}!)^{1/2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t=1}^{T}(e_{is}e_{js}-\delta_{ij})v_{it}v_{jt}\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)$, we can write $C_{5T}=C_{5T,a}+C_{5T,b}$, where

$$C_{5T,a} = T^{-1}(\mathbf{h}!)^{1/2} \times \sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}(e_{is}e_{js}-\delta_{ij})v_{it}v_{jt}\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)$$

and

$$C_{5T,b} = T^{-1}(\mathbf{h}!)^{1/2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{t=1}^{T}(e_{it}e_{jt}-\delta_{ij})v_{it}v_{jt}\overline{K}_{\mathbf{h}}(\mathbf{0}).$$

By Assumptions A3, A5, and A8, and the Markov inequality,

$$|C_{5T,a}| \le \frac{k(k+1)}{2}T^{-1}(\mathbf{h}!)^{1/2} \times \sum_{s=1}^{T}\sum_{t\neq s}^{T}(\|\mathbf{e}_s\|^2+1)\overline{D}_t^2\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)\|\hat\theta-\theta_*\|^2 = O_p\big(T(\mathbf{h}!)^{1/2}\big)O_p(T^{-1}) = o_P(1)$$

and

$$|C_{5T,b}| \le \frac{k(k+1)}{2}T^{-1}(\mathbf{h}!)^{1/2} \times \sum_{t=1}^{T}(\|\mathbf{e}_t\|^2+1)\overline{D}_t^2\overline{K}_{\mathbf{h}}(\mathbf{0})\|\hat\theta-\theta_*\|^2 = O_p\big((\mathbf{h}!)^{-1/2}\big)O_p(T^{-1}) = o_P(1).$$

Consequently, $C_{5T}=o_P(1)$. This concludes the proof of the theorem.

Proof of Theorem 3.4

Under $H_1(T^{-1/2}(\mathbf{h}!)^{-1/4})$, the expression $T(\mathbf{h}!)^{1/2}\hat\Gamma = \sum_{l=1}^{10}C_{lT}$ obtained in the proof of Theorem 3.3 continues to hold. In addition, one can verify that under $H_1(T^{-1/2}(\mathbf{h}!)^{-1/4})$, $C_{lT}=o_P(1)$ continues to hold for $l=2,3,\ldots,10$. The main change is associated with the term C_{1T}. Let $\epsilon_{ijt}=e_{it}e_{jt}-\delta_{ij}$. Let $E_t(\epsilon_{ijt})$ denote the conditional expectation of ϵ_{ijt} given $\mathcal{F}_{t-1}$ and $\overline{\epsilon}_{ijt}=\epsilon_{ijt}-E_t(\epsilon_{ijt})$. Then we can write $C_{1T}=C_{1T,a}+C_{1T,b}+C_{1T,c}$, where

$$C_{1T,a} = T^{-1}(\mathbf{h}!)^{1/2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}\overline{\epsilon}_{ijs}\overline{\epsilon}_{ijt}\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s),$$

$$C_{1T,b} = T^{-1}(\mathbf{h}!)^{1/2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}E_s(\epsilon_{ijs})E_t(\epsilon_{ijt})\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s),$$

and

$$C_{1T,c} = 2T^{-1}(\mathbf{h}!)^{1/2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}\overline{\epsilon}_{ijs}E_t(\epsilon_{ijt})\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s).$$

$C_{1T,a}$ now plays the role of C_{1T} in the proof of Theorem 3.3, and we can show that $C_{1T,a}\xrightarrow{d}\mathrm{N}(0,\sigma_0^2)$. Next, noting that under $H_1(T^{-1/2}(\mathbf{h}!)^{-1/4})$, $E_t(\epsilon_{ijt})=T^{-1/2}(\mathbf{h}!)^{-1/4}\Delta_{ij}(\mathbf{x}_t)$, we have

$$\begin{aligned} C_{1T,b} &= T^{-2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}\Delta_{ij}(\mathbf{x}_s)\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s) \\ &= \frac{T-1}{T}\frac{2}{T(T-1)}\sum_{1\le t<s\le T}\varphi(\mathbf{x}_t,\mathbf{x}_s) \\ &\equiv \frac{T-1}{T}\widetilde{C}_{1T,b}, \end{aligned} \tag{A.12}$$

where $\varphi(\mathbf{x}_t,\mathbf{x}_s)=\sum_{i=1}^{k-1}\sum_{j=i}^{k}\Delta_{ij}(\mathbf{x}_s)\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)$. Noticing that $\widetilde{C}_{1T,b}$ is a second order U-statistic, a typical WLLN for a U-statistic of strong mixing process (e.g., Borovkova, Burton, and Dehling 1999) will require that $\{\varphi(\mathbf{x}_t,\mathbf{x}_s):t,s\ge 1, t\neq s\}$ be uniformly integrable, which is difficult to verify here. By the H-decomposition, we can write

$$\widetilde{C}_{1T,b} = \vartheta_T + 2H_T^{(1)} + H_T^{(2)}, \tag{A.13}$$

where $\vartheta_T=\int\int\varphi(\mathbf{x}_t,\mathbf{x}_s)f(\mathbf{x}_t)f(\mathbf{x}_s)\,d\mathbf{x}_t\,d\mathbf{x}_s$, $H_T^{(1)}=(1/T)\times\sum_{t=1}^{T}\varphi_1(\mathbf{x}_t)-\vartheta_T$, $H_T^{(2)}=\frac{2}{T(T-1)}\sum_{1\le t<s\le T}\overline{\varphi}(\mathbf{x}_t,\mathbf{x}_s)$, $\varphi_1(\mathbf{x}_t)=\int\varphi(\mathbf{x}_t,\mathbf{x}_s)f(\mathbf{x}_s)\,d\mathbf{x}_s$, and $\overline{\varphi}(\mathbf{x}_t,\mathbf{x}_s)=\varphi(\mathbf{x}_t,\mathbf{x}_s)-\varphi_1(\mathbf{x}_t)-\varphi_1(\mathbf{x}_s)+\vartheta_T$. By the Fubini theorem and the change of variables, we have

$$\begin{aligned} \vartheta_T &= \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int\int\Delta_{ij}(\mathbf{x}_s)\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s) \times f(\mathbf{x}_t)f(\mathbf{x}_s)\,d\mathbf{x}_t\,d\mathbf{x}_s \\ &= \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int\Delta_{ij}^2(\mathbf{x})f^2(\mathbf{x})\,d\mathbf{x}+o(1). \end{aligned} \tag{A.14}$$

Note that $\varphi_1(\mathbf{x}_t)$ is a measurable function of $\mathbf{x}_t$ and inherits the α-mixing property of the latter. By Assumption A1, $\varphi_1(\mathbf{x}_t)$ is a strictly stationary α-mixing process with mixing coefficient $\alpha(j) \to 0$ as $j \to \infty$. By proposition 3.44 of White (2001), ς_t is also ergodic. Furthermore, it is easy to verify that $E|\varphi_1(\mathbf{x}_t)| < \infty$. It follows from the Ergodic theorem (e.g., White 2001, theorem 3.34) that

$$H_T^{(1)} \xrightarrow{p} 0. \tag{A.15}$$

Now, $H_T^{(2)}$ is a standard second order degenerate U-statistic with a symmetric kernel $\overline{\varphi}(\cdot,\cdot) : \overline{\varphi}(\mathbf{x}_t, \mathbf{x}_s) = \overline{\varphi}(\mathbf{x}_s, \mathbf{x}_t)$ and $E\overline{\varphi}(\mathbf{x}_1, \mathbf{a}) = 0$ for any nonrandom $\mathbf{a} \in \mathbb{R}^q$. Noting that

$$\max_{1<t\le T} \max\left\{E|\overline{\varphi}(\mathbf{x}_1,\mathbf{x}_t)|^{2(1+\delta)}, \int |\overline{\varphi}(\mathbf{x}_1,\mathbf{x}_t)|^{2(1+\delta)}\,dF(\mathbf{x}_1)\,dF(\mathbf{x}_t)\right\} = O\big((\mathbf{h}!)^{-(1+2\delta)}\big),$$

where $F(\cdot)$ is the distribution function of $\mathbf{x}_t$, it follows from lemma C.2 of Gao and King (2003) that

$$E\big[H_T^{(2)}\big]^2 \le C\left(\frac{2}{T(T-1)}\right)^2 T^2(\mathbf{h}!)^{-(1+2\delta)/(1+\delta)} = O\big(T^{-2}(\mathbf{h}!)^{-(1+2\delta)/(1+\delta)}\big) = o(1).$$

Hence by the Chebyshev inequality

$$H_T^{(2)} = o_P(1). \tag{A.16}$$

Combining Assumptions (A.12)–(A.16) yields $C_{1T,b} \xrightarrow{p} \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int \Delta_{ij}^2(\mathbf{x})f^2(\mathbf{x})\,d\mathbf{x} \equiv \Delta_0$.

Now, write $C_{1T,c} = C_{1T,c1} + C_{1T,c2}$, where $C_{1T,c1} = 2 \times T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\le t<s\le T}^{T}\overline{\epsilon}_{ijs}\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)$ and $C_{1T,c2} = 2T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\le s<t\le T}^{T}\overline{\epsilon}_{ijs}\Delta_{ij}(\mathbf{x}_t) \times \overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)$. By construction $E(\overline{\epsilon}_{ijs}|\mathcal{F}_{s-1}) = 0$. It follows that $E(C_{1T,c1}) = 0$ by the law of iterated expectations and the hypothesis that $(\mathbf{x}_t, \mathbf{x}_s) \in \mathcal{F}_{s-1}$ for $t < s$. By Davydov's inequality (e.g., Bosq 1996, p. 19), we have

$$\begin{aligned}
E(C_{1T,c2}) &= 2T^{-3/2}(\mathbf{h}!)^{1/4} \times \sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\le s<t\le T}^{T} E[\overline{\epsilon}_{ijs}\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x}_s)] \\
&= 2T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\le s<t\le T}^{T}\int E[\overline{\epsilon}_{ijs}K_{\mathbf{h}}(\mathbf{x}_s-\mathbf{x}) \times \Delta_{ij}(\mathbf{x}_t)K_{\mathbf{h}}(\mathbf{x}_t-\mathbf{x})]\,d\mathbf{x} \\
&\le CT^{-1/2}(\mathbf{h}!)^{1/4}(\mathbf{h}!)^{-2(1+\delta)/(2+\delta)}\sum_{j=1}^{T-1}[\alpha(i)]^{\delta/(2+\delta)} \\
&= o(1) \quad \text{for sufficiently small } \delta > 0.
\end{aligned}$$

Similarly, we can show that $E(C_{1T,c1}^2) = o(1)$ and $E(C_{1T,c2}^2) = o(1)$. Then $C_{1T,c} = o_P(1)$ by the Chebyshev inequality.

Consequently, $P(\hat{T} \ge z_\alpha | H_1(T^{-1/2}(\mathbf{h}!)^{-1/4})) \to 1 - \Phi(z_\alpha - \Delta_0/\sigma_0)$.

ACKNOWLEDGMENTS

The authors thank Serena Ng, the associate editor, and an anonymous referee for their many insightful comments that led to a great improvement of the presentation. The authors also thank the conference and seminar participants at the Far Eastern and South Asian Meeting of the Econometric Society (2008), the Australasian Meeting of Econometric Society (2006), Catholic University of Louvain, European FMA conference (2005), Forum of Interdisciplinary Mathematics (FIM) Portugal, Indiana University, University of Cambridge, and University of Oxford for their comments. All errors are the authors' responsibilities. The second author gratefully acknowledges the financial support from the NSFC under the grant numbers 70501001 and 70601001. The third author acknowledges the financial support from the academic senate, UCR.

[Received March 2007. Revised May 2009.]

REFERENCES

Ang, A., and Chen, J. (2002), "Asymmetric Correlations of Equity Portfolios," *Journal of Financial Economics,* 63 (3), 443–494. [109]

Awartani, B. M. A., and Corradi, V. (2005), "Predicting the Volatility of the S&P-500 Index via GARCH Models: The Role of Asymmetries," *International Journal of Forecasting*, 21 (1), 167–184. [118]

Baba, Y., Engle, R. F., Kraft, D. F., and Kroner, K. F. (1991), "Multivariate Simultaneous Generalized ARCH," manuscript, UCSD, Dept. of Economics. [109]

Bauwens, L., and Laurent, S. (2005), "A New Class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroscedasticity Models," *Journal of Business & Economic Statistics*, 23, 346–354. [116]

Bollerslev, T. (1990), "Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach," *Review of Economics and Statistics*, 72, 498–505. [111]

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988), "A Capital Asset Pricing Model With Time-Varying Covariance," *Journal of Political Economy*, 96 (1), 116–131. [109]

Borovkova, S., Burton, R., and Dehling, H. (1999), "Consistency of the Takens Estimator for the Correlation Dimension," *The Annals of Applied Probability*, 9, 376–390. [123]

Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, New York: Springer. [120,124]

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–956. [120]

Cappiello, L., Engle, R. F., and Sheppard, K. (2003), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," Working Paper 204, ECB. [109]

Engle, R. F. (1982), "Autoregressive, Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007. [109]

——— (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 20, 339–350. [109-111, 115,116]

Engle, R. F., and Kroner, K. F. (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150. [109]

Engle, R. F., and Sheppard, K. (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," Discussion Paper 2001-15, UCSD, Dept. of Economics. [109]

Engle, R. F., Ng, V. R., and Rothschild, M. (1990), "Asset Pricing With a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213–238. [109]

Fama, E., and French, K. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56. [109]

Gao, J., and King, M. (2003), "Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression Models," mimeo, University of Western Australia, Dept. of Mathematics and Statistics. [115,122,124]

Glad, I. K. (1998), "Parametrically Guided Non-Parametric Regression," *Scandinavian Journal of Statistics*, 25, 649–668. [109,111]

Hafner, C., van Dijk, D., and Franses, P. (2006), "Semiparametric Modelling of Correlation Dynamics," in *Advances in Econometrics*, Vol. 20, Part A, eds. T. Fomby and C. Hill, pp. 59–103. [110,116]

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–631. [116]

Lavergne, P., and Vuong, Q. (2000), "Nonparametric Significance Testing," *Econometric Theory*, 16, 576–601. [114]

Lee, T. H., and Long, X. D. (2009), "Copula-Based Multivariate GARCH Model With Uncorrelated Dependent Errors," *Journal of Econometrics*, 150, 207–218. [109]

Longin, F., and Solnik, B. (2001), "Extreme Correlation of International Equity Market," *Journal of Finance*, 56, 649–679. [109]

Magnus, J. R., and Neudecker, H. (1999), *Matrix Differential Calculus With Applications in Statistics and Econometrics*, New York: Wiley. [110]

Mashal, R., and Zeevi, A. (2002), "Beyond Correlation: Extreme Co-Movements Between Financial Assets," manuscript, Columbia University. [109]

Mishra, S., Su, L., and Ullah, A. (2010), "Combined Estimator of Time Series Conditional Heteroskedasticity," *Journal of Business & Economic Statistics*, to appear. [109-114]

Olkin, I., and Spiegelman, C. (1987), "A Semiparametric Approach to Density Estimation," *Journal of the American Statistical Association*, 88, 858–865. [109]

Pelletier, D. (2006), "Regime Switching for Dynamic Correlations," *Journal of Econometrics*, 131, 445–473. [109,118]

Richardson, M. P., and Smith, T. (1993), "A Test of Multivariate Normality of Stock Returns," *Journal of Business*, 66, 295–321. [109]

Silvennoinen, A., and Teräsvirta, T. (2005), "Multivariate Autoregressive Conditional Heteroskedasticity With Smooth Transitions in Conditional Correlations," Working Paper Series in Economics and Finance 577, Stockholm School of Economics. [109]

Su, L., and Ullah, A. (2009), "Testing Conditional Uncorrelatedness," *Journal of Business & Economic Statistics*, 27, 18–29. [114-116]

Su, L., and White, H. (2007), "A Consistent Characteristic Function-Based Test for Conditional Independence," *Journal of Econometrics*, 141, 807–834. [114]

Tse, Y. K., and Tsui, A. K. (2002), "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations," *Journal of Business & Economic Statistics*, 20, 351–362. [111]

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge: Cambridge University Press. [109,113]

——— (2001), *Asymptotic Theory for Econometricians* (2nd ed.), San Diego: Academic Press. [124]

Zangari, P. (1997), "Streamlining the Market Risk Measurement Process," *RiskMetrics Monitor*, 1, 29–35. [118]