We introduce dynamic orthogonal components (DOC) for multivariate time series and propose a procedure for estimating and testing the existence of DOCs for a given time series. We estimate the dynamic orthogonal components via a generalized decorrelation method that minimizes the linear and quadratic dependence across components and across time. We then use Ljung–Box type statistics to test the existence of dynamic orthogonal components. When DOCs exist, univariate analysis can be applied to build a model for each component. Those univariate models are then combined to obtain a multivariate model for the original time series. We demonstrate the usefulness of dynamic orthogonal components with two real examples and compare the proposed modeling method with other dimension-reduction methods available in the literature, including principal component and independent component analyses. We also prove consistency and asymptotic normality of the proposed estimator under some regularity conditions. We provide some technical details in online Supplementary Materials.

KEY WORDS: Conditional heteroscedasticity; Dimension reduction; Generalized decorrelation; Independent component analysis; Principal component analysis; Vector autoregression.

1. INTRODUCTION

Reducing the parameter space is essential for successfully modeling multivariate time series, because the number of parameters involved increases quickly with the dimension of the series. One commonly used approach for reducing the number of parameters is dimension reduction, which seeks a simplifying structure in a vector time series. For linear time series without conditional heteroscedasticity, several methods are available to perform dimension reduction, including the canonical correlation analysis (CCA) of Box and Tiao (1977), the factor models of Peña and Box (1987), the independent components analysis (ICA) of Back and Weigend (1997), and the principal components analysis (PCA) of Stock and Watson (2002). These methods seek linear combinations that have certain characteristics useful in model building; for instance, the CCA of Box and Tiao (1977) produces linear combinations that rank from the most predictable to the least predictable. An alternative approach to reducing the number of parameters is structural specification that identifies a parsimonious representation for the vector series, for example, the scalar component models of Tiao and Tsay (1989).

Conditional heteroscedasticity is common in economic and financial data, and in many applications (e.g., portfolio optimization) is the main focus of the analysis. A proper analysis of conditional heteroscedasticity requires modeling the volatility of the series. By volatility, we mean the conditional standard deviation of the series (see Bauwens, Laurent, and Rombouts 2006 for a review of multivariate volatility models). For a $d$-dimensional series, volatility analysis requires modeling $d(d+1)/2$ conditional variances and conditional covariances processes under a positive-definite constraint; thus dimension reduction becomes even more important. Indeed, PCA and ICA have been used by Alexander (2001) and Chen, Härdle, and Spokoiny (2007), respectively. An alternative approach is to enforce strong assumptions on the dynamic structure of the volatility; for example, Tse and Tsui (2002) and Engle (2002) used highly parameterized dynamic correlation models to govern the time evolution of conditional correlations after individual time series have been normalized by the generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986). Several orthogonal factor models that reduce the number of parameters by imposing a common dynamic structure on all elements of the volatility matrix have been proposed as well, including the K-factor GARCH model of Lin (1992), the full-factor GARCH model of Vrontos, Dellaportas, and Politis (2003), the orthogonal GARCH model of Alexander (2001), the generalized orthogonal GARCH model of van der Weide (2002), and the conditionally uncorrelated component (CUC) model of Fan, Wang, and Yao (2008).

PCA of spectral matrices is another approach to estimating the common factors within various factor models. Stoffer (1999) considered modeling multivariate time series that contain common cyclic components. Each series can be represented as a linear combination of independent stochastic cyclic components of different frequencies, in which each may have a different phase.

Forni et al. (2000) proposed a generalized dynamic factor model for high-dimensional multivariate time series. The observed processes are dynamically dependent on linear combinations of current and lagged values of common orthonormal white noise. The model is generalized to allow the idiosyncratic component to be a possibly correlated stationary vector process. These authors proposed an estimator based on dynamic eigenvectors, which are eigenvectors of the spectral density matrix as functions of the frequency, and proved consistency as both the size and the length of the series go to infinity. Eichler, Motta, and von Sachs (2011) further generalized the dynamic factor model for use with second-order nonstationary time series. Assuming local stationarity, estimation is performed using principal component regression to allow deterministic time variation.
in the factor loadings. Their approach uses two-sided filtering, however, which is not amenable to prediction.

Bai (2003) considered inferential theory for factor models as both the size and the length of the series go to infinity. The common factors are dynamic linearly, but the relationship between the observations and the factor is static. The idiosyncratic component is allowed to have contemporaneous and lagged cross-dependence and heteroscedasticity, under the assumption of finite eighth moments. Motta, Hafner, and von Sachs (2011) generalized this theory for static-stationary factor models to the nonstationary case. Ombao, von Sachs, and Guo (2005) also considered analysis of multivariate nonstationary time series. They applied a time-varying eigendecomposition of the smooth localized complex exponential (SLEX) spectral density matrix to conduct feature extraction. The SLEX principal components can then be used for optimal model selection and segmentation. Their approach is computationally efficient and allows for time-varying autocorrelation and cross-correlation, but it is not designed for forecasting.

In this article we propose a new approach to modeling the conditional mean vector and conditional covariance matrix of a stationary multivariate autoregressive and conditionally heteroscedastic time series that ameliorates some of the weaknesses of PCA and ICA. Our approach is an extension of PCA in which we also seek orthogonality in cross-dependence over time, but it is not as stringent as ICA, because we focus on orthogonality only up to the fourth order. Principal components are contemporaneously uncorrelated; however, lagged cross-correlations may be nonzero, conditional correlations may be nonzero, and cross-correlations of nonlinear transformations, such as the square process, may be nonzero. Independent components are orthogonal in all of these respects, but sufficient conditions for independence are much stronger. We seek a middle ground that allows us to focus on the time evolutions of the conditional mean and conditional covariance of a vector time series.

The common components may be autoregressive or conditionally heteroscedastic, and their relationship with the observations is static. We make assumptions about the observations directly, but not the common components, and derive our asymptotic results for a fixed dimension. In applications, we check the validity of a $d$-dimensional time series can be modeled by $d$ dynamic orthogonal component (DOC) series. This is different from an approach that uses common factors or common dynamic factors, which often involves testing for the number of common factors.

In statistical analysis, orthogonal components are often used to find suitable representations of multivariate data. The well-known PCA produces contemporaneously orthogonal components, but those components may have high-order cross-dependence over time. To overcome this weakness, we consider DOCs that have no cross-correlations over time. Specifically, we estimate the components via a generalized decorrelation method that minimizes the linear and quadratic dependence across components and across time. We do not assume the existence of DOCs for all vector time series. In an application, we apply a Ljung–Box type statistic to test the validity of the estimated DOCs. When DOCs exist, univariate analysis can be applied to build a model for each component. Those univariate models are then combined to obtain a multivariate model for the original time series. In this way, DOC is similar to ICA, although in practice testing the existence of independent components is much more difficult, given that robust tests of independence between multiple time series are not applicable without additional strong assumptions.

For a stationary vector time series $x_t$, assume, for simplicity, that there exist DOCs $s_t$ such that $x_t = Ms_t$, where $M$ is a nonsingular mixing matrix. We test the validity of this assumption in applications. For computational simplicity, let $U$ denote an uncorrelating matrix and let $z_t = Ux_t$ denote uncorrelated observations. In practice, $z_t$ may consist of the principal components of $x_t$. The relationship between $z_t$ and $s_t$ is then

$$s_t = M^{-1}x_t = M^{-1}U^{-1}z_t = Wz_t,$$

where $W = M^{-1}U^{-1}$ is referred to as the separating matrix. In this article we estimate $W$ to obtain the DOCs.

We use the dynamic structure intrinsic in time series data to estimate a separating matrix $W$ and identify DOCs. Unlike in ICA, Gaussian components are allowed as long as they are serially dependent. Our estimation method is robust to account for heavy-tailed distributions usually associated with financial data. We use a rotation-based parameterization for $W$ that has several computational advantages. We prove the consistency and asymptotic normality of our estimator under suitable regularity conditions. Finally, we consider the application of our method to economic and financial time series data and contrast our method with traditional methods.

In Section 2 we define DOCs, construct a test for checking the existence of DOCs, introduce our methodology, discuss parameterization and identifiability of the separating matrix, and state conditions for consistency and asymptotic normality of the proposed estimator. Our approach is computationally practical with a simple analytical gradient. In Section 3 we provide extensive data analysis to evaluate the performance of the proposed procedure. Commonly used multivariate diagnostic checking establishes the empirical adequacy of the model built via our proposed procedure. Various out-of-sample evaluations are conducted to contrast the proposed modeling approach with traditional methods. We present concluding remarks in Section 4, and provide technical proofs and additional details for the competing methods in the online Supplemental Material.

\section{Methodology}

Let $y_t = (y_{1t}, \ldots, y_{dt})'$ denote a stationary $d$-dimensional vector process with $E(y_t) = 0$. Let $\mathcal{F}_t$ denote the sigma-field generated by the past information until time index $t$, that is, $\sigma(y_1, y_{t-1}, \ldots)$. We partition the process as

$$y_t = \mu_t + e_t,$$

in which $\mu_t = E(y_t|\mathcal{F}_{t-1})$ is the conditional mean and $e_t$ is the serially uncorrelated disturbance vector with $E(e_t) = 0$. Let $\Sigma_t = \text{Cov}(y_t|\mathcal{F}_{t-1}) = \text{Cov}(e_t|\mathcal{F}_{t-1})$ denote the conditional covariance matrix. Multivariate time series modeling is concerned with the time evolutions of $\mu_t$ and $\Sigma_t$. Finally, let $\Sigma_e = \text{Cov}(e_t)$ and $\Sigma_y = \text{Cov}(y_t)$ be the unconditional covariance matrix of $e_t$ and $y_t$, respectively.
2.1 The Dynamic Orthogonal Components

Multivariate serial correlation and multivariate conditional heteroscedasticity are the predominant forms of dependence observed in most econometric applications. We propose two separate DOC specifications for Equation (2) to address each of these cases, DOCs in mean and DOCs in volatility. These two DOCs may be combined for time series exhibiting both forms of dependence. These specifications are applicable for linear autoregressive and generalized autoregressive dependence and are the focus of our applications. However, our methodology also may be extended to certain forms of nonlinear generalized autoregressive dependence.

We begin with the specification of DOCs in mean for time series with multivariate serial correlation. Here we work with the demeaned observations \( x_t = y_t - \bar{E}(y_t) \), and assume for simplicity that \( \Sigma_t \) is constant. If necessary, we perform a joint estimation of the mean and volatility equations once a model is specified. Here let \( M_1 \) denote the mixing matrix associated with DOCs in mean. Combining Equations (1) and (2), we have \( x_t = M_1 s_t \) and

\[
s_t = M_1^{-1} x_t = M_1^{-1} \mu_t + M_1^{-1} e_t = \tilde{\mu}_t + \tilde{e}_t. \tag{3}
\]

Here the components \( s_t \) are dynamically orthogonal in mean. Specifically, \( \text{Cov}(s_t, s_{t-\ell}) \) is diagonal for all lags \( \ell = 0, 1, 2, \ldots \), and univariate ARMA models can be used to model the remaining marginal serial correlation. In the corresponding vector ARMA model for Equation (3), \( \text{Cov}(\tilde{e}_t) \) is diagonal, and all of the coefficient matrices specified by \( \tilde{\mu}_t \) are also diagonal.

Next we specify DOCs in volatility for time series with multivariate conditional heteroscedasticity. Here we focus on the conditional covariance \( \Sigma_t \) and work with the disturbance process \( x_t = y_t - \bar{E}(y_t|F_{t-1}) = e_t \). For simplicity, we assume that \( \mu_t \) is known. In practice, it may be estimated using DOCs in mean, or by any other appropriate time series model. Let \( M_2 \) denote the mixing matrix associated with DOCs in volatility. Combining Equations (1) and (2), we have \( x_t = M_2 s_t \), \( s_t = M_2^{-1} x_t \), and

\[
\Sigma_t = \text{Cov}(x_t|F_{t-1}) = M_2 \text{Cov}(s_t|F_{t-1}) M_2'. \tag{4}
\]

Here the components \( s_t \) are dynamically orthogonal in volatility. Specifically, \( \text{Cov}(s_t', s_{t-\ell}') \) is diagonal and \( \text{Cov}(s_{t+hs_{t+hs}}|s_{t+\ell}) \) is 0 for \( \ell \neq j \), for all lags \( \ell = 0, 1, 2, \ldots \). For linear generalized autoregressive conditional heteroscedastic time series, this implies that the conditional covariance process \( \text{Cov}(s_t|F_{t-1}) \) is diagonal at all time points, and univariate models, such as GARCH or stochastic volatility models, can be used to model the remaining marginal conditional heteroscedasticity. All coefficient matrices for the corresponding mixing models of \( \text{Cov}(s_t|F_{t-1}) \) will be diagonal.

For any DOC specification, we require the \( x_t \) process to obey some standard regularity conditions:

**Assumption 1.** The process \( x_t \) is stationary and ergodic with \( \mathbb{E}|x_t|^2 < \infty \).

Let \( s_t = (s_{t1}, \ldots, s_{tp})' \) denote a vector time series of DOCs. Without loss of generality, \( s_t \) is assumed to be standardized such that \( \mathbb{E}[s_t] = 0 \) and \( \text{Var}(s_t) = 1 \), for \( i = 1, \ldots, d \). The fundamental motivation is that empirically, the dynamics of \( x_t \) often can be well approximated by an invertible linear combination of DOCs, \( x_t = M s_t \). Later, we check the adequacy of the approximation in applications.

2.2 Testing the Existence of Dynamic Orthogonal Components

To test for the existence of dynamic orthogonal components, we propose a Ljung–Box type statistic (Ljung and Box 1978). For DOCs in mean, we test for significant lagged cross-correlations in \( s_t \), and for DOCs in volatility, we test for significant lagged cross-correlations in \( s_t^2 \). Let \( h \) denote either the identity transformation (for DOCs in mean) or the square transformation (for DOCs in volatility), and let \( h_{t-\ell} = h(s_{t-\ell}) \) and \( \rho_{h}^{(\ell)} = \text{Corr}(h_{t-j}, h_{t-j-\ell}). \) The joint lag-\( m \) null and alternative hypotheses to test for the existence of DOCs are

\[
H_0: \rho_{h}^{(\ell)} = 0 \quad \text{for all } i \neq j, \ell = 0, \ldots, m; \tag{5}
\]

\[
H_A: \rho_{h}^{(\ell)} \neq 0 \quad \text{for some } i \neq j, \ell = 0, \ldots, m.
\]

We define our test statistic as

\[
Q_d^h(m) = n \sum_{i<j} \rho_{h}^{(\ell)}(0)^2 + n(n + 2) \sum_{k=1}^m \sum_{j\neq j} \rho_{h}^{(\ell)}(k)^2 / (n - k).
\]

Under \( H_0 \), \( Q_d^h(m) \) is asymptotically distributed as a \( \chi^2 \) distribution with \( d(d-1)/2 + md(d-1) \) degrees of freedom. This result can be obtained using the techniques of Li (2004). The null hypothesis is rejected for a large value of \( Q_d^h(m) \). When \( H_0 \) is rejected, an alternative modeling procedure must be sought.

For serially correlated processes, this test justifies specification of vector ARMA type models for DOCs, with diagonal coefficient matrices. Inclusion of lag 0 ensures that the components are contemporaneously uncorrelated as well. For conditionally heteroscedastic process, the proposed test justifies univariate specification for the conditional variance of the DOCs. To see that this also implies that the product processes \( s_t s_t' \) are serially uncorrelated, we note that under stationarity, the Cauchy–Schwarz inequality gives

\[
|\text{Cov}(s_t s_t')| \le \text{Var}(s_t) s_t'^2.
\]

Therefore, this test justifies specification of diagonal conditional covariance models with diagonal coefficient matrices, because we always include lag 0 in our hypothesis test.

2.3 Estimation of Dynamic Orthogonal Components

For theoretical and practical considerations, it is convenient to work with uncorrelated random variables. In Equation (1), it is useful to start with \( z_t \) instead of \( x_t \). Without loss of generality, we assume that \( \text{Cov}(z_t) = I_d \), the \( d \times d \) identity matrix. Let \( Y \) be the matrix of eigenvectors and \( A \) be the diagonal matrix of the corresponding eigenvalues of \( \Sigma_x \), the unconditional covariance matrix of \( x_t \). We then have \( U = A^{-1/2} Y' \).

Given the uncorrelated process \( z_t \), Equation (1) implies that the separating matrix \( W \) is necessarily orthogonal, because

\[
\mathbb{I} = \text{Var}(s_t) = W \text{Var}(z_t) W' = WW'. \]

Thus \( W \) has \( p = (d-1)/2 \) free elements, instead of \( d^2 \). The orthogonality of \( W \) means that it represents a rotation in the \( d \)-dimensional space and can be parameterized by the length \( p \) vector \( \theta \) of rotation angles as \( W_\theta \), defined later.
Parameterization. For $d \geq 2$, let $O(d)$ denote the group of all $d \times d$ orthogonal matrices, and let $SO(d)$ denote the subgroup (rotation group) with determinant equal to $+1$. [ Relevant properties of $SO(d)$ are discussed in the online Supplementary Materials.] Let $\xi_1, \ldots, \xi_d$ denote the canonical basis of $\mathbb{R}^d$. Let $Q_j(\psi)$ denote a rotation of all vectors lying in the $(\xi_j, \xi_j)$ plane of $\mathbb{R}^d$ by an angle $\psi$, oriented such that the rotation from $\xi_j$ to $\xi_j$ is assumed to be positive. Specifically, for $i \neq j$, $Q_j(\psi)$ is a Givens (plane) rotation matrix, that is, the identity matrix $I_d$ with the $(i, i)$ and $(j, j)$ elements replaced by $\cos(\psi)$, the $(i, j)$ element replaced by $-\sin(\psi)$, and the $(j, i)$ element replaced by $\sin(\psi)$.

Let $\theta$ denote a vector of rotation angles with length $p = (d-1)/2$, indexed by $i, j: 1 \leq i < j \leq d$, and note that any rotation $W \in SO(d)$ can be written in the form

$$ W = Q^{(1)} \cdots Q^{(2)}, $$

in which

$$ Q^{(j)} = Q_{1, j}(\theta_j) \cdots Q_{k-1, j}(\theta_{k-1,j}). $$

Furthermore, there exists a unique inverse mapping of $W \in SO(d)$ into $\theta \in (-\pi, \pi)^p$ such that the mapping is ensured to be continuous if either all elements on the main diagonal of $W$ are positive or all elements of $W$ are nonzero (see Matteson (2008)).

Estimation Equation. Given an uncorrelated mean-0 vector time series $z_t, t = 1, \ldots, n$, we begin by estimating the parameter vector of rotation angles $\theta$ defined earlier. Define $s_t(\theta) = W_{\theta} z_t$, and let $\hat{E} \{ \}$ denote the sample expectation operator.

For a suitably chosen vector function $h: \mathbb{R}^d \to \mathbb{R}^d$, such that $h_t(s) = h_t(s_t)$ for $i = 1, \ldots, d$, our objective in estimating $\theta$ is to make the lagged sample cross-covariance function of the transformed process

$$ \hat{\gamma}^{h(s(\theta))}(\ell) = \hat{E} \{ h_t(s_t) h_{t-\ell}(s_{t-\ell}) \}' \quad \text{as close to diagonal as possible for some prespecified finite set of lags $\ell \in \mathbb{N}_0 \subset \mathbb{N}$, in which the subscript 0 emphasizes the inclusion of 0.} $$

To estimate DOCs in mean, let $h_t(s) = s_t$, the identity function of each component. To estimate DOCs in volatility, let $h_t(s) = s_t^2$, or if the observations $z_t$ exhibit heavy tails, use a continuously differentiable version of Huber’s function, defined as

$$ \text{Huber}_c(s) = \begin{cases} s^2 & \text{if } |s| \leq c, \\ 2|s|c - c^2 & \text{if } |s| > c \end{cases} $$

for some $0 < c < \infty$. Huber’s function will considerably relax the moment conditions required for consistent estimation of DOCs in volatility. Several other functions with similar properties, such as those associated with M-estimators, also may be considered. Continuously differentiable functions are preferred for both theoretical and computational reasons.

Other choices of $h$ will give different forms of DOCs. For example, if a nonlinear autoregressive structure is exhibited by the observations $z_t$, then the DOC approach can be generalized by considering $\hat{E} \{ s_t(\theta) h(s_{t-\ell}(\theta)) \} - \hat{E} \{ s_t(\theta) \} \hat{E} \{ h(s_{t-\ell}(\theta)) \}'$ in lieu of Equation (7). We have taken $h$ as given in our analysis; however, such transformations can be estimated directly from the data.

We formalize the objective to conform with the generalized method-of-moments (GMM) literature. Let $f(z_t, \theta)$ denote a vectorized array of orthogonality constraints such that

$$ f_{ij}^*(z_t, \theta) = h_t(s_t(\theta)) h_{t-\ell}(s_{t-\ell}(\theta)) - \hat{E} \{ h_t(s_t(\theta)) \} \hat{E} \{ h(s_{t-\ell}(\theta)) \} \} \}. $$

Note that $f$ is indexed by $i < j$ for $\ell = 0$ and by $i \neq j$ for $\ell > 0$, reflecting the symmetry of the lagged cross-covariance function at lag 0. Let $|\mathbb{N}_0|$ denote the cardinality of the set $\mathbb{N}_0$; then the length of $f$ will be $q = p(2|\mathbb{N}_0| - 1)$.

Let $l_{ij}$ denote the indicator function. Lagged cross-dependence is typically strongest at low lags, so as we orthogonalize, we apply weights

$$ \phi_{\ell} = \frac{1 - \ell / |\mathbb{N}_0|}{\sum_{\ell}(1 - \ell / |\mathbb{N}_0|)} \quad \text{for $\ell \in \mathbb{N}_0$} $$

for successive lags accordingly. The weights are arranged into a diagonal weighting matrix as $\Phi = \text{diag} \{ \phi_{1,1}, \ldots, \phi_{1,1}, \ldots, \phi_{2,2}, \ldots, \phi_{p,0}, \ldots, \phi_{p,p} \}$. Let $\tilde{\theta}_t(\theta) = \hat{E} \{ f(z_t, \theta) \}$. We define our objective function as

$$ J_t(\theta) = \tilde{\theta}_t(\theta) \cdot \Phi \tilde{\theta}_t(\theta) $$

and our estimate of $\theta$ as $\hat{\theta}_t = \text{argmin}_\theta J_t(\theta)$. Given our estimate of $\theta$, we estimate the separating matrix as $W_{\hat{\theta}}$. Conditional on $\theta$, we estimate the vector time series of DOCs as $\hat{s}_t = W_{\hat{\theta}} z_t$. Equation (8) measures the predominant forms of dependence observed in most common econometric applications, multivariate lagged cross-correlation and conditional heteroscedasticity, for $h$ equal to the identity and square transformation, respectively.

2.4 Identifiability

Principal components have a natural ordering and can be standardized, but their signs are ambiguous. For estimation of DOCs, three ambiguities are associated with estimating $W$ and $s$ in general: the magnitude, sign, and order of the DOCs. The magnitude of the DOCs has been fixed by the assumption that $\text{Var}(s) = I$. Let $P_{\pm}$ denote a signed permutation matrix, and note that the mixing model $x = Ms$ is equivalent to

$$ x = MP_{\pm} s = (MP_{\pm} \cdot P_{\pm} s), $$

in which $P_{\pm}$ are new DOCs and $MP_{\pm}$ is the new mixing matrix. Identification of DOCs up to a signed permutation is sufficient for modeling and forecasting purposes if the univariate time series models specified for the individual DOCs are symmetric in their autoregressive nature. This is true for ARMA and GARCH models when considering DOCs in mean and volatility, respectively. When this is the case, we may construct an equivalence class and a canonical form for $W$ to conduct inference. We detail our approach to this in the online Supplementary Materials.

2.5 Asymptotic Properties of the Proposed Estimator

Asymptotic results for our estimator require basic assumptions about the transformed observations and the parameter space. Without loss of generality, assume throughout this section that $E[z_t] = 0$ and $\text{Cov}(z_t) = I_d$. 
Strong Consistency. Let

$$\mathcal{J}(\theta) = E[f(x_1, \theta)]' \Phi E[f(x_1, \theta)],$$

and let $\Theta$ denote a sufficiently large compact subset of the space $\Theta$, defined in the online Supplementary Materials.

Assumption 2.

$$\sup_{\theta \in \Theta} E \| \mathbf{h}(\mathbf{W}_n \mathbf{x}_1) \|^2 < \infty$$

and

$$\sup_{\theta \in \Theta} E \left\| \frac{\partial \mathbf{h}(\mathbf{W}_n \mathbf{x}_1)}{\partial \theta} \right\|^2 < \infty.$$  

Theorem 1. Suppose that $\mathbb{N}_0 \subset \mathbb{N}_0$ is fixed and finite, $\Phi$ is constant and positive definite, and $\mathbf{h}$ is measurable and continuously differentiable. Suppose that Assumption 1 holds for $\mathbf{x}_1$, and that $\Phi$ is chosen such that Assumption 2 is satisfied. Suppose that there exists a unique minimizer $\theta^0 \in \Theta$ of Equation (9), and that $\mathbf{W}_n\theta^0$ satisfies the conditions for a unique continuous inverse to exist. Then $\hat{\theta}_n \rightarrow^p \theta^0$ as $n \rightarrow \infty$.

Convergence is established on equivalence classes. Proofs of this and the subsequent theorem are in the online Supplementary Materials. Convergence in probability follows under weaker assumptions. Convergence also may be established for certain forms of nonstationarity. Note that when DOCs exist, then there exists a $\theta^0 \in \Theta$ such that $E[f(x_1, \theta^0)] = 0$.

Asymptotic Normality. Note that $\hat{f}_n(\theta)$ is continuously differentiable with respect to $\theta$ on $\Theta$ by assumption. We denote its $q \times p$ matrix gradient by $\nabla \hat{f}_n(\theta) = \frac{\partial \hat{f}_n(\theta)}{\partial \theta}$. Let $\nabla_n = n \text{Var}(\hat{f}_n(\theta^0))$, a nonrandom, positive definite $q \times q$ matrix. Because both $\nabla_n$ and $\theta^0$ are unknown, we require a weakly consistent estimator of the former to use the following theorem for statistical inference.

Theorem 2. If the conditions of Theorem 1 hold, if $\mathbf{h}(\mathbf{W}_n \mathbf{x}_1)$ is strong mixing with geometric rate, if $E[\mathbf{h}(\mathbf{W}_n \mathbf{x}_1)]$ has linearly independent columns, and if there exists a $\hat{\mathbf{V}}_n$ such that $\hat{\mathbf{V}}_n - \mathbf{V}_n \xrightarrow{p} \mathbf{0}$ as $n \rightarrow \infty$, then, as $n \rightarrow \infty$,

$$\left(\hat{\mathbf{V}}_n(\theta^0)' \Phi \hat{\mathbf{V}}_n(\theta^0) - \mathbf{V}_n(\theta^0)' \Phi \mathbf{V}_n(\theta^0)\right) \xrightarrow{d} \mathcal{N}(0, \mathbf{I}).$$

Note that $\mathbf{x}_1$ is strong mixing with geometric rate if there exists constants $K > 0$ and $a \in (0, 1)$ such that

$$\sup_{A \in \mathbb{S}(x_1,t \leq 0), B \in \mathbb{S}(x_1,t > k)} |P(A \cap B) - P(A)P(B)| = \phi_k \leq Ka^k,$n

in which $\phi_k$ is referred to as the mixing rate function (see Taniguchi and Kakizawa 2000).

In practice, $\hat{\mathbf{V}}_n$ is estimated consistently using $\hat{f}(x_1, \hat{\theta}_n)$. Mátějáš (1999) provided a survey of covariance matrix estimators arising from GMM estimation. The nonparametric heteroscedasticity- and autocorrelation-consistent estimator of Andrews (1991) or the Newey and West (1987) estimator are two of the more successful methods in use. This general approach is related to the method of Hansen (1982) in the GMM literature, as well as others in the generalized estimating equations literature.

Our suggestion for the weighting matrix $\Phi$ is intuitive and simple and has no effect on the consistency of the estimator. The choice of $\Phi$ does affect the variance of the asymptotic distribution of $\hat{\theta}_n$. Weighting by $\nabla_n^{-1}$ would minimize the asymptotic covariance matrix. This suggests that $\Phi = \nabla_n^{-1}$ might be preferred, given an asymptotically efficient estimate of $\theta$ within this class of estimators, which has asymptotic covariance matrix equal to $\mathbb{E}(\hat{f}_n(\theta^0)' \nabla_n^{-1} \nabla_n^{-1} \hat{f}_n(\theta^0))^{-1}$. However, implementing such a choice for $\Phi$ is not straightforward, given that consistent estimation of $\nabla_n$ requires a consistent estimator of $\theta^0$. Hansen, Heaton, and Yaron (1996) suggested two approaches for estimation. The first of these is to iterate estimation of the parameter vector and the weighting matrix until both converge. This has been applied extensively; however, care must be taken to ensure convergence. Alternatively, the objective function in Equation (8) may be altered to reflect the optimal weighting matrix’s dependence on the parameter vector. Referred to as the continuously updating estimator, its advantages are increased efficiency and estimators that are invariant to the scaling of the moment conditions (see Ni 1997). However, the added complexity will quickly become infeasible as the dimension increases.

2.6 Practical Implementation

In this section we discuss several issues regarding practical implementation of the proposed DOC methodology, including dimension reduction, numeric optimization, and general principles of choosing parameter settings. We present a simulation study to demonstrate the effects of these parameter choices.

The proposed DOC achieves simplification in multivariate modeling similar to that of independent component models, which differs from that of PCA. In practice, when the dimension is very high, PCA may be used as a screening device to reduce certain dimensions before applying DOC. The effectiveness of such a procedure warrants further investigation, however.

Implementing numeric optimization of Equation (8) is straightforward. We provide an analytic gradient in the online Supplementary Materials, and will make estimation algorithms written using the statistical software R (R Development Core Team 2010) publicly available. Equation (8) has $d(d - 1)/2$ unconstrained parameters and generally is only locally convex. In practice, multiple starting points can be considered because local minima may exist. Computational costs grow with the dimension, and our asymptotic results assume a fixed dimension; thus, the DOC method is not yet suited for extremely high-dimensional data. However, PCA offers a possible means of reducing the dimension of the time series before DOC estimation, as discussed earlier.

When testing for the existence of DOCs, the user must specify the number of lags $m$ to include in Equation (5). The same principals used for determining the number of lags to include in a Ljung–Box test remain applicable. Using as many lags as possible is preferable, but the power of the Ljung–Box test decreases as the number of lags increases. The maximum number of lags may be chosen relative to the length of the series; some authors have suggested $m = \log(n)$. Another consideration is that time series data often have a natural seasonal pattern, and as a practical rule of thumb, $m$ may be set to twice this value. In our implementation, we use the maximum of these two choices.

Specification of Equation (8) requires that the user specify the set $\mathbb{N}_0$ of lags to include in estimating DOCs. Lag 0 is al-
ways included in the set $\mathbb{N}_0$. For DOCs in mean, this is redundant when working with uncorrelated components $z_t$; however, it is essential when estimating DOCs in volatility to minimize the conditional correlations.

For DOCs in mean, we must include at least one lag in which there is serial correlation. If the data are seasonal, considering only one lag might not be sufficient. To automatically choose an appropriate number of lags, we recommend fitting a vector autoregressive model to the data using the Akaike information criterion (AIC) to choose the order. Then the lag order chosen by the AIC may be used as the maximum lag included in the model.

In all of the simulations that follow, the AR coefficients are randomly selected as Uniform(-0.5, 0.5) random variables, and the residual variance is 1, and the AR coefficients are again selected randomly. The first AR coefficient is selected as Uniform(-0.5, 0.5), and the second coefficient is selected as Uniform(-0.5, -0.1). These choices result in roughly half of the implied second-order difference equations having complex roots. Thus, approximately half of the series will have stochastic cycles with oscillating autocorrelations.

All simulations were replicated 1000 times. DOCs in mean were estimated using $h$ equal to the identity function and choosing $\mathbb{N}_0$ using the AIC-based procedure outlined earlier. The percentage of time the AIC correctly selected the order is reported in Table 1. The simulated data $x_t$ are decorrelated by $\hat{U}$ estimated using PCA, the minimum to Equation (8) is found, and then the mixing matrix is estimated as $\hat{M} = \hat{U}^{-1} \hat{W} \theta_n$. To account for the identifiability issues discussed in Section 2.4, we use the Amari metric (Bach and Jordan 2003) to measure the estimation accuracy, defined as

$$A(M_0, \hat{M}) = \frac{1}{2d} \sum_{i=1}^{d} \left( \sum_{j=1}^{d} \frac{|\tilde{m}_{ij}|}{\max_j |\tilde{m}_{ij}|} - 1 \right)$$

in which $\tilde{m}_{ij} = (M_0 \hat{M}^{-1})_{ij}$. The simulation results are summarized in Table 2, and the mean and standard deviation of the Amari error reported for each of the different cases. In general, the Amari error increases as the dimension $d$ increases. Simply performing PCA estimates the mixing matrix as $\hat{U}^{-1}$. This results in large error for every instance and does not decrease with increasing sample size. Estimation using our proposed method results in a much smaller error, and larger sample sizes further reduce the error in all cases.

| Table 1. Percentages AIC selects order of DOCs in mean correctly from simulation study in Section 2.6 |
|---|---|---|---|---|---|---|
|   | $d = 5$ |   | $d = 10$ |   |
|   | $n = 100$ | $n = 500$ | $n = 1000$ | $n = 100$ | $n = 500$ | $n = 1000$ |
| AR(1) | 96.8 | 99.6 | 99.6 | 80.5 | 100 | 100 |
| AR(2) | 88.4 | 99.2 | 99.7 | 31.3 | 100 | 100 |

**Note:** Here $d$ and $n$ denote the dimension and sample size, respectively.

| Table 2. Amari error: mean (SD) from DOCs in mean for simulation study in Section 2.6 |
|---|---|---|---|---|---|---|
|   | $d = 5$ |   | $d = 10$ |   |
|   | $n = 100$ | $n = 500$ | $n = 1000$ | $n = 100$ | $n = 500$ | $n = 1000$ |
| AR(1) | PCA | 1.76 (0.24) | 1.83 (0.24) | 1.86 (0.25) | 3.62 (0.27) | 3.82 (0.31) | 3.88 (0.32) |
|         | DOC | 1.34 (0.30) | 0.87 (0.31) | 0.71 (0.29) | 2.93 (0.32) | 2.10 (0.36) | 1.69 (0.35) |
| AR(2) | PCA | 1.86 (0.24) | 2.03 (0.22) | 2.07 (0.22) | 3.79 (0.28) | 4.23 (0.32) | 4.36 (0.33) |
|         | DOC | 1.24 (0.33) | 0.77 (0.30) | 0.60 (0.28) | 2.85 (0.32) | 1.87 (0.37) | 1.42 (0.33) |

**Note:** Here $d$ and $n$ denote the dimension and sample size, respectively.
to the Huber2.25 function, and the minimum to Equation (8) is found. The mixing matrix is estimated as before. The results of PCA and DOCs in variance are compared in Table 3, again using the Amari error. The Amari error increases with the dimension, and PCA again results in large error in every instance. Estimating the mixing matrix using our proposed method again results in a much smaller error, and larger sample sizes further reduce the error in all cases. The results for DOCs in volatility remain stable for each of the chosen sets \( \mathcal{F}_0 \).

### 3. APPLICATION

In this section we apply our proposed DOC approach to two multivariate time series. We first identify DOCs in mean for the quarterly change series of seasonally adjusted gross domestic product (GDP) of several countries. We compare the in-sample fit and out-sample forecast performance of our approach with several related models. We then identify DOCs in volatility for a three-dimensional series of stock returns. Although the volatility process of asset returns is latent, we provide some in-sample diagnostic checking of the proposed model. We compare our approach with several competing methods.

#### 3.1 Vector Autoregression via Dynamic Orthogonal Components

We identify DOCs in mean for the quarterly change in seasonally adjusted GDP, in percentage, for five countries: the United States (US), Canada (CA), United Kingdom (UK), South Korea (KO), and Taiwan (TW). After identifying DOCs in mean, we model the component process with univariate autoregressive models. The forecasts from these marginal models are combined to obtain forecasts for the original time series. We compare the in-sample fit and out-sample forecast performance of our approach with standard vector autoregressive (VAR) models as well as component models based on PCA and ICA.

The data were obtained from the Organization for Economic Cooperation and Development Web site. We use data from the first quarter of 1981 through the second quarter of 2009. The time series are plotted in Figure 1, with Asian countries in grey and the others in black. The non-Asian countries appear to have very similar evolutions. The Asian countries also exhibit considerable comovement, but this is obscured by their greater variability.

Let \( \mathbf{x}_t \) denote the demeaned multivariate process. We partition the data into training and forecasting subsamples in the second quarter of 2004. As a first step, we examine the original data for serial correlation. Quarterly economic data commonly exhibit seasonal behavior with period four; to ensure that we measure both regular and seasonal correlations, we consider two full seasonal periods. The lag eight multivariate Ljung–Box statistics for \( \mathbf{x}_t \) is \( Q_8(\mathbf{x}_t, 8) = 315.77 \), which is highly significant compared with the chi-squared distribution with 200 degrees of freedom. Indeed, the \( p \)-value of the statistic is \( 3.27 \times 10^{-7} \); thus there are serial and cross-correlations in the data. A visual inspection of autocorrelation and lagged cross-correlation plots shows geometric decay within the first four lags.

We now consider modeling the process as a linear combination of DOCs as \( \mathbf{x}_t = \mathbf{M} s_t \). We first check whether the components of \( \mathbf{x}_t \) are already DOCs in mean. The DOC test statistic is \( Q_8^M(s_t, 8) = 528.24 \), indicating that \( \mathbf{x}_t \) is not DOC in mean, and that a diagonal vector model will be insufficient. We next estimate a DOC in mean based VAR model in two stages. First, we estimate DOCs in mean via the proposed procedure of Section 2. For DOCs in mean, we use \( h \) equal to the component identity transformation. Second, following the procedure outlined in Section 2.6 shows that a VAR(1) model for \( \mathbf{x}_t \) has the lowest AIC, so we include lags \( \mathcal{F}_0 = \{0, 1\} \) in our objective function and estimate the mixing matrix \( \mathbf{M} \) and DOCs in mean \( s_t \). The DOC in mean test statistics for the estimated DOCs is \( Q_8^M(s_t, 8) = 178.24 \) with \( p = 0.86 \), indicating that DOCs exist

### Table 3. Amari error, in mean (SD) from DOCs in volatility for simulation study in Section 2.6

<table>
<thead>
<tr>
<th></th>
<th>( d = 5 )</th>
<th>( n = 1000 )</th>
<th>( n = 2000 )</th>
<th>( d = 10 )</th>
<th>( n = 1000 )</th>
<th>( n = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>1.74 (0.30)</td>
<td>1.83 (0.28)</td>
<td>3.64 (0.39)</td>
<td>3.78 (0.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOC (Huber2.25, ( \mathcal{F}_0 = {0, 1} ))</td>
<td>0.51 (0.11)</td>
<td>0.39 (0.08)</td>
<td>1.44 (0.24)</td>
<td>1.03 (0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOC (Huber2.25, ( \mathcal{F}_0 = {0, 1, 2} ))</td>
<td>0.55 (0.12)</td>
<td>0.42 (0.09)</td>
<td>1.63 (0.25)</td>
<td>1.15 (0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOC (Huber2.25, ( \mathcal{F}_0 = {0, 1, 2, 3} ))</td>
<td>0.59 (0.13)</td>
<td>0.44 (0.10)</td>
<td>1.77 (0.26)</td>
<td>1.25 (0.21)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Here \( d \) and \( n \) denote the dimension and sample size, respectively.

Figure 1. The quarterly change in seasonally adjusted GDP in percentage for the United States (US), Canada (CA), United Kingdom (UK), Korea (KO), Taiwan (TW) from 1981Q1 through 2009QII.

**NOTE:** The vertical line at 2004QII marks the beginning of recursive evaluation of out-of-sample one-step-ahead forecasts.
and that a diagonal model will be sufficient for the estimated DOCs.

We now fit a diagonal VAR(1) model to the estimated DOCs \( \hat{s}_t \). Let

\[
    s_t = \beta s_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} (0, \Sigma_\varepsilon),
\]

in which both \( \beta \) and \( \Sigma_\varepsilon \) are diagonal matrices. Because the class of VAR models is closed under linear transformations, the implied model for the observations \( x_t \) is given by

\[
    x_t = M \hat{s}_t = M (\beta s_{t-1} + \varepsilon_t) = M \hat{\beta} M^{-1} s_{t-1} + M \varepsilon_t.
\]

The implied VAR(1) model for the original data is \( x_t = \beta x_{t-1} + \varepsilon_t, \varepsilon_t \overset{iid}{\sim} (0, \Sigma_\varepsilon) \), in which \( \beta = M \hat{\beta} M^{-1}, \varepsilon_t = M \varepsilon_t, \) and \( \Sigma_\varepsilon = M \Sigma_\varepsilon M^T \). Because the mixing matrix must be estimated this model still has \( d^2 + (p)d + d = 35 \) total parameters. The reduction in parameters becomes even more significant for higher-order VAR models.

The estimated coefficient matrix \( \hat{\beta} \) implied by the DOCVAR(1) model is given in Table 4(b). These implied estimates are similar to those of an unconstrained VAR(1) model. However, in this particular instance, the DOC factorization, as expected, gives a more parsimonious representation. The multivariate Ljung–Box statistics for the residuals give \( Q_5(\hat{\varepsilon}_t, 8) = 223.37 \) with \( p = 0.12 \), indicating no significant serial and cross correlations in the residuals at the \( 5\% \) significance level. Details of competing approaches, including standard vector autoregressive (VAR) models and component models based on PCA and ICA, are given in the online Supplementary Materials. Some comments on their relative performance are summarized below.

The loadings on pairs of factors for each of the five countries are given in Figure 2 for selected components. Loadings on the first two principal components closely cluster the developed countries, whose GDPs are less variable. The two developing countries are largely separated by their loadings on the second component. The next plot depicts the loading from the FastICA method (Hyvärinen, Karhunen, and Oja 2001), restricted to only the first two principal components as inputs. This orthogonal rotation, with signed permutation, of the principal components would not give a substantially different fit.

The unrestricted FastICA method may estimate components sequentially via a deflation algorithm. The first two components selected do not show loadings related to the countries in an obvious way. Finally, two components selected from our generalized decorrelation method cluster the Asian countries while separating the developed countries by their loadings on the second component.

All models were fit using observations up to the second quarter of 2004. The AIC is computed using only data from the training period and is given in Table 5. The proposed DOC in mean approach has the lowest AIC value. One-step-ahead forecasts of \( \mu_t \) are computed and forecast errors are calculated. The models are then refit using an additional data point, and forecasts are recursively calculated. The following metrics were used to compare the out-of-sample forecast performance of the models: mean squared forecast error (MSFE), mean absolute forecast error (MAFE), generalized variance of forecast errors det(\( \hat{\Sigma}_e \)), and the total forecast error variance trace(\( \hat{\Sigma}_e \)). The out-of-sample performance is given in Table 5. Our proposed DOC in mean approach performs best in each out-of-sample criterion. The FastICA-VAR(1) model has a comparable MSFE, but it fares poorly based on the det(\( \hat{\Sigma}_e \)) criterion. The unconstrained VAR(1) model has the most parameters and fares relatively well except on the trace(\( \hat{\Sigma}_e \)) criterion. The diagonal VAR(1) was nearly uniformly the worst, followed by PC-VAR(1).

Recursive estimation was done to mitigate the possible influence of a particular forecast origin on the performance measure. In this particular instance, estimates of the DOCs in mean separating matrix were stable over the forecasting period. The positive results, both in-sample and out-of-sample, indicate that the GDP series may be modeled parsimoniously by the proposed DOCs in mean.

### 3.2 Multivariate Volatility via Dynamic Orthogonal Components in Volatility

We identify DOCs in volatility for a multivariate heteroscedastic time series of stock returns. We consider the daily log returns, in percentage, of the S&P 500 Index and Cisco Systems and Intel Corporation stocks. After identifying DOCs in volatility, we chose to model the component process with univariate GARCH models. The estimated component volatility
processes are combined to estimate the conditional covariance matrix of the original series $\Sigma_t$. Although the volatility process of asset returns is latent, we provide some in-sample diagnostic checking of the proposed model. We compare our approach with the orthogonal GARCH (O-GARCH), dynamic conditional correlation (DCC), and CUC models.

The data span January 2, 1991 through December 31, 1999, with $n = 2275$ observations. This dataset has been analyzed previously by Tsay (2005) and Fan, Wang, and Yao (2008). The return process is shown in Figure 3. The presence of volatility clustering is clear in each series. The volatilities of the series also generally move together; note the large increase in the fourth quarter of 1997 in each series. However, there are instances where the change in volatility is restricted to a single asset; note the large increase for Cisco stock in the second quarter of 1994 that is not present in the other series.

The sample correlations are 0.52, for S&P 500 and Intel, 0.50 for S&P 500 and Cisco, and 0.47 for Cisco and Intel. We use a restricted VAR(5) model for $\mu(t)$ that has 11 nonzero parameters and focus on volatility modeling of $x_t$. We examine serial correlations of $e_t$ and its squared series. The multivariate Ljung–Box statistics and $p$-values for the residuals $\hat{e}_t$ and for the squared residuals $\hat{e}_t^2$ are computed in Table 6(a). The VAR(5) model has removed the serial correlation, but significant serial correlation remains in the squared residuals, indicating conditional heteroscedasticity. Each residual series also exhibits significant excess kurtosis.

We now consider modeling the process as a linear combination of DOCs as $x_t = M_s t$. We first check whether the components of $x_t$ are already DOCs in volatility. The DOC test statistic is $Q_0^0(\hat{e}_t^2, 10) = 156.81$, which is very large relative to a $\chi^2$ with 63 degrees of freedom, indicating that $x_t$ is not DOC in volat-

| Table 5. Out-of-sample performance with respect to MSFE, MAFE, generalized variance of forecast errors $[\text{det}(\hat{\Sigma}_e)]$, and the forecast error variances $[\text{trace}(\hat{\Sigma}_e)]$ for the various fitted VAR(1) models in Section 3.1 |
|---------------------------------|----------------|----------------|----------------|----------------|
| Model                          | MSFE   | MAFE   | det($\hat{\Sigma}_e$) | trace($\hat{\Sigma}_e$) |
| Diagonal VAR(1)                | 1.978  | 0.803  | 0.027             | 8.926            |
| VAR(1)                         | 1.804  | 0.760  | 0.017             | 8.656            |
| DOC-VAR(1)                     | 1.730  | 0.748  | 0.014             | 8.073            |
| FastICA-VAR(1)                 | 1.774  | 0.807  | 0.042             | 8.253            |
| PC-VAR(1)                      | 1.956  | 0.808  | 0.026             | 8.660            |

NOTE: The AIC is computed using only data from the training period.
ity, and that a diagonal volatility model will be insufficient. We next estimate the orthogonal matrix of eigenvectors $\mathbf{Y}$ and the diagonal matrix of corresponding eigenvalues $\mathbf{\Lambda}$ from the estimated residual covariance matrix. We define an uncorrelated residual vector time series by

$$\hat{\mathbf{z}}_t = \mathbf{\Lambda}_t^{-1/2} \mathbf{\hat{e}}_t.$$  

We confirm that decorrelation is insufficient for estimating DOCs in volatility. The DOC test statistic is $Q^{(3)}(\hat{s}_t^2, 10) = 235.36$, indicating that a diagonal volatility model will be insufficient.

We next estimate a DOC in volatility-based GARCH model in two stages. We begin by estimating the mixing matrix $\mathbf{M}$ and DOCs in volatility $s_t$. Following the procedure outlined in Section 3.2, for DOCs in volatility, we use $\hat{h}$ equal to the component Huber$^{-2.5}$ transformation and include $\mathbb{N}_0 = \{0, 1\}$ in our objective function. The DOC in volatility test statistic is $Q^{(3)}(\hat{s}_t^2, 10) = 41.30$, indicating DOCs exist and that a diagonal volatility model will be sufficient. Now that DOCs in volatility are identified, we consider a univariate volatility model for each process. Depending on the application, several choices are available, including models from the GARCH family, stochastic volatility models, and nonparametric and semiparametric models.

We consider a version of the GARCH model (Bollerslev 1986), referred to as GARCH(1, 1)-$t$, for each DOCs in volatility $s_t$, $i = 1, \ldots, d$. Let $t_{(0, 1)}$ denote the standardized Student $t$ distribution with $\nu$ degrees of freedom. The implied multivariate volatility model for the original time series of innovations is

$$\mathbf{e}_t = \mathbf{M}_t \mathbf{S}_t = \mathbf{MV}_t^{1/2} \mathbf{e}_t,$$

$$\mathbf{V}_t = \text{diag}(\sigma^2_{1t}, \ldots, \sigma^2_{dt}), \quad \mathbf{e}_t \overset{iid}{\sim} t_{\nu}(0, 1),$$

$$\mathbf{\Sigma}_t = \mathbf{MV}_t \mathbf{M}'_t, \quad \sigma^2_{it} = \omega_i + \alpha_i \sigma^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}.$$  

We assume that $\omega_i > 0$, and $\alpha_i > 0, \beta_i < 1$ to ensure positiveness. We further assume $\nu_i > 2$ and $\omega_i + \alpha_i < 1$ to ensure a second-order stationary and ergodic process. To ensure coherence, we parametrize the intercept as $\omega_i = 1 - \alpha_i - \beta_i$, because it is assumed that $\text{Var}(s_{it}) = 1$. Finally, to satisfy Assumption 2 for the
Applying the theorems of He et al. (1999) and Carrasco and Chen (2002), we note that these are sufficient conditions for strict stationarity and strong mixing with a geometric rate for the GARCH(1, 1)-τ model, and necessary and sufficient conditions to ensure Assumption 2 for the Huber_{2,25}-based objective function is satisfied. We estimate each univariate model by maximizing a truncated version of the respective conditional log-likelihood function.

The estimated mixing matrix, its inverse, and the fitted univariate models’ parameters, including the estimated degrees of freedom, are listed in Table 7. All conditional distribution parameters are significantly different from 0 at the 5% level, and the stated parameter constraints, including Equations (14), have all been met for these estimates. Ljung–Box statistics and p-values for the standardized residuals \( \hat{\epsilon}_t \) and their squares \( \hat{\epsilon}_t^2 \) are computed in Table 6(b), with p-values based on 1000 simulations of the fitted process. The results indicate that the fitted model is adequate.

From the estimated inverse mixing matrix \( \hat{\mathbf{M}}^{-1} \), we see that the first DOC in volatility is weighted almost entirely by the S&P 500 Index (1.1831), with little weight given to the other two assets (0.0058 and −0.0344, respectively). The fitted volatility model for this DOC implies high persistence in the conditional variance. The third DOC in volatility largely consists of the difference between Intel Corporation and the Index; its fit is similar, but with less persistence and more excess kurtosis. The second DOC in volatility largely consists of the difference between Cisco Systems and the Index; the fit for this DOC is much more adaptive and has much less excess kurtosis. From the estimated mixing matrix \( \hat{\mathbf{M}} \), the conditional variance of the Index closely matches that of the first DOC. The volatility of Cisco stock is heavily weighted toward the second DOC (2.3755), but considerable weight is also given to the first DOC (1.4362) and a little weight is given to the third DOC (0.5550). Similarly, the volatility of Intel stock is heavily weighted toward the third DOC, but considerable weight is also given to the first DOC.

Figure 4(a) gives the fitted conditional standard deviation process for the S&P 500 Index return series. Figure 5(a) gives the estimated conditional correlations between the S&P 500 Index and Intel Corporation returns. A rolling correlation estimator with a 6-month window is plotted for comparison. Although the dynamics of the conditional correlations are not explicitly specified, they are clearly changing with time and generally follow the trends of the rolling estimator. Applications of estimation have been discussed by Matteson and Tsay (2007).

Details of the competing approaches (the O-GARCH, DCC, and CUC models) are given in the online Supplementary Materials and Figures 4(b)–(d) and 5(b)–(d). In this analysis, the DOC in volatility approach perform similarly to the DCC model, but without explicitly specifying common dynamics for the conditional correlation matrix. The O-GARCH model is not adequate for modeling dynamic correlations. Although the DOC and CUC methods perform similarly, the DOC in volatility approach estimates a smoother conditional correlation processes, and only DOCs in volatility can be characterized by a diagonal volatility model.

### 4. CONCLUSION

Modeling multivariate time series via DOCS reduces the curse of dimensionality, effectively reducing the multivariate estimation to a set of disjoint univariate models. In autoregressive modeling, our approach provides a parsimonious representation that has good in-sample and out-of-sample empirical performance. In multivariate volatility modeling, correlations evolve over time without explicit modeling, the volatility matrix is positive-definite at every time point, and the in-sample fit is adequate in our application. Consistent estimation using our proposed model is simple to achieve and allows numerous parameterizations. The assumption of a constant mixing parameters.
matrix, estimation based on more robust objective functions, and efficiency gains from joint estimation is left for further research.

SUPPLEMENTARY MATERIALS

Technical Arguments: Comments on the parameterization and identifiability, proofs of Theorems 1 and 2, and additional details for the competing methods discussed in Sections 3.1 and 3.2. (SupplementalMaterials.pdf)

[Received October 2010. Revised May 2011.]

REFERENCES

Figure 5. Fitted conditional correlations for S&P 500 Index and Intel Corporation daily percentage log returns: (a) DOC-GARCH(1, 1); (b) O-GARCH(1, 1); (c) DCC-GARCH(1, 1); and (d) CUC-GARCH(1, 1). A rolling correlation estimator with a 6-month window is plotted with dashed lines. The online version of this figure is in color.


