Econometric Theory, **28**, 2012, 629–669.
doi:10.1017/S0266466611000363

LOCAL INSTRUMENTAL VARIABLE METHOD FOR THE GENERALIZED ADDITIVE-INTERACTIVE NONLINEAR VOLATILITY MODEL ESTIMATION

MICHAEL LEVINE
Purdue University

JINGUANG (TONY) LI
Canadian Pension Fund

In this article we consider a new separable nonparametric volatility model that includes second-order interaction terms in both mean and conditional variance functions. This is a very flexible nonparametric ARCH model that can potentially explain the behavior of the wide variety of financial assets. The model is estimated using the generalized version of the local instrumental variable estimation method first introduced in Kim and Linton (2004, *Econometric Theory* 20, 1094–1139). This method is computationally more effective than most other nonparametric estimation methods that can potentially be used to estimate components of such a model. Asymptotic behavior of the resulting estimators is investigated and their asymptotic normality is established. Explicit expressions for asymptotic means and variances of these estimators are also obtained.

1. INTRODUCTION

Volatility modeling has been one of the most active research areas in empirical finance and time series econometrics in the past two decades. Given the importance of predicting volatility in a number of asset-pricing and portfolio management problems, this is hardly surprising. The most popular class of volatility models historically has been the class of ARCH/GARCH (autoregressive conditional heteroskedasticity/generalized ARCH) models, originally introduced by Engle (1982) and generalized to GARCH by Bollerslev (1986).

The parameterization employed in the classical ARCH and GARCH models cannot capture some of the salient features of financial data such as, for example, the leverage effect first documented in Black (1976). In the parametric setting, one possible way to capture leverage effect is to use models with cross-product

Address correspondence to Michael Levine, Department of Statistics, Purdue University 250 N. University St. West Lafayette, IN 47907-2066; e-mail: mlevins@purdue.edu

© Cambridge University Press 2011

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

terms in the conditional variance function. One of the most important models of that kind that generalizes the standard ARCH (GARCH) paradigm is the QARCH (Quadratic ARCH) model of Sentana (1995) that considers the conditional variance function to be a quadratic polynomial of the past q values of the process Y_t.

That model encompasses a number of earlier volatility models, such as the classical ARCH model of Engle (1982) and the augmented ARCH (AARCH) of Bera and Lee (1990) and others. Another example of an earlier model that has interaction terms in the conditional variance and is capable of capturing the leverage effect is the nonlinear asymmetric GARCH (NGARCH) model of Engle and Ng (1993).

There are a few reasons why parametric models are sometimes inadequate. First, there is often very little reason to choose one parametric specification over the other; as a result, parametric models are often subject to misspecification. Nonparametric specification, on the contrary, imposes only some basic smoothness constraints on the conditional variance function. In practice, in order to make the model tractable, additional assumptions (such as additivity) often need to be imposed; however, even in this case, it is a much more general approach to modeling than the parametric one. Second, many features present in the data (such as nonnormality, asymmetric cycles, nonlinearity between lagged variables) require nonlinear models to describe the law that generates the data. However, the number of such models is infinitely large, and time series analysts cannot explore all of them. Therefore, a more general nonparametric setting can be explored as an alternative.

The general nonparametric volatility model assumes that the conditional variance function depends on a given number of past lags of the process Y_t; in other words, the conditional variance function $v \equiv v(Y_{t-1}, \ldots, Y_{t-d})$ for some integer $d > 0$. In general, the unstructured version of this model suffers from the well-known "curse of dimensionality" problem. This means that the best possible rate of convergence for an estimator of v quickly decreases as the number of dimensions d increases. Such a model, under assumption of zero mean, was considered in Pagan and Schwert (1990) and Pagan and Hong (1991). Historically, the first way to get around the curse of dimensionality has been to assume the generalized additive structure of the function to be estimated. Thus, a simple generalized additive model for the conditional variance function is $v(Y_{t-1}, \ldots, Y_{t-d}) = v_1(Y_{t-1}^2) + \cdots + v(Y_{t-p}^2)$; such a model has been discussed in, for example, Fan and Yao (2003). This model is, clearly, a generalization of the ARCH(p) model. In many situations, however, a generalized additive structure is not sufficiently flexible, and other alternatives have to be considered.

Yang, Härdle, and Nielsen (1999) introduced the nonparametric volatility model with additive mean structure but multiplicative volatility; they argued that this is rather natural since volatility function must be presumed to be nonnegative. A rather general nonparametric volatility model was considered by Kim and Linton (2004), who called it GANARCH (generalized additive nonlinear ARCH). That model defines both conditional mean and

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

conditional variance as $m(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) = F_m(C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha}))$ and $v(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) = F_v(C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}))$, where $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ are any smooth but unknown functions, while $F_m(\cdot)$ and $F_v(\cdot)$ are known monotone transformations. Note also that the generalized additive model in both mean and variance is also a special case of the GANARCH model when both link functions are equal to 1. The GANARCH model assumes that the link function is known and, therefore, some information about the distribution of data is available. Such information is often very hard to obtain, particularly in multidimensional settings.

We are interested in studying a model that avoids the assumption of the known distribution of the data that is needed for the GANARCH model of Kim and Linton (2004). Because of that, we do not consider link functions for mean and variance. We would also like to avoid the fairly rigid assumption of additivity of the mean and variance function; therefore, we take a step ahead and introduce a model that contains nonparametric "interactions" in the volatility function as well as in the conditional mean function. The resulting model is indirectly related to the model of Kim and Linton and can be represented in the nonlinear ARCH form as

$$y_t = m(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) + v^{1/2}(y_{t-1}, y_{t-2}, \ldots, y_{t-d})\varepsilon_t, \tag{1}$$

$$m(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) = C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha}) + \sum_{1 \le \alpha < \beta \le d} m_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta}), \tag{2}$$

$$v(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) = C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}) + \sum_{1 \le \alpha < \beta \le d} v_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta}), \tag{3}$$

where $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ are smooth but unknown univariate functions while $m_{\alpha\beta}(\cdot)$ and $v_{\alpha\beta}(\cdot)$ are also smooth but unknown bivariate functions. Such a model can be viewed as a generalization of the QARCH model of Sentana (1995) and, indeed, nests the QARCH model of Sentana directly with the appropriate choice of functions v_α and $v_{\alpha\beta}$. As is the case with QARCH, this model allows us to consider the possibility of two lags of the process y_t with the same sign influencing the volatility more than the regular ARCH model, or even a generalized additive model for conditional variance, would allow. Due to the nonparametric specification of the conditional variance function, this influence can take on many different functional forms and is therefore more flexible than QARCH. As is the case with the QARCH model, the model (1)–(3) also allows for a dynamic asymmetric effect of positive and negative values of y_t on the conditional variance; thus, it is capable of capturing the leverage effect that may have an

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

arbitrary functional form. In this way it is also more flexible than the QARCH model.

In terms of estimation, note that, according to Stone (1994), interactive components in such a model can be estimated at the same rate as a regular two-dimensional nonparametric smoothing problem; more specifically, the interactive terms can be estimated at the optimal rate $O(n^{-q/(2q+2)})$ whenever the function to be estimated is q-smooth in the sense of Stone. In other words, the interactive effect estimation is as hard a problem as a two-dimensional nonparametric smoothing. Note also that the expansion of functions m and v used in (1) can be interpreted as a functional analysis of variance (ANOVA) representation for general functions $m(y_{t-1}, \ldots, y_{t-d})$ and $v(y_{t-1}, \ldots, y_{t-d})$.

The specific contribution of this paper consists in proposing an easy to implement and computationally effective estimation method for the model (1) that represents a nontrivial extension of the instrumental variable method of Kim and Linton (2004). Moreover, we prove asymptotic normality of the resulting estimators under a fairly mild set of regularity conditions and provide explicit expressions of asymptotic mean and variance for all of the functional components' estimators.

The rest of the paper is organized as follows. In Section 2 we make some comments on stationarity and strong mixing property of our model. Section 3 describes the main estimation idea and defines the local instrumental variable estimators. In Section 4 we describe our main theoretical results and then give the proof of the first one in Section 5. The last section contains discussion of the results. Simulation Appendix A and an additional theoretical Appendix B containing proof of Theorem 2 are available on the first author's professional webpage at http://www.stat.purdue.edu/∼mlevins/Articles/simulation_appendix.pdf and http://www.stat.purdue.edu/∼mlevins/Articles/Proof_Theorem2.pdf, respectively.

2. STATIONARITY, STRONG MIXING, AND IDENTIFIABILITY PROPERTIES OF THE MODEL (1)–(3)

Under some weak assumptions, the general nonlinear autoregressive time series model can be shown to be stationary and strongly mixing with mixing coefficients decaying exponentially fast. Auestad and Tjostheim (1990) used α-mixing or geometric ergodicity to identify the nonlinear time series model. Ango Nze (1992) studied the L_1 geometric ergodicity of the multivariate model $X_t = f(X_{t-1}, \ldots, X_{t-p}) + H(X_{t-1}, \ldots, X_{t-q})\epsilon_t$, where X_t and ϵ_t are two sequences of m-dimensional random variables defined on a common probability space and ϵ_t is an m-dimensional white noise process. Ango Nze gave, probably, the first in the literature sufficient condition that ensures L_1 geometric ergodicity of such a model. Lu and Jiang (2001) derived another sufficient condition that also ensures L_1 geometric ergodicity of the same model but that is much less restrictive. In this paper we impose constraints from Lu and Jiang and, in doing

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

so, assume that the conditions for strict stationarity and strong mixing property of the process $\{y_t\}_{t=1}^n$ in (1)–(3) are met.

The model (1)–(3) is not identifiable unless specific conditions are met. These conditions are similar to those of Kim and Linton (2004); they can be summarized as

$$\mathrm{E}[m_\alpha(Y_{t-\alpha})] = 0, \qquad \alpha = 1, \ldots, d, \tag{4}$$

$$\mathrm{E}[v_\alpha(Y_{t-\alpha})] = 0, \qquad \alpha = 1, \ldots, d, \tag{5}$$

and

$$\mathrm{E}[m_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\alpha} = y_\alpha] = \mathrm{E}[m_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\beta} = y_\beta] = 0, \tag{6}$$

$$\mathrm{E}[v_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\alpha} = y_\alpha] = \mathrm{E}[v_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\beta} = y_\beta] = 0, \tag{7}$$

where $1 \le \alpha < \beta \le d$.

3. THE LOCAL INSTRUMENTAL VARIABLE ESTIMATORS

3.1. Estimation Methods

The generalized additive model literature suggests several possible approaches that can be conceptually extended to our model. The first is the so-called backfitting algorithm of Breiman and Friedman (1985); see also Hastie and Tibshirani (1987, 1990), as well as Buja, Hastie, and Tibshirani (1989). It can be extended to our model but it is rather hard to analyze theoretically. Another possibility is to extend the marginal integration method that was introduced independently by Newey (1994), Tjostheim and Auestad (1994), and Linton and Nielsen (1995). However, marginal integration is relatively computationally expensive even in the case of a generalized additive model, and this shortcoming is only going to become more dramatic if it is used to fit our proposed model.

In this paper we suggest an alternative approach that is a generalization of the LIVE idea suggested first in Kim and Linton (2004). There, instrumental variables were used to estimate each additive component of the mean and/or variance function separately. We extend this approach by suggesting an extra set of instruments that can be used to estimate the interactive components of both functions as well. One of the most important advantages of this approach is that it reduces the number of smoothings required to estimate a model component by a factor of n, compared to the similar (and related) marginal integration method; for example, it takes only $O(n^2)$ smoothings to estimate an additive component in the example we give later in the next section as opposed to $O(n^3)$ when using marginal integration. Also, the asymptotic properties of the LIVE are fairly easy to derive and, in this regard, it is much more tractable than the backfitting. The rest of this section is dedicated to the description of the LIVE approach in our context.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

3.2. LIVE Algorithm for the Additive-Interactive Model (1)–(3)

We begin with introducing the notation that will be used repeatedly throughout the paper. We denote $\mathbf{y}_t = (y_{t-1}, \ldots, y_{t-d})$, $\mathbf{y} = (y_1, \ldots, y_d)$. We use underscore to imply that a particular direction α or directions α and β have been omitted; boldface is used for all multidimensional quantities. Thus, $\mathbf{y}_{t,\underline{\alpha}} = (y_{t-1}, \ldots, y_{t-\alpha+1}, y_{t-\alpha-1}, \ldots, y_{t-d})$ and $\mathbf{y}_{\underline{\alpha}} = (y_1, \ldots, y_{\alpha-1}, y_{\alpha+1}, \ldots, y_d)$. Analogously, $\mathbf{y}_{t,\alpha\beta} = (y_{t-\alpha}, y_{t-\beta})$, $\mathbf{y}_{\alpha\beta} = (y_\alpha, y_\beta)$, while $\mathbf{y}_{t,\underline{\alpha\beta}}$ and $\mathbf{y}_{\underline{\alpha\beta}}$ are defined analogously to $\mathbf{y}_{t,\underline{\alpha}}$ and $\mathbf{y}_{\underline{\alpha}}$. Let $p_\alpha(y_\alpha)$ be the marginal density of $y_{t-\alpha}$, while $p_{\underline{\alpha}}(\mathbf{y}_{\underline{\alpha}})$, $p_{\alpha\beta}(y_\alpha, y_\beta)$, $p_{\underline{\alpha\beta}}(\mathbf{y}_{\underline{\alpha\beta}})$, and $p(\mathbf{y})$ are joint densities of $\mathbf{y}_{t,\underline{\alpha}}$, $\mathbf{y}_{t,\alpha\beta}$, $\mathbf{y}_{t,\underline{\alpha\beta}}$, and $\mathbf{y}_t$, respectively.

1. *Preliminary Density Estimation.* The instrumental variables we are going to define depend on several marginal and joint densities. Thus, these density functions need to be estimated first. As we mentioned before, we use regular product kernel density estimators. Specifically, we estimate the marginal density $p_\alpha(\cdot)$ as

$$\hat{p}_\alpha(y_\alpha) = \frac{1}{ng_1}\sum_{t=1}^{n} L\left(\frac{y_{t-\alpha} - y_\alpha}{g_1}\right), \qquad \alpha = 1, 2, \ldots, d,$$

and, analogously, joint densities $p_{\alpha\beta}$, $p_{\underline{\alpha\beta}}$, $p_{\underline{\alpha}}$, and $p(\mathbf{y})$ as

$$\hat{p}_{\alpha\beta}(y_\alpha, y_\beta) = \frac{1}{ng_\alpha g_\beta}\sum_{t=1}^{n} L\left(\frac{y_{t-\alpha} - y_\alpha}{g_\alpha}\right) L\left(\frac{y_{t-\beta} - y_\beta}{g_\beta}\right), \qquad 1 \le \alpha < \beta \le d,$$

$$\hat{p}_{\underline{\alpha\beta}}(y_{\underline{\alpha}}, y_{\underline{\beta}}) = \frac{1}{n\prod\limits_{\substack{\lambda=1\\ \lambda\notin\{\alpha,\beta\}}}^{d} g_\lambda}\sum_{t=1}^{n}\prod_{\substack{\lambda=1\\ \lambda\notin\{\alpha,\beta\}}}^{d} L\left(\frac{y_{t-\lambda} - y_\lambda}{g_\lambda}\right), \qquad 1 \le \alpha < \beta \le d,$$

$$\hat{p}_{\underline{\alpha}}(y_{\underline{\alpha}}) = \frac{1}{n\prod\limits_{\substack{\lambda=1\\ \lambda\neq\alpha}}^{d} g_\lambda}\sum_{t=1}^{n}\prod_{\substack{\lambda=1\\ \lambda\neq\alpha}}^{d} L\left(\frac{y_{t-\lambda} - y_\lambda}{g_\lambda}\right), \qquad \alpha = 1, 2, \ldots, d \quad \text{and}$$

$$\hat{p}(\mathbf{y}) = \frac{1}{n\prod\limits_{\alpha=1}^{d} g_\alpha}\sum_{t=1}^{n}\prod_{\alpha=1}^{d} L\left(\frac{y_{t-\alpha} - y_\alpha}{g_\alpha}\right).$$

In the above, $g_i = g_i(n)$ are the bandwidths, $i = 1, \ldots, d$, and $L(\cdot)$ is the unimodal one-dimensional symmetric kernel function.

Remark 1. Of course, multivariate product kernels are not the only possibility we could have considered. In general, two popular ways of constructing multivariate kernels are usually considered. The product kernel is the first, and the second is the so-called spherically symmetric multivariate kernel. In general, multivariate product kernel-based estimators are slightly less efficient than those based on spherically symmetric kernels (for details, see, e.g., Wand and Jones, 1995).

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

However, since the observed loss of efficiency is rather minor, we prefer to use the product kernel that implies an easy and straightforward notation.

2. *Estimation of the Constant Component of the Mean C_m.* C_m can be directly estimated as $\hat{C}_m = \frac{1}{n}\sum_{t=1}^{n} y_t$.

3. *Estimation of the Additive Components of the Mean $m_\alpha(\cdot)$.* Define the instrumental variable

$$\hat{W}_\alpha(\mathbf{y}_t) = \frac{\hat{p}_\alpha(y_{t-\alpha})\hat{p}_{\underline{\alpha}}(\mathbf{y}_{t,\underline{\alpha}})}{\hat{p}(\mathbf{y}_t)}, \qquad \alpha = 1, 2, \ldots, d,$$

denote $\tilde{y}_t = y_t - \hat{C}_m$, and use it to estimate $m_\alpha(y_\alpha)$ as

$$\hat{m}_\alpha(y_\alpha) = \mathrm{E}[\hat{W}_\alpha(\mathbf{y}_t)\tilde{y}_t | y_{t-\alpha} = y_\alpha], \qquad \alpha = 1, 2, \ldots, d.$$

The conditional expectation above is estimated using the local linear regression; in other words, $\hat{m}_\alpha(y_\alpha)$ in the above is the minimizer a_α of the kernel-weighted least squares problem $\min_{a_\alpha, b_\alpha} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha)\{\hat{W}_\alpha(\mathbf{y}_t)\tilde{y}_t - a_\alpha - b_\alpha(y_{t-\alpha} - y_\alpha)\}^2$. Here and further, we use h to denote the bandwidth used to estimate functional components of the conditional mean and variance functions. Also, from this point on we denote the kernel function $K(\cdot)$ to suggest that it is not necessarily the same as the kernel $L(\cdot)$ used to estimate marginal and joint densities earlier.

4. *Estimation of the Interactive Components of the Mean $m_{\alpha\beta}(\cdot)$.* Let us denote $\bar{y}_t = y_t - \left[\hat{C}_m + \sum_{\alpha=1}^{d} \hat{m}_\alpha(y_{t-\alpha})\right]$. Define the instrumental variable

$$\hat{W}_{\alpha\beta}(\mathbf{y}_t) = \frac{\hat{p}_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta})\hat{p}_{\underline{\alpha\beta}}(\mathbf{y}_{t,\underline{\alpha\beta}})}{\hat{p}(\mathbf{y}_t)}, \qquad 1 \le \alpha < \beta \le d,$$

and estimate the interactive component $m_{\alpha\beta}$ by means of the minimizer $a_{\alpha\beta}$ of the two-dimensional kernel-weighted least squares problem

$$\min_{a_{\alpha\beta}, \mathbf{b}_{\alpha\beta}} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha) K_h(y_{t-\beta} - y_\beta)$$
$$\times \{\hat{W}_{\alpha\beta}(\mathbf{y}_t)\bar{y}_t - a_{\alpha\beta} - b_{\alpha\beta,\alpha}(y_{t-\alpha} - y_\alpha) - b_{\alpha\beta,\beta}(y_{t-\beta} - y_\beta)\}^2.$$

In the above, the vector "slope" $\mathbf{b}_{\alpha\beta} = (b_{\alpha\beta,\alpha}, b_{\alpha\beta,\beta})'$.

5. *Estimation of the Constant Component of the Variance C_v.* Denote the squared residuals from the mean estimation

$$y_t^* = \left(\bar{y}_t - \sum_{1 \le \alpha < \beta \le d} \hat{m}_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta})\right)^2,$$

and estimate C_v as $\hat{C}_v = \frac{1}{n}\sum_{t=1}^{n} y_t^*$.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

6. *Estimation of the Additive Components of the Variance* $v_\alpha(\cdot)$. Using the instrumental variables defined in step (3) we can estimate $v_\alpha(\cdot)$ as the minimizer of the localized least squares problem,

$$\min_{a_\alpha, b_\alpha} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha)\{\hat{W}_\alpha(\mathbf{y}_t) y_t^* - a_\alpha - b_\alpha(y_{t-\alpha} - y_\alpha)\}^2.$$

7. *Estimation of the Interactive Components of the Variance* $v_{\alpha\beta}(\cdot)$. Analogously to estimation of the interactive components of the mean, denote $\tilde{y}_t^* = y_t^* - \left[\hat{C}_v + \sum_{\alpha=1}^d \hat{v}_\alpha(y_{t-\alpha})\right]$ and estimate interactive components $v_{\alpha\beta}(\cdot)$ as

$$\min_{a_{\alpha\beta}, \mathbf{b}_{\alpha\beta}} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha) K_h(y_{t-\beta} - y_\beta)$$

$$\times \{\hat{W}_{\alpha\beta}(\mathbf{y}_t)\tilde{y}_t^* - a_{\alpha\beta} - b_{\alpha\beta,\alpha}(y_{t-\alpha} - y_\alpha) - b_{\alpha\beta,\beta}(y_{t-\beta} - y_\beta)\}^2.$$

4. MAIN RESULTS

In this section we state the main results for estimation in our additive-interactive nonlinear ARCH model. For this, we need the following definitions and assumptions. Let $\mathcal{F}_a^b$ be the σ-algebra generated by $\{y_t\}_{t=a}^{t=b}$ and $\alpha(k)$ the strong mixing coefficient of $\{y_t\}$ defined as $\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |\mathrm{P}(\mathrm{A} \cap \mathrm{B}) - \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})|$.

Condition 1. $\{y_t\}_{t=1}^\infty$ is a stationary and strongly mixing process generated by (1)–(3), with a mixing coefficient $\alpha(k)$ such that $\sum_{k=0}^\infty k^a \{\alpha(k)\}^{1-2/\nu} < \infty$, for some $\nu > 2$ and $a > (1 - 2/\nu)$. For simplicity, we assume that the process $\{y_t\}_{t=1}^\infty$ has a compact support.

Condition 2. The functions $m_\alpha(\cdot), m_{\alpha\beta}(\cdot), v_\alpha(\cdot), v_{\alpha\beta}(\cdot)$, $1 \le \alpha, \beta \le d$, are continuous and twice differentiable with bounded (partial) derivatives on the compact support. Also, the joint and marginal density functions $p(\cdot)$, $p_\alpha(\cdot)$, $p_{\underline{\alpha}}(\cdot)$, $p_{\alpha\beta}(\cdot)$, and $p_{\underline{\alpha\beta}}(\cdot)$ are twice continuously differentiable and have bounded third derivatives. All of the density functions above are also bounded away from zero on the compact support.

Condition 3. The functions $L(\cdot)$ and $K(\cdot)$ are bounded, nonnegative, symmetric around zero, compactly supported, Lipschitz continuous first-order kernels. Furthermore, we assume their moments of order higher than 2 are not equal to zero and $\| \underline{u} \|^2 L(\underline{u}) \in L_1$, $\| \underline{u} \|^4 K(\underline{u}) \in L_1$, and $\| \underline{u} \|^{(2\nu+d)} K(\underline{u}) \to 0$ as $\| \underline{u} \| \to \infty$.

Condition 4. As $g \to 0$, $h \to 0$, and $n \to \infty$ we have $ng^d \to \infty$, $nh^s \to \infty$, and $\frac{g^{2\nu}}{h^s} \to 0$. Also, there exists a sequence of positive integers satisfying $t(n) \to \infty$ and $t(n) = o\left(\sqrt{nh^s}\right)$ such that $\sqrt{\frac{n}{h^s}}\alpha(t(n)) \to 0$. Finally, $\sqrt{\frac{\log n}{nh^s}} \to 0$, as $n \to \infty$, $h \to 0$, and $nh^s \to \infty$, where $s \in \{1, 2\}$.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

Condition 1 implies that strong mixing coefficients decay at the polynomial rate. This is milder than the usual assumption for a strongly mixing process where the rate is exponential. This assumption is similar to Assumption A1 in Kim and Linton (2004). Conditions 2 and 3 are standard in kernel and local polynomial estimation. Condition 4 on the bandwidths is needed to show asymptotic negligibility of the stochastic error terms resulting from the preliminary estimation of marginal densities $p_\alpha(\cdot)$, $p_{\alpha\beta}(\cdot,\cdot)$, and the functions $m(\cdot)$ and $\upsilon(\cdot)$. It is also needed to take into account the effect of dependence of the mixing process on the asymptotic results. Before presenting the main result, some extra notation has to be introduced. Let us denote $\mu_K^l \equiv \int K(u)u^l du$ and $\mu_L^l \equiv \int L(u)u^l du$ for $l = 2, 3$. Also, $\| K \|_2^2 \equiv \int K^2(u)du$, $\kappa_3(y_\alpha, z_{\underline{\alpha}}) \equiv \mathrm{E}\left[\varepsilon_k^3 | (y_{t-\alpha}, \mathbf{y}_{t,\underline{\alpha}}) = (y_\alpha, z_{\underline{\alpha}})\right]$, and $\kappa_4(y_\alpha, z_{\underline{\alpha}}) \equiv \mathrm{E}\left[(\varepsilon_k^2 - 1)^2 | (y_{t-\alpha}, \mathbf{y}_{t,\underline{\alpha}}) = (y_\alpha, z_{\underline{\alpha}})\right]$. Let $p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha)$ be the conditional density of $\mathbf{y}_{\underline{\alpha}}$ given y_α. Define

$$\hat{\phi}_\alpha(y_\alpha) = \begin{pmatrix} \hat{m}_\alpha(y_\alpha) + \hat{C}_m \\ \hat{\upsilon}_\alpha(y_\alpha) + \hat{C}_\upsilon \end{pmatrix}, \qquad \phi_\alpha(y_\alpha) = \begin{pmatrix} m_\alpha(y_\alpha) + C_m \\ \upsilon_\alpha(y_\alpha) + C_\upsilon \end{pmatrix},$$

$$B_\alpha(y_\alpha) = \begin{pmatrix} b_\alpha^m(y_\alpha) \\ b_\alpha^\upsilon(y_\alpha) \end{pmatrix}, \quad \text{and} \quad \Sigma_\alpha^*(y_\alpha) = \begin{pmatrix} \sigma_\alpha^m(y_\alpha) & \sigma_\alpha^{m\upsilon}(y_\alpha) \\ \sigma_\alpha^{m\upsilon}(y_\alpha) & \sigma_\alpha^\upsilon(y_\alpha) \end{pmatrix},$$

where

$$b_\alpha^m(y_\alpha) = \frac{1}{2}h^2\mu_K^2 m_\alpha^{(2)}(y_\alpha) + \frac{1}{2}g^2\mu_L^2 \int \left[p_{\underline{\alpha}}^{(2)}(z_{\underline{\alpha}}) + \frac{p_\alpha^{(2)}(y_\alpha)}{p_\alpha(y_\alpha)} p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} p^{(2)}(y_\alpha, z_{\underline{\alpha}}) \right] m(y_\alpha, z_{\underline{\alpha}}) dz_{\underline{\alpha}},$$

and $b_\alpha^\upsilon(y_\alpha)$ is the same as above with υ substituted for m. Let

$$\sigma_\alpha^m(y_\alpha) = \| K \|_2^2 \int \left\{ \frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} \upsilon(y_\alpha, z_{\underline{\alpha}}) + \left[p_{\underline{\alpha}}(z_{\underline{\alpha}}) m(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha) m_\alpha(y_\alpha) \right]^2 \right\} dz_{\underline{\alpha}},$$

$$\sigma_\alpha^\upsilon(y_\alpha) = \| K \|_2^2 \int \left\{ \frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} \upsilon^2(y_\alpha, z_{\underline{\alpha}}) \kappa_4(y_\alpha, z_{\underline{\alpha}}) + \left[p_{\underline{\alpha}}(z_{\underline{\alpha}}) \upsilon(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha) \upsilon_\alpha(y_\alpha) \right]^2 \right\} dz_{\underline{\alpha}},$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

and

$$\sigma_\alpha^{mv}(y_\alpha) = \| K \|_2^2 \int \left[p_{\underline{\alpha}}(z_{\underline{\alpha}})m(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha)m_\alpha(y_\alpha)\right]$$

$$\left[p_{\underline{\alpha}}(z_{\underline{\alpha}})v(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha)v_\alpha(y_\alpha)\right]$$

$$+\frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})}v^{3/2}(y_\alpha, z_{\underline{\alpha}})\kappa_3(y_\alpha, z_{\underline{\alpha}})dz_{\underline{\alpha}}.$$

THEOREM 1. *Let y_α be in the interior of the support of $p_\alpha(\cdot)$. Then, under Conditions* 1–4 *with $s = 1$, we have*

$$\sqrt{nh}[\hat{\phi}_\alpha(y_\alpha) - \phi_\alpha(y_\alpha) - B_\alpha(y_\alpha)] \xrightarrow{d} N[0, \Sigma_\alpha^*(y_\alpha)].$$

Now define analogously

$$\hat{\phi}_{\alpha\beta}(y_\alpha, y_\beta) = \begin{pmatrix} \hat{m}_{\alpha\beta}(y_\alpha, y_\beta) + \hat{C}_m \\ \hat{v}_{\alpha\beta}(y_\alpha, y_\beta) + \hat{C}_v \end{pmatrix}, \qquad \phi_{\alpha\beta}(y_\alpha, y_\beta) = \begin{pmatrix} m_{\alpha\beta}(y_\alpha, y_\beta) + C_m \\ v_{\alpha\beta}(y_\alpha, y_\beta) + C_v \end{pmatrix},$$

$$B_{\alpha\beta}(y_\alpha, y_\beta) = \begin{pmatrix} b_{\alpha\beta}^m(y_\alpha, y_\beta) \\ b_{\alpha\beta}^v(y_\alpha, y_\beta) \end{pmatrix}, \quad \text{and}$$

$$\Sigma_{\alpha\beta}^*(y_\alpha, y_\beta) = \begin{pmatrix} \sigma_{\alpha\beta}^m(y_\alpha, y_\beta) & \sigma_{\alpha\beta}^{mv}(y_\alpha, y_\beta) \\ \sigma_{\alpha\beta}^{mv}(y_\alpha, y_\beta) & \sigma_{\alpha\beta}^v(y_\alpha, y_\beta) \end{pmatrix},$$

where

$$B_{\alpha\beta}(y_\alpha, y_\beta) = \frac{h^2}{2}\mu_K^2\left[\frac{\partial^2\phi_{\alpha\beta}}{\partial y_{t-\alpha}^2}(y_\alpha, y_\beta) + \frac{\partial^2\phi_{\alpha\beta}}{\partial y_{t-\beta}^2}(y_\alpha, y_\beta)\right]$$

$$+\frac{g^2}{2}\mu_L^2\int\left[p_{\underline{\alpha\beta}}^{(2)}(z_{\underline{\alpha\beta}}) + \frac{p_{\alpha\beta}^{(2)}(y_\alpha, y_\beta)}{p_{\alpha\beta}(y_\alpha, y_\beta)}p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})\right.$$

$$\left. - \frac{p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})}p^{(2)}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})\right]\Delta_{\underline{\alpha\beta}}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})dz_{\underline{\alpha\beta}}$$

$$+\int p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})\Delta_{\alpha\beta}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})dz_{\underline{\alpha\beta}},$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$\sigma^{m}_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Bigg\{ \frac{p^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} v(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) + \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) m(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) m_{\alpha\beta}(y_\alpha, y_\beta) \Big]^2 \Bigg\} dz_{\underline{\alpha\beta}},$$

$$\sigma^{v}_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Bigg\{ \frac{p^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} v^2(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) \kappa_4(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) + \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) v(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) v_{\alpha\beta}(y_\alpha, y_\beta) \Big]^2 \Bigg\} dz_{\underline{\alpha\beta}},$$

$$\sigma^{mv}_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Bigg\{ \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) m(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) m_{\alpha\beta}(y_\alpha, y_\beta) \Big] \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) v(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) v_{\alpha\beta}(y_\alpha, y_\beta) \Big] + \frac{p^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} v^{3/2}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) \kappa_3(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) \Bigg\} dz_{\underline{\alpha\beta}},$$

and

$$\Delta_{\alpha\beta}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) = \begin{pmatrix} b^m_\alpha(y_\alpha) + b^m_\beta(y_\beta) + \sum^d_{\lambda\neq\alpha,\beta} b^m_\lambda(z_\lambda) \\ b^v_\alpha(y_\alpha) + b^v_\beta(y_\beta) + \sum_{\lambda\neq\alpha,\beta} b^v_\lambda(z_\lambda) \end{pmatrix},$$

$$\Delta_{\underline{\alpha\beta}}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) = \begin{pmatrix} m_{\alpha\beta}(y_\alpha, y_\beta) - b^m_\alpha(y_\alpha) - b^m_\beta(y_\beta) + \sum_{\lambda,\theta\neq\alpha,\beta;\lambda<\theta} m_{\lambda\theta}(z_\lambda, z_\theta) - \sum_{\lambda\neq\alpha,\beta} b^m_\lambda(z_\lambda) \\ v_{\alpha\beta}(y_\alpha, y_\beta) - b^v_\alpha(y_\alpha) - b^v_\beta(y_\beta) + \sum_{\lambda,\theta\neq\alpha,\beta;\lambda<\theta} v_{\lambda\theta}(z_\lambda, z_\theta) - \sum_{\lambda\neq\alpha,\beta} b^v_\lambda(z_\lambda) \end{pmatrix}.$$

Define $p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta)$ as the conditional density of $y_{\underline{\alpha\beta}}$ given y_α and y_β.

THEOREM 2. *Let* (y_α, y_β) *be in the interior of the support of* $p_{\alpha\beta}(\cdot)$. *Then under Conditions* 1–4 *with* $s = 2$, *we have that*

$$\sqrt{nh^2}[\hat{\phi}_{\alpha\beta}(y_\alpha, y_\beta) - \phi_{\alpha\beta}(y_\alpha, y_\beta) - B_{\alpha\beta}(y_\alpha, y_\beta)] \xrightarrow{d} N[0, \Sigma^*_{\alpha\beta}(y_\alpha, y_\beta)].$$

Remark 2. Note that constants C_m and C_v are estimated by $\hat{C}_m$ and $\hat{C}_v$ with the degree of precision of $O_p\left(\frac{1}{\sqrt{n}}\right)$ and, therefore, the individual additive and interactive components $m_\alpha(y_\alpha)$, $m_{\alpha\beta}(y_\alpha, y_\beta)$, $v_\alpha(y_\alpha)$, and $v_{\alpha\beta}(y_\alpha, y_\beta)$ have the same asymptotic bias and variance as $\phi_\alpha(y_\alpha)$ and $\phi_{\alpha\beta}(y_\alpha, y_\beta)$, respectively.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

Remark 3. The asymptotic bias expressions $b_{\alpha}^{m}(y_{\alpha})$ and $b_{\alpha}^{v}(y_{\alpha})$ consist of two main terms. The first term is of the order $O(h^2)$, which is the same as in the ordinary local linear regression. This would have been the order of the asymptotic bias if the densities used in constructing instrumental variables were known. The second term in both cases has the order of $O(g^2)$ and is the penalty we pay for not knowing these densities.

Remark 4. The method that is most relevant for comparison purposes here is probably the traditional version of the marginal integration method. As described in Linton and Nielsen (1995), in the two-dimensional case it requires the practitioner to choose two different bandwidths h_1 and h_2 for the two directions. The resulting bias is then of the order $O(h_1^2 + h_2^2)$. This is the same situation one encounters in LIVE algorithm since, for an additive component that is twice continuously differentiable and is defined on the compact support, the resulting bias is then $O(h^2 + g^2)$, which is conceptually the same as in standard marginal integration. As we noted earlier, the main benefit of the, LIVE method is its speed of computation; whereas marginal integration requires $O(n^3)$ smoothings to estimate a linear component, the LIVE method only takes $O(n^2)$ smoothings.

Remark 5. It is also fairly clear that there is some room for improvement when it comes to asymptotic variances of LIVE. When compared to the regular marginal integration estimators of Linton and Nielsen (1995), one can see that, for example, the asymptotic variance of the additive mean component for LIVE is larger than the one for the marginal integration (MI) estimator. More precisely, if one assumes constant conditional variance function $v(\cdot)$, the first term of $\sigma_{\alpha}^{m}(y_{\alpha})$, which is $\| K \|_2^2 \int \left\{ \frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_{\alpha}, z_{\underline{\alpha}})} v(y_{\alpha}, z_{\underline{\alpha}}) \right\} dz_{\underline{\alpha}}$, coincides with the asymptotic variance of the MI estimator. This is related to the fact that the LIVE is based on the omitted variable regression idea that treats everything except the additive component of interest as the error term. Such an error term is not conditionally independent of the component of interest. Finally, as in Kim and Linton (2004), a number of possible efficiency standards can be considered in order to construct possible oracle estimators. This is one of the topics we intend to consider as part of the future research.

Remark 6. Note that one of the conditions of Theorems 1 and 2 is that $\frac{g^{2v}}{h} \to 0$ as $g \to 0$ and $h \to 0$. Recall that $v > 2$ is the strong mixing index that characterizes the strength of dependence of the process y_t. In practice, this means that the bandwidth g can be selected much smaller than h. In other words, undersmoothing of the densities that comprise instrumental variables is possible. If g is thus selected to be of the smaller order of magnitude then h, the penalty for not knowing true marginal and joint densities of the process y_t becomes of the order $o(h^2)$ and thus can be disregarded asymptotically.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

5. PROOF OF THEOREM 1

The proof of Theorem 1 contains four steps: first we decompose the estimation error into the main stochastic term and bias; second, we approximate each component, then we compute the asymptotic bias; and finally, we establish the asymptotic normality. Without loss of generality, we assume that $\alpha = 1$.

For expositional convenience we use the subscript 2 to denote the nuisance directions. For example, $p_2(\mathbf{y}_{t,\underline{1}}) = p_{\underline{1}}(\mathbf{y}_{t,\underline{1}})$ in the case of density function, and $m_2(\mathbf{y}_{t,\underline{1}}) = m_{\underline{1}}(\mathbf{y}_{t,\underline{1}})$, $\upsilon_2(\mathbf{y}_{t,\underline{1}}) = \upsilon_{\underline{1}}(\mathbf{y}_{t,\underline{1}})$ in the case of component functions. We use $X_n \simeq Y_n$ to signify $X_n = Y_n\{1 + o_p(1)\}$. Let $vec(X)$ denote the vector of matrix X arranged by columns.

Proof.

Step I: Decompositions. The LIVE of $\phi_1(y_1)$, as any local linear regression estimator, can be expressed in the closed form

$$\hat{\phi}_1(y_1)^T = e_1^T(\mathbf{Y}_-^T\mathbf{K}\mathbf{Y}_-)^{-1}\mathbf{Y}_-^T\mathbf{K}\mathbf{W}\tilde{\mathbf{R}}, \tag{8}$$

where

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \mathbf{Y}_- = \begin{pmatrix} 1 & y_{d+1-1} - y_1 \\ 1 & y_{d+2-1} - y_1 \\ \vdots & \vdots \\ 1 & y_{d+n-1} - y_1 \end{pmatrix}, \qquad \tilde{\mathbf{R}} = \begin{pmatrix} \tilde{r}_1 \\ \tilde{r}_2 \\ \vdots \\ \tilde{r}_n \end{pmatrix} = \begin{pmatrix} y_{d+1} & \tilde{\varepsilon}_{d+1}^2 \\ y_{d+2} & \tilde{\varepsilon}_{d+2}^2 \\ \vdots & \vdots \\ y_{d+n} & \tilde{\varepsilon}_{d+n}^2 \end{pmatrix},$$

$\mathbf{K} = diag\{K_h(y_{d+l-1} - y_1)\}$, $l = 1, \ldots, n$, $\mathbf{W} = diag\{\hat{W}_{d+k}\}$, $k = 1, \ldots, n$, and $\tilde{\varepsilon}_t^2 = (y_t - \tilde{m}(\mathbf{y}_t))^2$. The preliminary density estimators used are $\hat{p}_1(y_1) = \frac{1}{ng}\sum_{t=d+1}^{d+n} L_g(y_{t-1} - y_1)$, $\hat{p}_2(\mathbf{y}_{\underline{1}}) = \frac{1}{ng^{d-1}}\sum_{t=d+1}^{d+n}\prod_{\alpha=2}^{d} L_g(y_{t-\alpha} - y_\alpha)$, and $\hat{p}(\mathbf{y}) = \frac{1}{ng^d}\sum_{t=d+1}^{d+n}\prod_{\alpha=1}^{d} L_g(y_{t-\alpha} - y_\alpha)$, while $\hat{W}_t = \frac{\hat{p}_1(y_{t-1})\hat{p}_2(\mathbf{y}_{t,\underline{1}})}{\hat{p}(\mathbf{y}_t)}$.

Using standard properties of the *vec* operator, it is easy to verify that (8) can be written as $\hat{\phi}_1(y_1) = [I_2 \otimes e_1^T(\mathbf{Y}_-^T\mathbf{K}\mathbf{Y}_-)^{-1}][I_2 \otimes \mathbf{Y}_-^T\mathbf{K}]\upsilon ec[\mathbf{W}\tilde{\mathbf{R}}]$, while the true function $\phi_1(y_1)$ satisfies the identity

$$\phi_1(y_1) = \left[I_2 \otimes e_1^T(\mathbf{Y}_-^T\mathbf{K}\mathbf{Y}_-)^{-1}\right]\left[I_2 \otimes \mathbf{Y}_-^T\mathbf{K}\right]\upsilon ec\left[l\phi_1^T(y_1) + Y_- \bigtriangledown \phi_1^T(y_1)\right], \tag{9}$$

where

$$l = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \qquad Y_- = \begin{pmatrix} y_{d+1-1} - y_1 \\ y_{d+2-1} - y_1 \\ \vdots \\ y_{d+n-1} - y_1 \end{pmatrix}.$$

It follows easily from the above results that the estimation error can be represented as

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$\hat{\phi}_1(y_1) - \phi_1(y_1) = [I_2 \otimes e_1^T Q_n^{-1}]\tau_n, \tag{10}$$

where $Q_n = \frac{1}{n} D_h^{-1} \mathbf{Y}_-^T \mathbf{K} \mathbf{Y}_- D_h^{-1}$ and $\tau_n = \frac{1}{n}[I_2 \otimes D_h^{-1} \mathbf{Y}_-^T \mathbf{K}] vec[\mathbf{W}\tilde{\mathbf{R}} - l\phi_1^T(y_1) - Y_- \triangledown \phi_1^T(y_1)]$ with $D_h = diag\{1, h\}$.

Step II: Approximations. We rewrite τ_n in (5) as

$$\tau_n = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1)[\hat{W}_k \tilde{r}_k - \phi_1(y_1) - (y_{k-1} - y_1) \triangledown \phi_1(y_1)] \right.$$
$$\left. \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\}$$

and expand it next by adding and subtracting $\phi_1(y_{k-1})$ within the inner factor enclosed in brackets as

$$\tau_n = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1)[\hat{W}_k \tilde{r}_k - \phi_1(y_1) - (y_{k-1} - y_1) \nabla \phi_1(y_1)] \right.$$
$$\left. \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\} \tag{11}$$
$$= \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1)[\hat{W}_k \tilde{r}_k - \phi_1(y_{k-1})] \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\}$$
$$+ \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1)[\phi_1(y_{k-1}) - \phi_1(y_1) - (y_{k-1} - y_1) \nabla \phi_1(y_1)] \right.$$
$$\left. \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\}.$$

At this point we introduce the notation $r_k = [m(\mathbf{y}_k), \ v(\mathbf{y}_k)]^T$ and $\tilde{r}_k^* = \begin{pmatrix} v^{1/2}(\mathbf{y}_k)\varepsilon_k \\ v(\mathbf{y}_k)(\varepsilon_k^2 - 1) \end{pmatrix}$, which will prove useful in deriving approximations of the two terms in the last formula. By plugging in the estimated instruments and bounding the difference between the true and estimated values of instruments $W_k, k = 1, \ldots, n$, we find in a way similar to Kim and Linton (2004) that the bias can be viewed as consisting of the four components

$$T_{1n} = \frac{h^2}{2n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1) \left(\frac{y_{k-1} - y_1}{h}\right)^2 D^2 \phi_1(y_1) \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\},$$

$$T_{2n} = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1) \left[\frac{\hat{p}_1(y_{k-1}) \hat{p}_2(\mathbf{y}_{k,\underline{1}})}{\hat{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1}) p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \right.$$
$$\left. \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\},$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$T_{3n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\frac{p_1(y_{k-1})p_2\left(\mathbf{y}_{k,\underline{1}}\right)}{p(\mathbf{y}_k)}\tilde{r}_k^* \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right\},$$

$$T_{4n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}r_k - \phi_1(y_{k-1})\right] \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right\},$$

Step III: Asymptotic Bias. As a reminder, we assume that the sufficient conditions of Lu and Jiang (2001) are satisfied and, therefore, the process y_t is geometrically ergodic. This implies strict stationarity and β-mixing (therefore, also α-mixing).

Since measurable function of strictly stationary and α-mixing process is again strictly stationary and α-mixing, we can apply the ergodic theorem for α-mixing stationary processes to T_{1n} (see, e.g., Fan and Yao, 2003, for details). As a result, using the Taylor expansion of the density function $p_1(\cdot)$, we obtain

$$\begin{aligned} T_{1n} &= \frac{h^2}{2n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2 D^2\phi_1(y_1)\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\} \\ &\xrightarrow{p} \frac{h^2}{2}\mathrm{E}\left[K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2 D^2\phi_1(y_1)\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right] \\ &= \frac{h^2}{2}p_1(y_1)D^2\phi_1(y_1)\otimes(\mu_K^2,\ \mu_K^3)^T + o(h^2). \end{aligned}$$

Next, we define marginal expectations of estimated density functions $\hat{p}_1(\cdot)$, $\hat{p}_2(\cdot)$, and $\hat{p}(\cdot)$ as $\bar{p}_1(y_{k-1}) = \int L_g(z_1 - y_{k-1})p_1(z_1)dz_1$, $\bar{p}_2(\mathbf{y}_{k,\underline{1}}) = \int L_g(z_2 - \mathbf{y}_{k,\underline{1}})p_2(z_2)dz_2$, and $\bar{p}_(\mathbf{y}_k) = \int L_g(z_1 - y_{k-1})L_g(z_2 - \mathbf{y}_{k,\underline{1}})p(z_1,z_2)dz_1dz_2$.

It can be shown analogously to Kim and Linton (2004) that T_{2n} can be approximated as

$$T_{2n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k \otimes \left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\} + o_p(1/\sqrt{nh}).$$

Applying the ergodic theorem for α-mixing stationary process and using the fact that $\mathrm{E}\left[K_h(y_{k-1}-y_1)\left[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\left[\tilde{m}(\mathbf{y}_k) - m(\mathbf{y}_k)\right]^2 \otimes \left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right] = o(1/\sqrt{nh})$, we have

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$T_{2n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right\} + o_p(1/\sqrt{nh})$$

$$\xrightarrow{p} \mathrm{E}\left[K_h(y_{k-1}-y_1)\left[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right]$$

$$\simeq \frac{g^2}{2}p_1(y_1)\int\left[p_2^{(2)}(z_2) + \frac{p_1^{(2)}(y_1)}{p_1(y_1)}p_2(z_2) - \frac{p_2(z_2)}{p(y_1,z_2)}p^{(2)}(y_1,z_2)\right] \times \phi(y_1,z_2)dz_2 \otimes (\mu_L^2, 0)^T.$$

As a next step, we note that the terms T_{3n} and T_{4n} are $\mathcal{F}_{k-1}$-measurable for any k, and, therefore, we can argue that, by the ergodic theorem for α-mixing stationary processes, both T_{3n} and T_{4n} converge in probability to zero: $T_{mn} \xrightarrow{p} 0$, for $m = 3$ and $m = 4$.

Finally, for the probability limit of $[I_2 \otimes e_1^T Q_n^{-1}]$ we have $Q_n = \frac{1}{n}D_h^{-1}\mathbf{Y}^T\mathbf{K}\mathbf{Y}D_h^{-1} = \begin{pmatrix} q_{n0} & q_{n1} \\ q_{n1} & q_{n2}\end{pmatrix}$, where $q_{ni} = \frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^i$, $i = 0, 1, 2$. Using the Taylor expansion of $p_1(y_1+uh)$, it is easy to obtain $q_{ni} \xrightarrow{p} q_i$, where $q_0 = p_1(y_1)\mu_K^0 = p_1(y_1)$, $q_1 = p_1(y_1)\mu_K^1 = 0$, $q_2 = p_1(y_1)\mu_K^2$.

Clearly, $e_1^T Q_n^{-1} \xrightarrow{p} \frac{1}{p_1(y_1)}e_1^T$. By the Slutzky theorem, we conclude that the bias of the estimator $\hat{\phi}_1(y_1)$ is

$$B_{1n}(y_1) = [I_2 \otimes e_1^T Q_n^{-1}](T_{1n}+T_{2n}+T_{3n}+T_{4n}) \xrightarrow{p} \begin{pmatrix} \frac{1}{2}h^2\mu_K^2 m_1^{(2)}(y_1) + \frac{1}{2}g^2\mu_L^2\int\left[p_2^{(2)}(z_2) + \frac{p_1^{(2)}(y_1)}{p_1(y_1)}p_2(z_2) - \frac{p_2(z_2)}{p(y_1,z_2)}p^{(2)}(y_1,z_2)\right] m(y_1,z_2)dz_2 \\ \frac{1}{2}h^2\mu_K^2 v_1^{(2)}(y_1) + \frac{1}{2}g^2\mu_L^2\int\left[p_2^{(2)}(z_2) + \frac{p_1^{(2)}(y_1)}{p_1(y_1)}p_2(z_2) - \frac{p_2(z_2)}{p(y_1,z_2)}p^{(2)}(y_1,z_2)\right] v(y_1,z_2)dz_2 \end{pmatrix} \equiv B_1(y_1).$$

Step IV: Asymptotic Normality. As a final step, we derive the asymptotic distribution of $\sqrt{nh}[I_2 \otimes e_1^T Q_n^{-1}](T_{1n}+T_{2n}+T_{3n}+T_{4n})$.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

We rewrite T_{1n} as

$$T_{1n} = \frac{h^2}{2n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2 D^2\phi_1(y_1)\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}$$

$$\equiv \frac{h^2}{2}(r_{1n},\ r_{2n},\ r_{3n},\ r_{4n})^T,$$

where $r_{qn} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^{q+1} m_1^{(2)}(y_1)\right\}$, $q = 1, 2$, and $r_{qn} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^{q-1} v_1^{(2)}(y_1)\right\}$, $q = 3, 4$. Applying Theorem 1 in Masry (1996a) directly, we can easily find that each of the $nh\cdot\mathrm{var}(r_{in})$, $i = 1,\ldots,4$ is uniformly bounded: $\sup_{y_1\in R} nh\cdot\mathrm{var}(r_{1n}) \le \sup_{y_1\in R} m_1^{(2)}(y_1)\cdot C_2\nu^4(1+o(1))$, $\sup_{y_1\in R} nh\cdot\mathrm{var}(r_{2n}) \le \sup_{y_1\in R} m_1^{(2)}(y_1)\cdot C_2\nu^6(1+o(1))$, $\sup_{y_1\in R} nh\cdot\mathrm{var}(r_{3n}) \le \sup_{y_1\in R} v_1^{(2)}(y_1)\cdot C_2\nu^4(1+o(1))$, and $\sup_{y_1\in R} nh\cdot\mathrm{var}(r_{4n}) \le \sup_{y_1\in R} v_1^{(2)}(y_1)\cdot C_2\nu^6(1+o(1))$, where $\nu_i \equiv \int K^2(u)u^i du, i = 1,\ldots,6$, and C_2 is the upper bound of the density function $p_1(\cdot)$; recall that functions m_1 and $v_1(\cdot)$ have bounded second derivatives by Condition 2 and observe that, as a result, we have $\mathrm{var}(\sqrt{nh}T_{1n}) = O(h^4)$ uniformly over $y_1 \in R$. Note that the orders of the covariances are $O(h^4)$ due to the Cauchy-Schwartz inequality; i.e., $\mathrm{cov}(\sqrt{nh}r_{in},\ \sqrt{nh}r_{jn}) \le \sqrt{\mathrm{var}(\sqrt{nh}r_{in})\cdot\mathrm{var}(\sqrt{nh}r_{jn})}$, for $1 \le i < j \le 4$. It can be similarly shown that $\mathrm{var}(\sqrt{nh}T_{2n})$ is asymptotically negligible as well. To see this, rewrite T_{2n} as

$$T_{2n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{\hat{p}_1(y_{k-1})\hat{p}_2(\mathbf{y}_{k,\underline{1}})}{\hat{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k \otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}$$

$$\equiv (x_{1n},\ x_{2n},\ x_{3n},\ x_{4n})^T,$$

where

$$x_{1n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K\left(\frac{y_{k-1}-y_1}{h}\right)\left[\frac{\hat{p}_1(y_{k-1})\hat{p}_2(\mathbf{y}_{k,\underline{1}})}{\hat{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]y_k\right\}$$

$$= \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{1k}$$

$$\equiv \frac{1}{h}\frac{1}{n}\xi_1,$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$x_{3n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{3k} \equiv \frac{1}{h}\frac{1}{n}\xi_3$, with ξ_{3k} being the same as ξ_1 except that $\tilde{\epsilon}_t^2$ is substituted for y_t, and $x_{2n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{2k} \equiv \frac{1}{h}\frac{1}{n}\xi_2$ with $\xi_{2k} = \xi_{1k}\left(\frac{y_{k-1}-y_1}{h}\right)$, and $x_{4n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{4k} \equiv \frac{1}{h}\frac{1}{n}\xi_4$ with $\xi_{4k} = \xi_{3k}\left(\frac{y_{k-1}-y_1}{h}\right)$ Applying Davydov's lemma (see Rio, 1993, Thm. 1.0), we obtain $\frac{1}{n}\text{var}(\xi_1) \le 2CM_{v,1}\sum_{i\in[0,n+d-1]}\alpha(i)^{1-2/v}$, where $M_{v,1} = \text{E}[|\xi_{1k}|^v] = O(g^{2v})$ and C is a nonnegative constant. Thus we find that $\text{var}(\sqrt{nh}x_{1n}) = O\left(\frac{g^{2v}}{h}\right) \to 0$, as $n \to \infty$; it can be shown analogously that $\text{var}(\sqrt{nh}x_{in}) \to 0$, as $n \to \infty$, for $i = 2, 3, 4$. By the Cauchy-Schwartz inequality, $\text{cov}(\sqrt{nh}x_{in},\ \sqrt{nh}x_{jn}) \to 0$ as $n \to \infty$, for $1 \le i < j \le 4$ as well.

We rewrite T_{3n} as

$$T_{3n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\tilde{r}_k^* \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right\}$$

$$\equiv (s_{0n},\ s_{1n},\ t_{0n},\ t_{1n})^T,$$

where

$$s_{ln} = \frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^l\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\upsilon^{1/2}(\mathbf{y}_k)\varepsilon_k,$$

$$t_{mn} = \frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^m\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\upsilon(\mathbf{y}_k)(\varepsilon_k^2-1),$$

$l = 0, 1$, and $m = 0, 1$. We first examine the term s_{0n}. It is easy to see that $\text{E}\left[s_{0n}|\mathcal{F}_{n+d}\right] = 0$ and

$$\text{var}\left[s_{0n}|\mathcal{F}_{n+d}\right] = \text{E}\left[s_{0n}^2|\mathcal{F}_{n+d}\right]$$

$$= \frac{1}{nh}\left\{\frac{1}{n}\sum_{k=d+1}^{d+n}\frac{1}{h}K^2\left(\frac{y_{k-1}-y_1}{h}\right)\frac{p_1^2(y_{k-1})p_2^2(\mathbf{y}_{k,\underline{1}})}{p^2(\mathbf{y}_k)}\upsilon(\mathbf{y}_k)\right\}.$$

Applying the ergodic theorem for α-mixing processes and changing variables, we find that

$$\text{var}\left[\sqrt{nh}s_{0n}\right] = \text{E}\left[nh\cdot\text{var}\left[s_{0n}|\mathcal{F}_{n+d}\right]\right]$$

$$= \text{E}\left[\frac{1}{h}K^2\left(\frac{y_{k-1}-y_1}{h}\right)\frac{p_1^2(y_{k-1})p_2^2(\mathbf{y}_{k,\underline{1}})}{p^2(\mathbf{y}_k)}\upsilon(\mathbf{y}_k)\right]$$

$$= \| K \|_2^2\, p_1^2(y_1)\int\frac{p_2^2(z_2)}{p(y_1,z_2)}\upsilon(y_1,z_2)dz_2[1+o(1)].$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

Applying the central limit theorem in Fan and Yao (2003, Thm. 2.22), we have that

$$\sqrt{nh}s_{0n} \xrightarrow{d} N[0,\ \sigma^m_{0,s}(y_1)],$$

where

$$\sigma^m_{0,s}(y_1) = \| K \|_2^2\ p_1^2(y_1) \int \frac{p_2^2(z_2)}{p(y_1,z_2)} v(y_1,z_2)dz_2. \tag{12}$$

It is equally easy to conclude that $\mathrm{E}\left[t_{0n}|\mathcal{F}_{n+d}\right] = 0$ and

$$\begin{aligned}\mathrm{var}\left[t_{0n}|\mathcal{F}_{n+d}\right] &= \mathrm{E}\left[t_{0n}^2|\mathcal{F}_{n+d}\right] \\ &= \frac{1}{n^2}\mathrm{E}\left[\left(\sum_{k=d+1}^{d+n} K_h(y_{k-1}-y_1)\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} v(\mathbf{y}_k)(\varepsilon_k^2-1)\right)^2 |\mathcal{F}_{n+d}\right] \\ &= \frac{1}{n^2}\sum_{k=d+1}^{d+n} K_h^2(y_{k-1}-y_1)\frac{p_1^2(y_{k-1})p_2^2(\mathbf{y}_{k,\underline{1}})}{p^2(\mathbf{y}_k)} v^2(\mathbf{y}_k)\kappa_4(\mathbf{y}_k).\end{aligned}$$

In much the same way as before for s_{0n}, applying the ergodic theorem for α-mixing process, changing variables, and using the central limit theorem in Fan and Yao (2003), we have that

$$\sqrt{nh}t_{0n} \xrightarrow{d} N[0,\ \sigma^v_{0,s}(y_1)], \tag{13}$$

$$\text{where}\quad \sigma^v_{0,s}(y_1) = \| K \|_2^2\ p_1^2(y_1) \int \frac{p_2^2(z_2)}{p(y_1,z_2)} v^2(y_1,z_2)\kappa_4(y_1,z_2)dz_2. \tag{14}$$

Note that $\mathrm{cov}(\sqrt{nh}s_{0n}, \sqrt{nh}t_{0n}) = \mathrm{E}(nhs_{0n}t_{0n})$ since $\mathrm{E}(s_{0n}) = \mathrm{E}(t_{0n}) = 0$. Using the same approach as before, we can easily show that

$$\begin{aligned}\mathrm{cov}(\sqrt{nh}s_{0n}, \sqrt{nh}t_{0n}) &= \mathrm{E}\left[\mathrm{E}\left[nhs_{0n}t_{0n}|\mathcal{F}_{n+d}\right]\right] \\ &= \| K \|_2^2\ p_1^2(y_1) \int \frac{p_2^2(z_2)}{p(y_1,z_2)} v^{3/2}(y_1,z_2)\kappa_3(y_1,z_2)dz_2[1+o(1)].\end{aligned}$$

Combining the results for s_{0n} and t_{0n}, we have the asymptotic normality for T_{3n},

$$\sqrt{nh}T_{3n} \xrightarrow{d} N(0, \Sigma_{1,T3}), \tag{15}$$

where

$$\Sigma_{1,T3} = \begin{pmatrix} \sigma^m_{0,s}(y_1) & * & \sigma^{mv}_{0,s}(y_1) & * \\ * & * & * & * \\ \sigma^{mv}_{0,s}(y_1) & * & \sigma^v_{0,s}(y_1) & * \\ * & * & * & * \end{pmatrix}, \tag{16}$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

with $\sigma^{m}_{0,s}(y_1)$ and $\sigma^{v}_{0,s}(y_1)$ defined as in (12) and (14), while

$$\sigma^{mv}_{1,s}(y_1) \equiv \| K \|_2^2 \, p_1^2(y_1) \int \frac{p_2^2(z_2)}{p(y_1, z_2)} v^{3/2}(y_1, z_2) \kappa_3(y_1, z_2) dz_2. \tag{17}$$

The rest of the elements of the matrix $\Sigma_{1,T3}$ are left blank since they are not needed to formulate the final result.

Next, we rewrite T_{4n} as

$$\begin{aligned} T_{4n} &= \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1) \left[\frac{p_1(y_{k-1}) p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} r_k - \phi_1(y_{k-1}) \right] \right. \\ &\qquad \left. \otimes \left(1, \frac{y_{k-1} - y_1}{h} \right)^T \right\} \\ &\equiv (u_{0n}, \, u_{1n}, \, w_{0n}, \, w_{1n})^T, \end{aligned}$$

where

$$\begin{aligned} u_{ln} &= \frac{1}{n} \sum_{k=d+1}^{d+n} K_h(y_{k-1} - y_1) \left(\frac{y_{k-1} - y_1}{h} \right)^l \\ &\quad \times \left[\frac{p_1(y_{k-1}) p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} m(\mathbf{y}_k) - m_1(y_{k-1}) \right], \\ w_{mn} &= \frac{1}{n} \sum_{k=d+1}^{d+n} K_h(y_{k-1} - y_1) \left(\frac{y_{k-1} - y_1}{h} \right)^m \\ &\quad \times \left[\frac{p_1(y_{k-1}) p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} v(\mathbf{y}_k) - v_1(y_{k-1}) \right], \end{aligned}$$

for $l = 0, 1$ and $m = 0, 1$. Corollary 2 in Masry (1996b) provides us with the tool we need to establish the asymptotic normality for T_{4n},

$$\sqrt{nh} T_{4n} \xrightarrow{d} N(0, \Sigma_{1,T4}), \tag{18}$$

where

$$\Sigma_{1,T4} = \begin{pmatrix} \sigma^{m}_{0,u}(y_1) & * & \sigma^{mv}_{0,uw}(y_1) & * \\ * & * & * & * \\ \sigma^{mv}_{0,uw}(y_1) & * & \sigma^{v}_{0,w}(y_1) & * \\ * & * & * & * \end{pmatrix}, \tag{19}$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

with

$$\sigma^{m}_{0,u}(y_1) = \| K \|_2^2 \, p_1^2(y_1) \int \left[p_2(z_2) m(y_1, z_2) - p_{2|1}(z_2|y_1) m_1(y_1) \right]^2 dz_2,$$

$$\sigma^{v}_{0,w}(y_1) = \| K \|_2^2 \, p_1^2(y_1) \int \left[p_2(z_2) v(y_1, z_2) - p_{2|1}(z_2|y_1) v_1(y_1) \right]^2 dz_2,$$

$$\sigma^{mv}_{0,uw}(y_1) = \| K \|_2^2 \, p_1^2(y_1) \int \left[p_2(z_2) m(y_1, z_2) - p_{2|1}(z_2|y_1) m_1(y_1) \right] \times \left[p_2(z_2) v(y_1, z_2) - p_{2|1}(z_2|y_1) v_1(y_1) \right] dz_2.$$

Finally, it is easy to notice that $\operatorname{cov}(T_{3n}, T_{4n}) = \mathrm{E}[T_{3n} T_{4n}] = \mathrm{E}[\mathrm{E}[T_{3n} T_{4n} | \mathcal{F}_{n+d}]] = 0$. Now the asymptotic normality of an additive component is almost at hand. Here $e_1^T Q_n^{-1} \xrightarrow{p} \frac{1}{p_1(y_1)} e_1^T$ implies that $I_2 \otimes e_1^T Q_n^{-1} \xrightarrow{p} I_2 \otimes \frac{1}{p_1(y_1)} e_1^T$. Applying the Slutzky theorem, we find that

$$\sqrt{nh}[I_2 \otimes e_1^T Q_n^{-1}](T_{1n} + T_{2n} + T_{3n} + T_{4n}) \xrightarrow{d} N(0, \Sigma_1^*(y_1)),$$

and

$$\sqrt{nh}[\hat{\phi}_1(y_1) - \phi_1(y_1) - B_1(y_1)] \xrightarrow{d} N[0, \Sigma_1^*(y_1)],$$

where

$$B_1(y_1) = \begin{pmatrix} b_1^m(y_1) \\ b_1^v(y_1) \end{pmatrix}, \qquad \Sigma_1^*(y_1) = \begin{pmatrix} \sigma_1^m(y_1) & \sigma_1^{mv}(y_1) \\ \sigma_1^{mv}(y_1) & \sigma_1^v(y_1) \end{pmatrix},$$

with the components as in the statement of Theorem 1 for $\alpha = 1$.

6. DISCUSSION

The additive-interactive model (1) represents a further step in the development of the nonparametric volatility model theory. The article provides the instrumental variable-based algorithm that can be easily used to fit such a model. The algorithm is computationally efficient and easy to implement. At the same time, central limit theorems for the estimators of the individual components are obtained, and closed-form expressions for asymptotic biases and variances of these estimators are given. Among several interesting questions that remain unanswered for now in the context of model (1) is the question of testing the statistical significance of individual additive and interactive components. This is the question of obvious practical interest. It has some prior history in the cross-sectional context. Specifically, a test that can handle the separability hypothesis in the mean function under a specific alternative (inclusion of second-order interactions) for cross-sectional data had been proposed in Sperlich, Tjostheim, and Yang (2002). Consistent specification tests for nonparametric/semiparametric models proposed in Li, Hsiao, and Zinn (2003) are designed for null models that may include, among

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

other possible nonparametric components, second-order interactions. However, not much is known about similar testing problems in the time series context. Note that many of the modern applications are concerned with situations where the number of lags d considered can be quite large. Even in the cross-sectional context, multicollinearity among many different explanatory variables is very much a commonplace; in the time series context, it is always the case. Therefore, multiple hypotheses testing is, probably, much more important under these circumstances. For example, to test the separability hypothesis in the mean(variance) function for model (1), it is necessary to test $m_{\alpha\beta} \equiv 0$, $1 \leq \alpha < \beta \leq d$ ($v_{\alpha\beta} \equiv 0$, $1 \leq \alpha < \beta \leq d$, respectively). It may also be of interest to test the null hypothesis that includes both additive and interactive components. Thus, the design of the F-type tests here seems to be an important issue.

REFERENCES

Ango Nze, P. (1992) Criteres d'ergodicite de quelques modeles a representation markovenne. *Les Comptes Rendus de l'Académie des Sciences Paris* 315(1), 1301–1304.

Auestad, B. & D. Tjostheim (1990) Identification of nonlinear time series: First order characterization and order estimation. *Biometrika* 77, 669–687.

Bera, A.K. & S. Lee (1990) On the Formulation of a General Structure for Conditional Heteroskedasticity. Mimeo, University of Illinois.

Black, F. (1976) Studies in stock price volatility changes. *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, pp. 177–181. American Statistical Association.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Breiman, L. & J.H. Friedman (1985) Estimating optimal transformations for multiple regression and correlation (with discussion). *Journal of the American Statistical Association* 80, 580–619.

Buja, A., T. Hastie, & R. Tibshirani (1989) Linear smoothers and additive models (with discussion). *Annals of Statistics* 17, 453–555.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.

Engle, R.F. & V.K. Ng, (1993) Measuring and testing the impact of news on volatility. *The Journal of Finance* 48(5), 1749–1778.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Hastie, T. & R. Tibshirani (1987) Generalized additive models: Some applications. *Journal of the American Statistical Association* 82, 371–386.

Hastie, T. & R. Tibshirani (1990) *Generalized Additive Models*. Chapman & Hall.

Kim, W. & O.B. Linton (2004) The LIVE method for generalized additive volatility models. *Econometric Theory* 20(6), 1094–1139.

Li, Q., C. Hsiao, & J. Zinn (2003) Consistent specification test for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.

Linton, O.B. & J. Nielsen (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.

Lu, Z. & Z. Jiang (2001) L1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. *Statistics & Probability Letters* 51, 121–130.

Masry, E. (1996a) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.

Masry, E. (1996b) Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81–101.

Newey, W.K. (1994) Kernel estimation of partial means. *Econometric Theory* 10, 233–253.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

Pagan, A.R. & Y.S. Hong (1991) Nonparametric estimation and the risk premium. In W. Barnett, J. Powell, & G.E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Pagan, A.R. & G.W. Schwert (1990) Alternative models for conditional stock volatility. *Journal of Econometrics* 45, 267–290.

Rio, E. (1993) Covariance inequalities for strongly mixing processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques* 29(4), 587–597.

Sentana, E. (1995) Quadratic ARCH models. *Review of Economic Studies* 62, 639–661.

Sperlich, S., D. Tjostheim, & L. Yang (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* 18, 197–251.

Stone, C.J. (1994) The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Annals of Statistics* 22, 118–184.

Tjostheim, D. & B. Auestad (1994) Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association* 89, 1398–1409.

Wand, M.P. & M.C. Jones (1995) *Kernel Smoothing*. Chapman & Hall.

Yang, L., W. Härdle, & J.P. Nielsen (1999) Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* 20, 579–604.

APPENDIX A: Simulation

This Appendix is dedicated to a simulation study that illustrates the finite-sample behavior of the LIVE $\hat{v}_\alpha(y_\alpha)$ and $\hat{v}_{\alpha\beta}(y_\alpha, y_\beta)$. For simplicity, we consider examples with zero conditional mean function only.

Example A1
We define the process

$$y_t = \sqrt{4 + v_1(y_{t-1}) + v_2(y_{t-2}) + v_3(y_{t-3}) + v_{12}(y_{t-1}, y_{t-2}) + v_{13}(y_{t-1}, y_{t-3}) + v_{23}(y_{t-2}, y_{t-3})}\varepsilon_t$$

where $v_1(u) = v_2(u) = -v_3(u) = 0.5\sin(u)$,
$v_{12}(u,v) = v_{13}(u,v) = v_{23}(u,v) = 0.5\arctan(u)\arctan(v)$, ε_t independent of $\mathcal{F}_{t-1}$ and $\varepsilon_t \sim Uniform(-\sqrt{3}, \sqrt{3})$.

The components of the volatility function are selected to satisfy conditions of Lu and Jiang (2001) to ensure the geometric ergodicity (and, therefore, the strict stationarity) of the process y_t. Condition (B1) of Lu and Jiang (2001) reduces in the one-dimensional case to the requirement that the growth rate in each coordinate must not exceed the linear one; apparently, the product of arctan functions satisfies this condition. Based on this model, we simulate 500 samples with sample size $n = 500$. For each realization of the ARCH process, we apply the instrumental variable estimation procedure from Section 3.2 to obtain estimates of $v_\alpha(\cdot)$ and $v_{\alpha\beta}(\cdot)$, $1 \le \alpha, \beta \le 3$. Gaussian kernels (univariate or product multivariate as needed) are used for all of the nonparametric estimates. Bandwidth g is selected according to the Gaussian rule of thumb as $g = cn^{-1/(l+4)}$, where $c = (\frac{4}{l+2})^{1/(l+4)}$ and l is the number of dimensions of the function to be estimated. For example, $l = 1$ for any one-dimensional marginal density, $l = 2$ for $\hat{p}_{\alpha\beta}(y_\alpha, y_\beta)$, $l = d$ for $\hat{p}(\mathbf{y})$, etc. (for details, see, e.g., Wand and Jones, 1995). The constant c is selected to ensure that the bandwidth g is asymptotically optimal in the mean squared error (MSE) sense under the assumption that the true density is Gaussian. The same rule is used to select the second bandwidth h, where $l = 1$ or $l = 2$ for additive and interactive components, respectively. To evaluate the

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

performance of the estimators, the MSE and the mean absolute deviation error (MAE) are computed for each simulated sample. They are defined as

$$MSE(v_\alpha) = \left\{ \frac{1}{101} \sum_{j=1}^{101} [v_\alpha(x_j) - \hat{v}_\alpha(x_j)]^2 \right\}^{1/2},$$

$$MSE(v_{\alpha\beta}) = \left\{ \frac{1}{2601} \sum_{j=1}^{2601} [v_{\alpha\beta}(x_{j,1}, x_{j,2}) - \hat{v}_{\alpha\beta}(x_{j,1}, x_{j,2})]^2 \right\}^{1/2},$$

$$MAE(v_\alpha) = \frac{1}{101} \sum_{j=1}^{101} |v_\alpha(x_j) - \hat{v}_\alpha(x_j)|,$$

$$MAE(v_{\alpha\beta}) = \frac{1}{2601} \sum_{j=1}^{2601} |v_{\alpha\beta}(x_{j,1}, x_{j,2}) - \hat{v}_{\alpha\beta}(x_{j,1}, x_{j,2})|,$$

TABLE A1. Averaged MSE and MAE for six volatility estimators in the uniform distribution case

	v_1	v_2	v_3	v_{12}	v_{13}	v_{23}
MSE	0.33	0.31	0.30	0.54	0.56	0.53
MAE	0.25	0.24	0.24	0.36	0.37	0.36

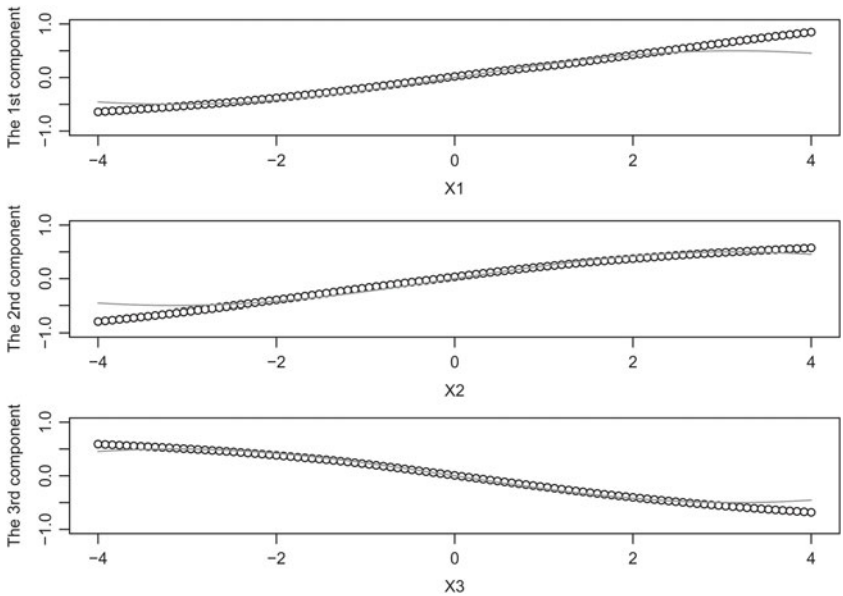


FIGURE A1. Estimates of three additive components in Example A1.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

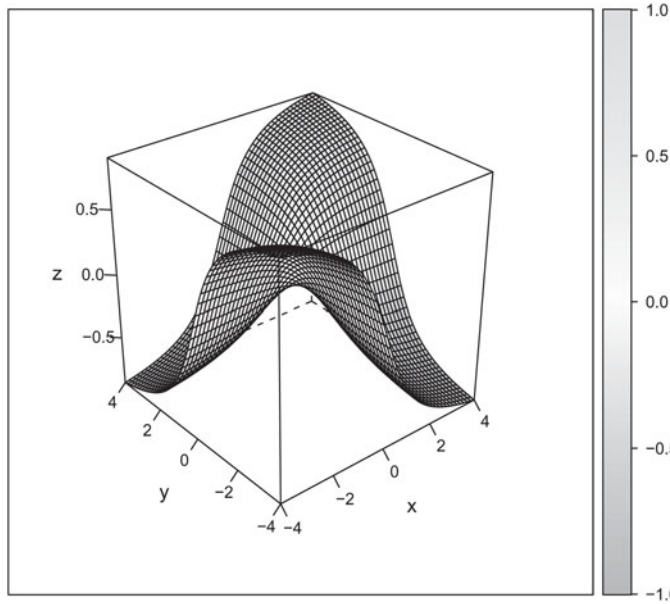


FIGURE A2. True surface of the interactive components in Example A1.

$1 \le \alpha < \beta \le 3$. In the above, $\{x_j\}_{j=1}^{101}$ is an equispaced grid on $[-4,4]$, and $\{x_{j,1}\}_{j=1}^{51} \times \{x_{j,2}\}_{j=1}^{51}$ is an equispaced grid on $[-4,4] \times [-4,4]$. The grid range covers more than 99% of all observations in both one and two dimensions, so very little information is lost. Table A1 gives averages of MSEs and MAEs for all six components from 500 repetitions. Figure A1 shows the averaged estimates of three additive components as well as the true functions; the solid lines are the true curves and the dotted ones are the estimates averaged over 500 repetitions. Figures A2–A5 show the true surface of the interactive components next to the estimates averaged over 500 simulations. In general, the results show a very good fit for both additive and interactive components; among interactive components, $v_{12}(\cdot)$ and $v_{23}(\cdot)$ seem to have been fit particularly well. It can be clearly seen that the use of local polynomial regression eliminated boundary effects to a great extent in both additive and interactive component estimation.

Example A2
Again, consider the model

$$y_t = \sqrt{4 + v_1(y_{t-1}) + v_2(y_{t-2}) + v_3(y_{t-3}) + v_{12}(y_{t-1}, y_{t-2}) + v_{13}(y_{t-1}, y_{t-3}) + v_{23}(y_{t-2}, y_{t-3})}\varepsilon_t,$$

where
$v_1(u) = v_2(u) = -v_3(u) = 0.5\sin(u)$,
$v_{12}(u,v) = v_{13}(u,v) = v_{23}(u,v) = 0.5\arctan(u)\arctan(v)$,
ε_t independent of $\mathcal{F}_{t-1}$ and $\varepsilon_t \sim N(0,1)$.

The previous example has the finitely supported error (innovation) distribution, and this may not be realistic enough in practice. Therefore, we are interested in testing the

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

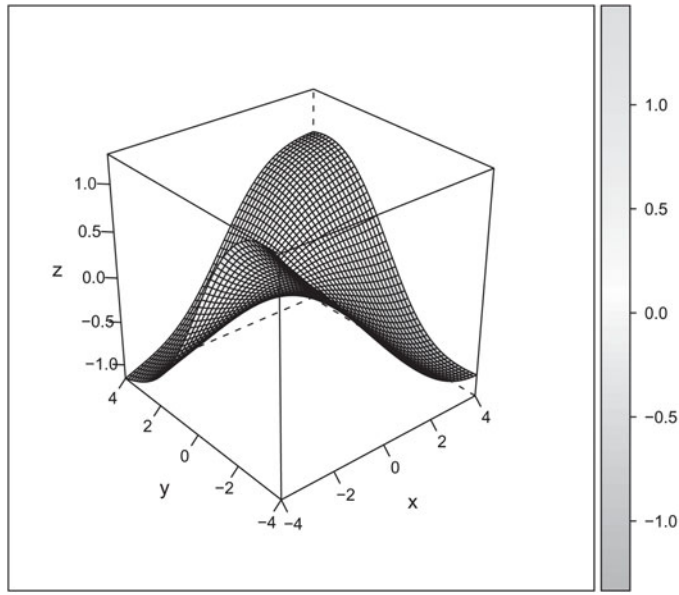


FIGURE A3. Estimated surface of $v_{12}(\cdot)$ in Example A1.

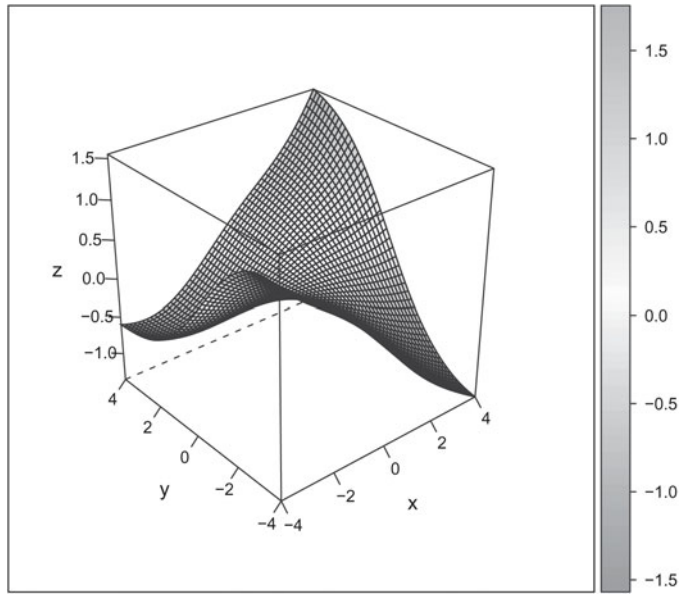


FIGURE A4. Estimated surface of $v_{13}(\cdot)$ in Example A1.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

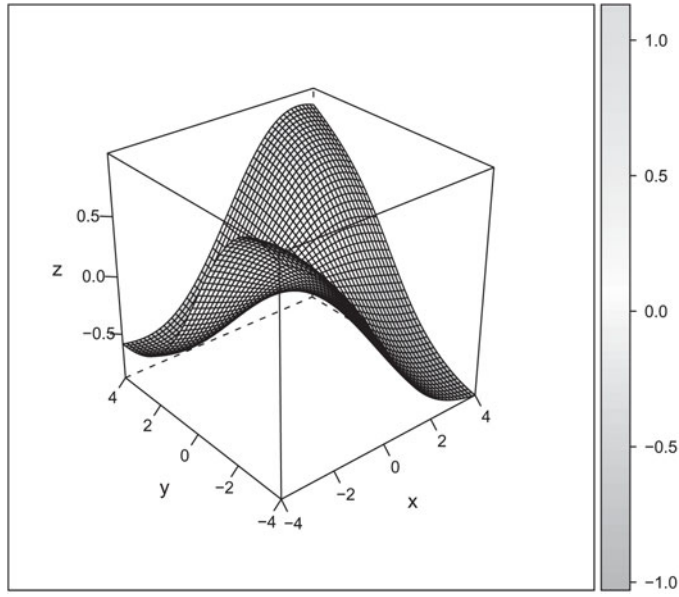


FIGURE A5. Estimated surface of $v_{23}(\cdot)$ in Example A1.

TABLE A2. Averaged MSE and MAE for six volatility estimators in the normal distribution case

	v_1	v_2	v_3	v_{12}	v_{13}	v_{23}
MSE	0.40	0.40	0.39	0.41	0.40	0.42
MAE	0.32	0.32	0.32	0.31	0.30	0.31

performance of the method in the case where the innovation distribution does not have a compact support. Obviously, the most intuitive choice is the standard normal distribution. The grid ranges we chose are $[-3.2, 3.2]$ in one-dimensional regressions and $[-3.2, 3.2] \times [-3.2, 3.2]$ in two-dimensional ones. These ranges cover approximately 90% and 80% of all observations, respectively. The averages of MSEs and MAEs for all six components from 500 repetitions are shown in Table A2. Note that the performance of the method does not seem to be any worse compared to the previous example. While additive components seem to be estimated with slightly less precision, the opposite is true when it comes to interactive components for either choice of the loss function. The averaged estimates of six volatility components as well as the true ones are presented in Figures A6–A10.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

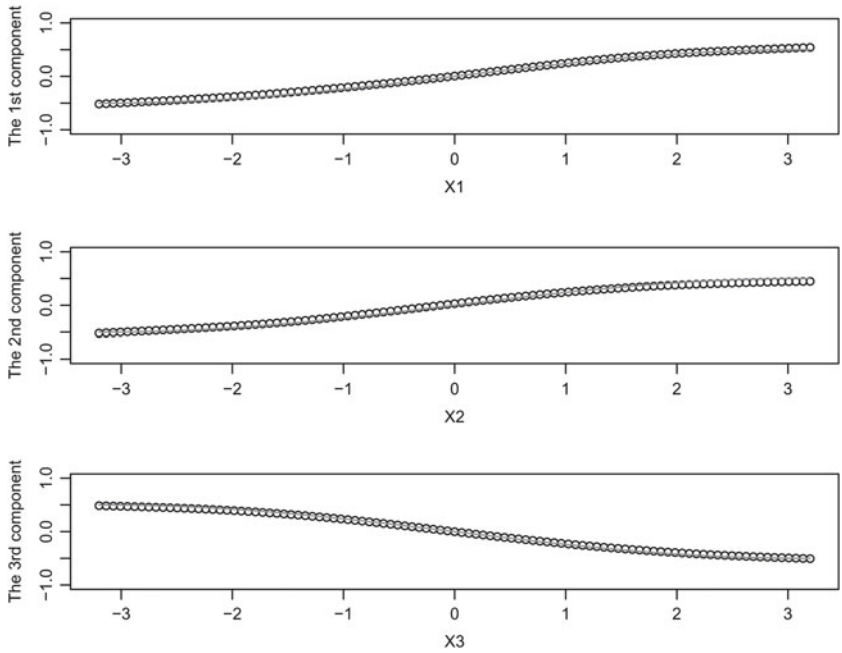


FIGURE A6. Estimates of three additive components in Example A2.

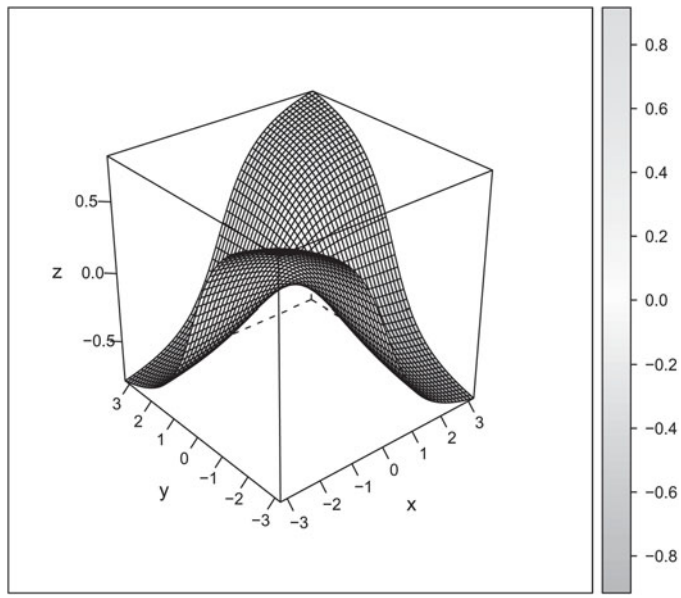


FIGURE A7. True surface of the interactive components in Example A2.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

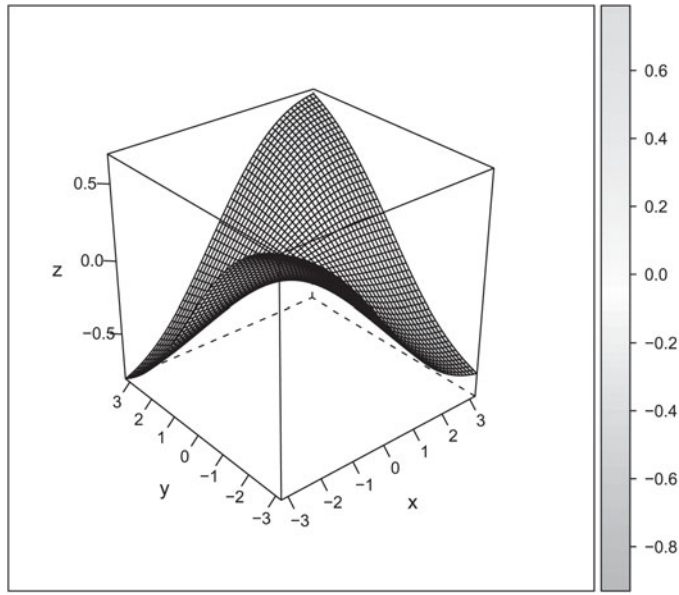


FIGURE A8. Estimated surface of $v_{12}(\cdot)$ in Example A2.

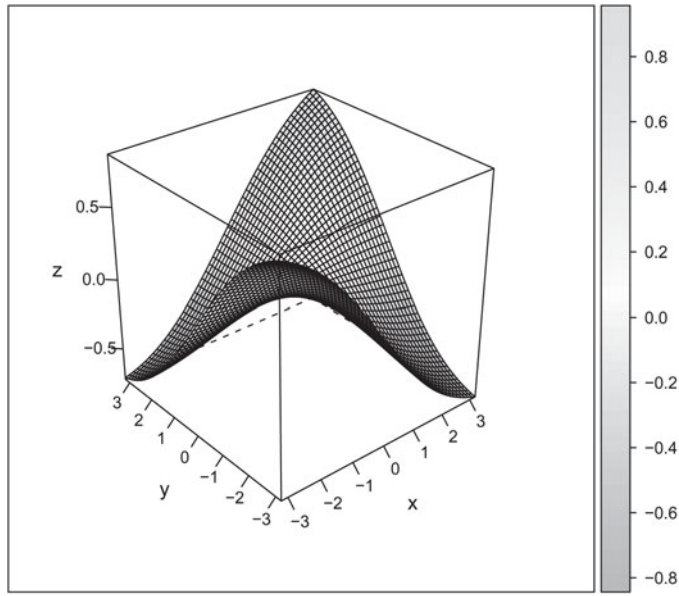


FIGURE A9. Estimated surface of $v_{13}(\cdot)$ in Example A2.

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

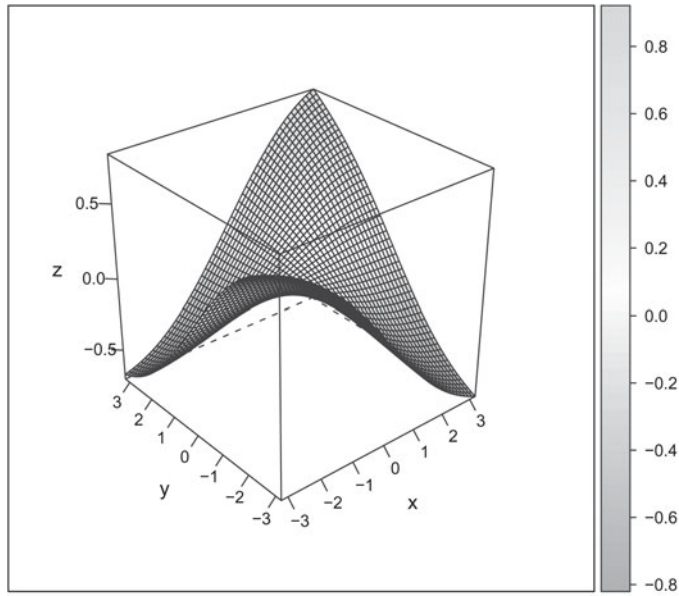


FIGURE A10. Estimated surface of $\upsilon_{23}(\cdot)$ in Example A2.

APPENDIX B: Proof of Theorem 2

Proof. We prove Theorem 2 using the same four steps as in the proof of Theorem 1. Without loss of generality, we prove it for the case of $\alpha = 1,\ \beta = 2$. For expositional convenience we use the subscript 3 to denote the nuisance directions.

Step I: Decompositions. The closed form of the LIVE for $\phi_{12}(y_1, y_2)$ is

$$\hat{\phi}_{12}(y_1, y_2)^T = e_1^T (\mathbf{Y}_-^T \mathbf{K} \mathbf{Y}_-)^{-1} \mathbf{Y}_-^T \mathbf{K} \mathbf{W} \tilde{\mathbf{R}}, \tag{B.1}$$

where

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{Y}_- = \begin{pmatrix} 1 & y_{d+1-1} - y_1 & y_{d+1-2} - y_2 \\ 1 & y_{d+2-1} - y_1 & y_{d+2-2} - y_2 \\ \vdots & \vdots & \vdots \\ 1 & y_{d+n-1} - y_1 & y_{d+n-2} - y_2 \end{pmatrix}, \quad \tilde{\mathbf{R}} = \begin{pmatrix} \tilde{r}_1 \\ \tilde{r}_2 \\ \vdots \\ \tilde{r}_n \end{pmatrix} = \begin{pmatrix} \tilde{y}_{d+1} & \tilde{\varepsilon}_{d+1}^2 \\ \tilde{y}_{d+2} & \tilde{\varepsilon}_{d+2}^2 \\ \vdots & \vdots \\ \tilde{y}_{d+n} & \tilde{\varepsilon}_{d+n}^2 \end{pmatrix},$$

$\mathbf{K} = diag\{K_h(y_{d+l-1} - y_1) K_h(y_{d+m-1} - y_1)\},\ l, m = 1, \ldots, n,\ \mathbf{W} = diag\{\hat{W}_{d+k}\}$, $k = 1, \ldots, n$, and $\tilde{\varepsilon}_t^2 = (y_t - \tilde{m}(\mathbf{y}_t))^2$, $\tilde{y}_t = y_t - \sum_{\alpha=1}^d \hat{m}_\alpha(y_{t-\alpha})$, and $\tilde{\varepsilon}_t^2 = (y_t - \tilde{m}(\mathbf{y}_t))^2 - \sum_{\alpha=1}^d \hat{\upsilon}_\alpha(y_{t-\alpha})$. The preliminary density estimators used are

$$\hat{p}_{12}(y_1, y_2) = \frac{1}{ng^2} \sum_{t=d+1}^{d+n} L_g(y_{t-1} - y_1) L_g(y_{t-2} - y_2),$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$\hat{p}_3(\mathbf{y}_{\underline{12}}) = \frac{1}{ng^{d-2}} \sum_{t=d+1}^{d+n} \prod_{\alpha=3}^{d} L_g(y_{t-\alpha} - y_\alpha),$$

$$\hat{p}(\mathbf{y}) = \frac{1}{ng^{d}} \sum_{t=d+1}^{d+n} \prod_{\alpha=1}^{d} L_g(y_{t-\alpha} - y_\alpha),$$

$$\hat{W}_t = \frac{\hat{p}_{12}(y_{t-1}, y_{t-2})\hat{p}_3(\mathbf{y}_{t,\underline{12}})}{\hat{p}(\mathbf{y}_t)}.$$

In exactly the same way as in Theorem 1, we can decompose the estimation error as

$$\hat{\phi}_{12}(y_1, y_2) - \phi_{12}(y_1, y_2) = [I_2 \otimes e_1^T Q_n^{-1}]\tau_n, \tag{B.2}$$

where

$$Q_n = \frac{1}{n} D_h^{-1} \mathbf{Y}_-^T \mathbf{K} \mathbf{Y}_- D_h^{-1}, \tag{B.3}$$

$$\tau_n = \frac{1}{n} [I_2 \otimes D_h^{-1} \mathbf{Y}_-^T \mathbf{K}] vec\left[\mathbf{W}\tilde{\mathbf{R}} - l\phi_{12}^T(y_1, y_2) - Y_- \bigtriangledown \phi_{12}^T(y_1, y_2)\right], \tag{B.4}$$

where $D_h = diag\{1,\ h,\ h\}$.

Step II: Approximations. τ_n from (B.4) can be represented as

$$\tau_n = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1) K_h(y_{k-2} - y_2) e_k \otimes \left(1, \frac{y_{k-1} - y_1}{h}, \frac{y_{k-2} - y_2}{h}\right)^T \right\},$$

where

$$e_k = \hat{W}_k \tilde{r}_k - \phi_{12}(y_1, y_2) - (y_{k-1} - y_1)\frac{\partial \phi_{12}}{\partial y_{t-1}}(y_1, y_2) - (y_{k-2} - y_2)\frac{\partial \phi_{12}}{\partial y_{t-2}}(y_1, y_2).$$

Denote $K_h(y_{k-1} - y_1)K_h(y_{k-2} - y_2)$ by $KK_h(k)$, and $\left(1, \frac{y_{k-1}-y_1}{h}, \frac{y_{k-2}-y_2}{h}\right)^T$ by $X(k)$; then τ_n can be represented as

$$\tau_n = \frac{1}{n} \sum_{k=d+1}^{d+n} \{KK_h(k)[\hat{W}_k \tilde{r}_k - \phi_{12}(y_{k-1}, y_{k-2})] \otimes X(k)\} \tag{B.5}$$

$$+ \frac{1}{n} \sum_{k=d+1}^{d+n} \{KK_h(k) e_k^* \otimes X(k)\}, \tag{B.6}$$

where

$$e_k^* = \phi_{12}(y_{k-1}, y_{k-2}) - \phi_{12}(y_1, y_2) - (y_{k-1} - y_1)\frac{\partial \phi_{12}}{\partial y_{t-1}}(y_1, y_2)$$

$$- (y_{k-2} - y_2)\frac{\partial \phi_{12}}{\partial y_{t-2}}(y_1, y_2).$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

Define $r_k = r(\mathbf{y}_k) = \sum_{1\le\alpha<\beta\le d} \left(m_{\alpha\beta}(y_{k-\alpha}, y_{k-\beta}),\ v_{\alpha\beta}(y_{k-\alpha}, y_{k-\beta})\right)^T = (r_k^m,\ r_k^v)^T$, Let us denote

$$\tilde{r}_k^* = \begin{pmatrix} \sum_{\alpha=1}^d [\hat{m}_\alpha(y_{k-\alpha}) - m_\alpha(y_{k-\alpha})] + v^{1/2}(\mathbf{y}_k)\varepsilon_k \\ \sum_{\alpha=1}^d [\hat{v}_\alpha(y_{k-\alpha}) - v_\alpha(y_{k-\alpha})] + v(\mathbf{y}_k)(\varepsilon_k^2 - 1) \end{pmatrix},$$

$$D_Y^2 = \left(\frac{y_{k-1}-y_1}{h}\right)^2 \frac{\partial^2\phi_{12}}{\partial y_{t-1}^2}(y_1, y_2) + \left(\frac{y_{k-2}-y_2}{h}\right)^2 \frac{\partial^2\phi_{12}}{\partial y_{t-2}^2}(y_1, y_2)$$

$$+ 2\frac{(y_{k-1}-y_1)(y_{k-2}-y_2)}{h^2} \frac{\partial^2\phi_{12}}{\partial y_{t-1}\partial y_{t-2}}(y_1, y_2)$$

$$\equiv (D_Y^{2,m},\ D_Y^{2,v})^T.$$

The two components of τ_n can now be approximated in the same way as it was done for the two terms of (11) in the proof of Theorem 1; specifically, (B.5) becomes

$$\frac{1}{n}\sum_{k=d+1}^{d+n} \left\{ KK_h(k) \left[\frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} r_k - \phi_{12}(y_{k-1}, y_{k-2}) \right] \otimes X(k) \right\}$$

$$+ \frac{1}{n}\sum_{k=d+1}^{d+n} \left\{ KK_h(k) \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \tilde{r}_k^* \otimes X(k) \right\}$$

$$+ \frac{1}{n}\sum_{k=d+1}^{d+n} \left\{ KK_h(k) \left[\frac{\hat{p}_{12}(y_{k-1}, y_{k-2}) \hat{p}_3(\mathbf{y}_{k,\underline{12}})}{\hat{p}(\mathbf{y}_k)} \right.\right.$$

$$\left.\left. - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \otimes X(k) \right\}$$

$$+ o_p(1/\sqrt{nh^2}),$$

while (B.6) becomes $\frac{h^2}{2n}\sum_{k=d+1}^{d+n} \{KK_h(k) D_Y^2 \otimes X(k)\} + o_p(1/\sqrt{nh^2})$. Thus, the four components that need to be analyzed in order to obtain the asymptotic distribution and the bias of the estimator $\hat{\phi}_{12}(y_1, y_2)$ are

$$R_{1n} = \frac{h^2}{2n}\sum_{k=d+1}^{d+n} \{KK_h(k) D_Y^2 \otimes X(k)\},$$

$$R_{2n} = \frac{1}{n}\sum_{k=d+1}^{d+n} \left\{ KK_h(k) \left[\frac{\hat{p}_{12}(y_{k-1}, y_{k-2}) \hat{p}_3(\mathbf{y}_{k,\underline{12}})}{\hat{p}(\mathbf{y}_k)} \right.\right.$$

$$\left.\left. - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \otimes X(k) \right\},$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$R_{3n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{KK_h(k)\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\tilde{r}_k^*\otimes X(k)\right\},$$

$$R_{4n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{KK_h(k)\left[\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}r_k - \phi_{12}(y_{k-1},y_{k-2})\right]\otimes X(k)\right\}.$$

Step III: Asymptotic Bias. Applying ergodic theorem for α-mixing stationary processes to R_{1n} and using the Taylor expansion of $p_{12}(\cdot,\cdot)$, we obtain

$$\begin{aligned}
R_{1n} &= \frac{h^2}{2n}\sum_{k=d+1}^{d+n}\{KK_h(k)D_Y^2\otimes X(k)\}\\
&\xrightarrow{p}\frac{h^2}{2}\mathrm{E}\left[K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\left(\frac{y_{k-1}-y_1}{h}\right)^2\frac{\partial^2\phi_{12}}{\partial y_{t-1}^2}(y_1,y_2)\right.\\
&\qquad\left.\otimes\left(1,\frac{y_{k-1}-y_1}{h},\frac{y_{k-2}-y_2}{h}\right)^T\right]\\
&\quad+\frac{h^2}{2}\mathrm{E}\left[K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\left(\frac{y_{k-2}-y_2}{h}\right)^2\frac{\partial^2\phi_{12}}{\partial y_{t-2}^2}(y_1,y_2)\right.\\
&\qquad\left.\otimes\left(1,\frac{y_{k-1}-y_1}{h},\frac{y_{k-2}-y_2}{h}\right)^T\right]\\
&\quad+\frac{h^2}{2}\mathrm{E}\left[K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\frac{2(y_{k-1}-y_1)(y_{k-2}-y_2)}{h^2}\right.\\
&\qquad\left.\frac{\partial^2\phi_{12}}{\partial y_{t-1}\partial y_{t-2}}(y_1,y_2)\otimes\left(1,\frac{y_{k-1}-y_1}{h},\frac{y_{k-2}-y_2}{h}\right)^T\right]\\
&=\frac{h^2}{2}p_{12}(y_1,y_2)\left[\frac{\partial^2\phi_{12}}{\partial y_{t-1}^2}(y_1,y_2)\otimes\left(\mu_K^2,\ \mu_K^3,0\right)^T\right.\\
&\qquad\left.+\frac{\partial^2\phi_{12}}{\partial y_{t-2}^2}(y_1,y_2)\otimes\left(\mu_K^2,\ 0,\ \mu_K^3\right)^T\right]+o(h^2).
\end{aligned}$$

In the same way as before, we denote marginal expectations of estimated density functions $\hat{p}_{12}(\cdot)$, $\hat{p}_3(\cdot)$, and $\hat{p}(\cdot)$,

$$\bar{p}_{12}(y_{k-1},y_{k-2}) = \int L_g(z_1-y_{k-1})L_g(z_2-y_{k-2})p_{12}(z_1,z_2)dz_1dz_2,$$

$$\bar{p}_3(\mathbf{y}_{k,\underline{12}}) = \int L_g(z_3-\mathbf{y}_{k,\underline{12}})p_3(z_3)dz_3,$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$\bar{p}(\mathbf{y}_k) = \int L_g(z_1 - y_{k-1}) L_g(z_2 - y_{k-2}) L_g(z_3 - \mathbf{y}_{k,\underline{12}}) p(z_1, z_2, z_3) dz_1 dz_2 dz_3.$$

Using the fact that

$$\frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K K_h(k) \left[\frac{\hat{p}_{12}(y_{k-1}, y_{k-2}) \hat{p}_3(\mathbf{y}_{k,\underline{12}})}{\hat{p}(\mathbf{y}_k)} - \frac{\bar{p}_{12}(y_{k-1}, y_{k-2}) \bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)} \right] \right.$$

$$\left. \tilde{r}_k \otimes X(k) \right\} = o_p\left(1/\sqrt{nh^2}\right).$$

we can approximate R_{2n} as

$$R_{2n} = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K K_h(k) \left[\frac{\hat{p}_{12}(y_{k-1}, y_{k-2}) \hat{p}_3(\mathbf{y}_{k,\underline{12}})}{\hat{p}(\mathbf{y}_k)} - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \right.$$

$$\left. \otimes X(k) \right\}$$

$$= \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K K_h(k) \left[\frac{\bar{p}_{12}(y_{k-1}, y_{k-2}) \bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \right.$$

$$\left. \otimes X(k) \right\} + o_p\left(1/\sqrt{nh^2}\right).$$

Next, we use yet again the uniform convergence result from Masry (1996a) in combination with the ergodic theorem for α-mixing stationary processes from Fan and Yao (2003); this allows us to conclude that $K K_h(k) \left[\frac{\bar{p}_{12}(y_{k-1}, y_{k-2}) \bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] [\tilde{m}(\mathbf{y}_k) - m(\mathbf{y}_k)]^2 \otimes X(k) = o_p(1/\sqrt{nh^2})$ and the left-hand side has finite expectation. Thus,

$$R_{2n} = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K K_h(k) \left[\frac{\bar{p}_{12}(y_{k-1}, y_{k-2}) \bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \right.$$

$$\left. \otimes X(k) \right\} + o_p\left(1/\sqrt{nh^2}\right)$$

$$\xrightarrow{p} \mathrm{E}\left[K K_h(k) \left[\frac{\bar{p}_{12}(y_{k-1}, y_{k-2}) \bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \right.$$

$$\left. \otimes X(k) \right]$$

$$= \mathrm{E}\left[\mathrm{E}\left[K K_h(k) \left[\frac{\bar{p}_{12}(y_{k-1}, y_{k-2}) \bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_{12}(y_{k-1}, y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \right] \tilde{r}_k \right.\right.$$

$$\left.\left. \otimes X(k) | \mathcal{F}_k \right] \right]$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$
\begin{aligned}
&\simeq \mathrm{E}\left[K K_h(k)\left[\frac{\bar{p}_{12}(y_{k-1},y_{k-2})\bar{p}_3(\mathbf{y}_{k,\underline{12}})}{\bar{p}(\mathbf{y}_k)}-\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\right]\right.\\
&\quad\cdot\left(\sum_{\alpha=1}^{d}[m_\alpha(y_{k-\alpha})-\hat{m}_\alpha(y_{k-\alpha})]+r_k^m,\ \sum_{\alpha=1}^{d}[v_\alpha(y_{k-\alpha})-\hat{v}_\alpha(y_{k-\alpha})]+r_k^v\right)^T\\
&\quad\left.\otimes X(k)\right]\\
&\simeq -\int K_h(z_1-y_1)K_h(z_2-y_2)\left[\frac{\bar{p}_{12}(z_1,z_2)\bar{p}_3(z_3)}{p(z)}-\frac{p_{12}(z_1,z_2)p_3(z_3)\bar{p}(z)}{p^2(z)}\right]\\
&\quad\times\left(\sum_{\alpha=1}^{d}b_\alpha^m(z_\alpha),\sum_{\alpha=1}^{d}b_\alpha^v(z_\alpha)\right)^T\otimes\left(1,\frac{z_1-y_1}{h},\frac{z_2-y_2}{h}\right)^T p(z)dz\\
&\quad+\int K_h(z_1-y_1)K_h(z_2-y_2)\left[\frac{\bar{p}_{12}(z_1,z_2)\bar{p}_3(z_3)}{p(z)}-\frac{p_{12}(z_1,z_2)p_3(z_3)\bar{p}(z)}{p^2(z)}\right]r(z)\\
&\quad\otimes\left(1,\frac{z_1-y_1}{h},\frac{z_2-y_2}{h}\right)^T p(z)dz\\
&\simeq -\int K_h(z_1-y_1)K_h(z_2-y_2)\left[p_{12}(z_1,z_2)[\bar{p}_3(z_3)-p_3(z_3)]\right.\\
&\qquad\qquad+p_3(z_3)[\bar{p}_{12}(z_1,z_2)-p_{12}(z_1,z_2)]\\
&\qquad\qquad\left.-\frac{p_{12}(z_1,z_2)p_3(z_3)}{p(z)}[\bar{p}(z)-p(z)]\right]\\
&\quad\times\left(\sum_{\alpha=1}^{d}b_\alpha^m(z_\alpha),\sum_{\alpha=1}^{d}b_\alpha^v(z_\alpha)\right)^T\otimes\left(1,\frac{z_1-y_1}{h},\frac{z_2-y_2}{h}\right)^T dz\\
&+\int K_h(z_1-y_1)K_h(z_2-y_2)\left[p_{12}(z_1,z_2)[\bar{p}_3(z_3)-p_3(z_3)]+p_3(z_3)[\bar{p}_{12}(z_1,z_2)\right.\\
&\qquad\qquad\left.-p_{12}(z_1,z_2)]-\frac{p_{12}(z_1,z_2)p_3(z_3)}{p(z)}[\bar{p}(z)-p(z)]\right]r(z)\\
&\qquad\qquad\otimes\left(1,\frac{z_1-y_1}{h},\frac{z_2-y_2}{h}\right)^T dz\\
&\qquad\qquad=\frac{g^2}{2}p_{12}(y_1,y_2)\\
&\times\int\left[p_3^{(2)}(z_3)+\frac{p_{12}^{(2)}(y_1,y_2)}{p_{12}(y_1,y_2)}p_3(z_3)-\frac{p_3(z_3)}{p(y_1,y_2,z_3)}p^{(2)}(y_1,y_2,z_3)\right]\\
&\Delta_{\underline{12}}(y_1,y_2,z_3)dz_3\otimes(\mu_L^2,0,0)^T,
\end{aligned}
$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

where

$$\Delta_{\underline{12}}(y_1, y_2, z_3) = \begin{pmatrix} m_{12}(y_1, y_2) - b_1^m(y_1) - b_2^m(y_2) + \sum_{3\le\alpha<\beta\le d} m_{\alpha\beta}(z_\alpha, z_\beta) - \sum_{\alpha=3}^{d} b_\alpha^m(z_\alpha) \\ v_{12}(y_1, y_2) - b_1^v(y_1) - b_2^v(y_2) + \sum_{3\le\alpha<\beta\le d} v_{\alpha\beta}(z_\alpha, z_\beta) - \sum_{\alpha=3}^{d} b_\alpha^v(z_\alpha) \end{pmatrix}.$$

The term R_{3n} is easy to analyze; again, we apply the ergodic theorem for α-mixing stationary processes on R_{3n} to obtain

$$R_{3n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K K_h(k)\frac{p_{12}(y_{k-1}, y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\tilde{r}_k^* \otimes X(k)\right\}$$

$$\xrightarrow{p} \mathrm{E}\left[K K_h(k)\frac{p_{12}(y_{k-1}, y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\tilde{r}_k^* \otimes X(k)\right]$$

$$= p_{12}(y_1, y_2)\int p_3(z_3)\Delta_{12}(y_1, y_2, z_3)dz_3 \otimes (1, 0, 0)^T,$$

where

$$\Delta_{12}(y_1, y_2, z_3) = \begin{pmatrix} b_1^m(y_1) + b_2^m(y_2) + \sum_{\alpha=3}^{d} b_\alpha^m(z_\alpha) \\ b_1^v(y_1) + b_2^v(y_2) + \sum_{\alpha=3}^{d} b_\alpha^v(z_\alpha) \end{pmatrix}.$$

Finally, it is easy to realize that the term R_{4n} does not contribute to the asymptotic bias since it converges in probability to zero due to the same ergodic theorem,

$$R_{4n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K K_h(k)\left[\frac{p_{12}(y_{k-1}, y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}r_k - \phi_{12}(y_{k-1}, y_{k-2})\right] \otimes X(k)\right\}$$

$$\xrightarrow{p} (0\,0\,0\,0\,0\,0)^T.$$

The analysis of $[I_2 \otimes e_1^T Q_n^{-1}]$ is almost exactly the same as in Theorem 1; it is easy to show that

$$Q_n \xrightarrow{p} \begin{pmatrix} p_{12}(y_1, y_2) & 0 & 0 \\ 0 & p_{12}(y_1, y_2)\mu_K^2 & 0 \\ 0 & 0 & p_{12}(y_1, y_2)\mu_K^2 \end{pmatrix}$$

$$Q_n^{-1} \xrightarrow{p} \begin{pmatrix} \frac{1}{p_{12}(y_1, y_2)} & 0 & 0 \\ 0 & \frac{1}{p_{12}(y_1, y_2)\mu_K^2} & 0 \\ 0 & 0 & \frac{1}{p_{12}(y_1, y_2)\mu_K^2} \end{pmatrix}$$

$$e_1^T Q_n^{-1} \xrightarrow{p} \frac{1}{p_{12}(y_1, y_2)} e_1^T.$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

Therefore, using Slutzky's theorem we conclude that the bias of our estimator is

$$
\begin{aligned}
B_{12,n}(y_1,y_2) &= [I_2 \otimes e_1^T Q_n^{-1}](R_{1n}+R_{2n}+R_{3n}+R_{4n}) \\
&\xrightarrow{p} \left[I_2 \otimes \left(\frac{1}{p_{12}(y_1,y_2)} \;\; 0 \;\; 0\right)\right] \frac{h^2}{2} p_{12}(y_1,y_2) \left[\frac{\partial^2 \phi_{12}}{\partial y_{t-1}^2}(y_1,y_2)\right. \\
&\otimes \left(\mu_K^2,\ \mu_K^3, 0\right)^T + \left.\frac{\partial^2 \phi_{12}}{\partial y_{t-2}^2}(y_1,y_2) \otimes \left(\mu_K^2,\ 0,\ \mu_K^3\right)^T\right] \\
&+ \left[I_2 \otimes \left(\frac{1}{p_{12}(y_1,y_2)} \;\; 0 \;\; 0\right)\right] \frac{g^2}{2} p_{12}(y_1,y_2) \\
&\cdot \int \left[p_3^{(2)}(z_3) + \frac{p_{12}^{(2)}(y_1,y_2)}{p_{12}(y_1,y_2)} p_3(z_3) - \frac{p_3(z_3)}{p(y_1,y_2,z_3)} p^{(2)}(y_1,y_2,z_3)\right] \\
&\quad \Delta_{\underline{12}}(y_1,y_2,z_3) dz_3 \otimes (\mu_L^2,0,0)^T \\
&+ \left[I_2 \otimes \left(\frac{1}{p_{12}(y_1,y_2)} \;\; 0 \;\; 0\right)\right] p_{12}(y_1,y_2) \\
&\cdot \int p_3(z_3) \Delta_{12}(y_1,y_2,z_3) dz_3 \otimes (1,0,0)^T \\
&= \frac{h^2}{2} \mu_K^2 \left[\frac{\partial^2 \phi_{12}}{\partial y_{t-1}^2}(y_1,y_2) + \frac{\partial^2 \phi_{12}}{\partial y_{t-2}^2}(y_1,y_2)\right] \\
&+ \frac{g^2}{2} \mu_L^2 \int \left[p_3^{(2)}(z_3) + \frac{p_{12}^{(2)}(y_1,y_2)}{p_{12}(y_1,y_2)} p_3(z_3) - \frac{p_3(z_3)}{p(y_1,y_2,z_3)} p^{(2)}(y_1,y_2,z_3)\right] \\
&\quad \Delta_{\underline{12}}(y_1,y_2,z_3) dz_3 \\
&+ \int p_3(z_3) \Delta_{12}(y_1,y_2,z_3) dz_3 \\
&\equiv B_{12}(y_1,y_2).
\end{aligned}
$$

Step IV: Asymptotic Normality. Finally, we derive the asymptotic distribution of $\sqrt{nh^2}[I_2 \otimes e_1^T Q_n^{-1}](R_{1n}+R_{2n}+R_{3n}+R_{4n})$. Using the same technics as the ones used in Theorem 1 to treat the terms T_{1n} and T_{2n}, it is easy to show that both $\text{var}(\sqrt{nh^2}R_{1n})$ and $\text{var}(\sqrt{nh^2}R_{2n})$ are negligible, with orders of $O(h^4)$ and $O\left(\frac{g^{2\nu}}{h^2}\right)$, respectively.

To analyze R_{3n} we expand it as

$$
R_{3n} = \frac{1}{n} \sum_{k=d+1}^{d+n} \left\{ K K_h(k) \frac{p_{12}(y_{k-1},y_{k-2}) p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \tilde{r}_k^* \otimes X(k) \right\}
$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{KK_h(k)\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\tilde{r}^*_{k,1}\otimes X(k)\right\}$$

$$+\frac{1}{n}\sum_{k=d+1}^{d+n}\left\{KK_h(k)\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\tilde{r}^*_{k,2}\otimes X(k)\right\}$$

$$\equiv R_{3n,1}+R_{3n,2},$$

where

$$\tilde{r}^*_{k,1}=\begin{pmatrix}\sum_{\alpha=1}^{d}[\hat{m}_\alpha(y_{k-\alpha})-m_\alpha(y_{k-\alpha})]\\ \sum_{\alpha=1}^{d}[\hat{v}_\alpha(y_{k-\alpha})-v_\alpha(y_{k-\alpha})]\end{pmatrix}$$

$$\tilde{r}^*_{k,2}=\begin{pmatrix}v^{1/2}(\mathbf{y}_k)\varepsilon_k\\ v(\mathbf{y}_k)(\varepsilon_k^2-1)\end{pmatrix}.$$

We claim that $\mathrm{var}(\sqrt{nh^2}R_{3n,1})$ is negligible too. To see why, rewrite $R_{3n,1}$ as

$$R_{3n,1}=(x_{1n},\ x_{2n},\ x_{3n},\ x_{4n},\ x_{5n},\ x_{6n})^T,$$

where

$$x_{1n}=\frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}$$
$$\sum_{\alpha=1}^{d}[\hat{m}_\alpha(y_{k-\alpha})-m_\alpha(y_{k-\alpha})],$$

$$x_{2n}=\frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\left(\frac{y_{k-1}-y_1}{h}\right)$$
$$\cdot\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\sum_{\alpha=1}^{d}[\hat{m}_\alpha(y_{k-\alpha})-m_\alpha(y_{k-\alpha})],$$

$$x_{3n}=\frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\left(\frac{y_{k-2}-y_2}{h}\right)$$
$$\cdot\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}\sum_{\alpha=1}^{d}[\hat{m}_\alpha(y_{k-\alpha})-m_\alpha(y_{k-\alpha})],$$

$$x_{4n}=\frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)}$$
$$\sum_{\alpha=1}^{d}[\hat{v}_\alpha(y_{k-\alpha})-v_\alpha(y_{k-\alpha})],$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

$$x_{5n} = \frac{1}{n}\sum_{k=d+1}^{d+n} K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\left(\frac{y_{k-1}-y_1}{h}\right)$$

$$\cdot \frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \sum_{\alpha=1}^{d}[\hat{v}_\alpha(y_{k-\alpha}) - v_\alpha(y_{k-\alpha})],$$

$$x_{6n} = \frac{1}{n}\sum_{k=d+1}^{d+n} K_h(y_{k-1}-y_1)K_h(y_{k-2}-y_2)\left(\frac{y_{k-2}-y_2}{h}\right)$$

$$\cdot \frac{p_{12}(y_{k-1},y_{k-2})p_3(\mathbf{y}_{k,\underline{12}})}{p(\mathbf{y}_k)} \sum_{\alpha=1}^{d}[\hat{v}_\alpha(y_{k-\alpha}) - v_\alpha(y_{k-\alpha})].$$

We claim that $\mathrm{var}(\sqrt{nh^2}x_{in})$, $i = 1,\ldots,6$ are all negligible, with orders of $O(h^2)$. This is due to Davydov's lemma and the fact that $\hat{m}_\alpha(y_{k-\alpha}) - m_\alpha(y_{k-\alpha}) = O_p(h^2)$, $\hat{v}_\alpha(y_{k-\alpha}) - v_\alpha(y_{k-\alpha}) = O_p(h^2)$ uniformly over all possible values of $y_{k-\alpha}$, with the latter following directly from Theorem 1. Here $\mathrm{var}(\sqrt{nh^2}R_{3n,2})$ can be analyzed in exactly the same way as $\mathrm{var}(\sqrt{nh}T_{3n})$ in Theorem 1; the outcome is that

$$\sqrt{nh^2}R_{3n,2} \xrightarrow{d} N(0, \Sigma_{1,R3}), \tag{B.7}$$

where

$$\Sigma_{1,R3} = \begin{pmatrix} \sigma^m_{1,s}(y_1,y_2) & * & * & \sigma^{mv}_{1,s}(y_1,y_2) & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ \sigma^{mv}_{1,s}(y_1,y_2) & * & * & \sigma^{v}_{1,s}(y_1,y_2) & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix}, \tag{B.8}$$

with

$$\sigma^m_{1,s}(y_1,y_2) = \| K \|_2^2\, p_{12}^2(y_1,y_2)\int \frac{p_3^2(z_3)}{p(y_1,y_2,z_3)} v(y_1,y_2,z_3)dz_3, \tag{B.9}$$

$$\sigma^v_{1,s}(y_1,y_2) = \| K \|_2^2\, p_{12}^2(y_1,y_2)\int \frac{p_3^2(z_3)}{p(y_1,y_2,z_3)} v^2(y_1,y_2,z_3)\kappa_4(y_1,y_2,z_3)dz_3, \tag{B.10}$$

$$\sigma^{mv}_{1,s}(y_1,y_2) = \| K \|_2^2\, p_{12}^2(y_1,y_2)\int \frac{p_3^2(z_3)}{p(y_1,y_2,z_3)} v^{3/2}(y_1,y_2,z_3)\kappa_3(y_1,y_2,z_3)dz_3. \tag{B.11}$$

Also, $\mathrm{var}(\sqrt{nh^2}R_{4n})$ can again be handled in the same way as $\mathrm{var}(\sqrt{nh}T_{4n})$ in Theorem 1; the result is also an asymptotic normality for R_{4n},

$$\sqrt{nh^2}R_{4n} \xrightarrow{d} N(0, \Sigma_{1,R4}), \tag{B.12}$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

where

$$\Sigma_{1,R4} = \begin{pmatrix} \sigma_{1,t}^{m}(y_1,y_2) & * & * & \sigma_{1,t}^{mv}(y_1,y_2) & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ \sigma_{1,t}^{mv}(y_1,y_2) & * & * & \sigma_{1,t}^{v}(y_1,y_2) & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix}, \tag{B.13}$$

with

$$\sigma_{1,t}^{m}(y_1,y_2) = \| K \|_2^2 \, p_{12}^2(y_1,y_2) \int \left[p_3(z_3) m(y_1,y_2,z_3) - p_{3|1,2}(z_3|y_1,y_2) m_{12}(y_1,y_2) \right]^2 dz_3, \tag{B.14}$$

$$\sigma_{1,t}^{v}(y_1,y_2) = \| K \|_2^2 \, p_{12}^2(y_1,y_2) \int \left[p_3(z_3) v(y_1,y_2,z_3) - p_{3|1,2}(z_3|y_1,y_2) v_{12}(y_1,y_2) \right]^2 dz_3, \tag{B.15}$$

$$\sigma_{1,t}^{mv}(y_1,y_2) = \| K \|_2^2 \, p_{12}^2(y_1,y_2) \int \left[p_3(z_3) m(y_1,y_2,z_3) - p_{3|1,2}(z_3|y_1,y_2) m_{12}(y_1,y_2) \right] \cdot \left[p_3(z_3) v(y_1,y_2,z_3) - p_{3|1,2}(z_3|y_1,y_2) v_{12}(y_1,y_2) \right] dz_3. \tag{B.16}$$

In both (B.8) and (B.13) those covariance matrix elements that are not important for the eventual result are replaced with $*$. It is also easy to conclude that the covariance between $R_{3n,2}$ and R_{4n} is 0 since $\mathrm{cov}(R_{3n,2}, R_{4n}) = \mathrm{E}[R_{3n,2} R_{4n}] = \mathrm{E}[\mathrm{E}[R_{3n,2} R_{4n} | \mathcal{F}_{n+d}]] = 0$.

Combining the four terms and using Slutzky's theorem, we can find that

$$\sqrt{nh^2}[I_2 \otimes e_1^T Q_n^{-1}](R_{1n} + R_{2n} + R_{3n} + R_{4n}) \xrightarrow{d} N(0, \Sigma_{12}^*(y_1,y_2)),$$

and, as a consequence,

$$\sqrt{nh^2}[\hat{\phi}_{12}(y_1,y_2) - \phi_{12}(y_1,y_2) - B_{12}(y_1,y_2)] \xrightarrow{d} N[0, \Sigma_{12}^*(y_1,y_2)],$$

where

$$B_{12}(y_1,y_2) = \begin{pmatrix} b_{12}^{m}(y_1,y_2) \\ b_{12}^{v}(y_1,y_2) \end{pmatrix}, \qquad \Sigma_{12}^*(y_1,y_2) = \begin{pmatrix} \sigma_{12}^{m}(y_1,y_2) & \sigma_{12}^{mv}(y_1,y_2) \\ \sigma_{12}^{mv}(y_1,y_2) & \sigma_{12}^{v}(y_1,y_2) \end{pmatrix},$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press

with

$$B_{12}(y_1,y_2)=\frac{h^2}{2}\mu_K^2\left[\frac{\partial^2\phi_{12}}{\partial y_{t-1}^2}(y_1,y_2)+\frac{\partial^2\phi_{12}}{\partial y_{t-2}^2}(y_1,y_2)\right]$$
$$+\frac{g^2}{2}\mu_L^2\int\left[p_3^{(2)}(z_3)+\frac{p_{12}^{(2)}(y_1,y_2)}{p_{12}(y_1,y_2)}p_3(z_3)-\frac{p_3(z_3)}{p(y_1,y_2,z_3)}p^{(2)}(y_1,y_2,z_3)\right]$$
$$\times\Delta_{\underline{12}}(y_1,y_2,z_3)dz_3+\int p_3(z_3)\Delta_{12}(y_1,y_2,z_3)dz_3,$$

$$\sigma_{12}^m(y_1,y_2)=\|K\|_2^2\int\left\{\frac{p_3^2(z_3)}{p(y_1,y_2,z_3)}v(y_1,y_2,z_3)\right.$$
$$\left.+\left[p_3(z_3)m(y_1,y_2,z_3)-p_{3|1,2}(z_3|y_1,y_2)m_{12}(y_1,y_2)\right]^2\right\}dz_3,$$

$$\sigma_{12}^v(y_1,y_2)=\|K\|_2^2\int\left\{\frac{p_3^2(z_3)}{p(y_1,y_2,z_3)}v^2(y_1,y_2,z_3)\kappa_4(y_1,y_2,z_3)\right.$$
$$\left.+\left[p_3(z_3)v(y_1,y_2,z_3)-p_{3|1,2}(z_3|y_1,y_2)v_{12}(y_1,y_2)\right]^2\right\}dz_3,$$

$$\sigma_{12}^{mv}(y_1,y_2)=\|K\|_2^2\int\left[p_3(z_3)m(y_1,y_2,z_3)-p_{3|1,2}(z_3|y_1,y_2)m_{12}(y_1,y_2)\right]$$
$$\times\left[p_3(z_3)v(y_1,y_2,z_3)-p_{3|1,2}(z_3|y_1,y_2)v_{12}(y_1,y_2)\right]$$
$$+\frac{p_3^2(z_3)}{p(y_1,y_2,z_3)}v^{3/2}(y_1,y_2,z_3)\kappa_3(y_1,y_2,z_3)dz_3,$$

and

$$\kappa_3(y_1,y_2,z_3)\equiv\mathrm{E}\left[\varepsilon_k^3|(y_{k-1},y_{k-2},\mathbf{y}_{k,3})=(y_1,y_2,z_3)\right],$$

$$\kappa_4(y_1,y_2,z_3)\equiv\mathrm{E}\left[(\varepsilon_k^2-1)^2|(y_{k-1},y_{k-2},\mathbf{y}_{k,3})=(y_1,y_2,z_3)\right].$$

https://doi.org/10.1017/S0266466611000363 Published online by Cambridge University Press