Econometric Theory, **27**, 2011, 639–661.
doi:10.1017/S0266466610000435

ESTIMATION OF A SEMIPARAMETRIC IGARCH(1,1) MODEL

WOOCHEOL KIM
Korea Institute of Public Finance

OLIVER LINTON
London School of Economics

We propose a semiparametric IGARCH model that allows for persistence in variance but also allows for more flexible functional form. We assume that the difference of the squared process is weakly stationary. We propose an estimation strategy based on the nonparametric instrumental variable method. We establish the rate of convergence of our estimator.

1. INTRODUCTION

A number of authors have found parameter estimates in GARCH(1,1) models close to the unit root region and have proposed using the integrated generalized autoregressive conditional heteroskedasticity (GARCH) or IGARCH process, which imposes this restriction; see, for example, Engle and Bollerslev (1986). The exponentially weighted moving average model (EWMA) (aka the "JP Morgan" model), which is the special case in which the intercept is set to zero, is in wide use by practitioners. The IGARCH process, although it does not possess a finite (limiting) variance, can be strongly stationary; see Nelson (1990). In fact, IGARCH processes can also be geometrically strong mixing; see Meitz and Saikonnen (2004).

We propose a semiparametric extension of the IGARCH model. Our model nests the standard IGARCH(1,1) model, but it allows more flexibility in functional form. It extends the recent model of Linton and Mammen (2005) to the case where the unconditional variance of the process does not exist. We propose an estimation method that involves solving a type one integral equation with estimated operator; see Carrasco, Florens, and Renault (2006). We establish the rate of uniform convergence of the nonparametric part of our model and the consistency of the parametric part.

We would like to thank Xiaohong Chen, Joel Horowitz, and Eric Renault for interesting discussions. We also thank Jean-Pierre Florens and two referees for helpful comments. Thanks to the ESRC and Leverhulme foundations for financial support. This paper was partly written while Linton was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and he thanks them for financial support. Address correspondence to Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; e-mail: o.linton@lse.ac.uk.

© Cambridge University Press 2010

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

All proofs are given in the Appendix.

2. THE MODEL

We suppose throughout that $\{y_t\}_{t=-\infty}^{\infty}$ is an observed, strongly stationary process. We suppose that

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \beta \sigma_{t-1}^2 + (1-\beta) y_{t-1}^2 + m(y_{t-1}), \end{aligned} \tag{1}$$

where ε_t and $\varepsilon_t^2 - 1$ are martingale difference sequences; i.e., $\mathrm{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ and $\mathrm{E}(\varepsilon_t^2 - 1 | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t$ contains all information up to the current period, and $m(\cdot)$ is an unknown function. When $m(y) = \omega$ for some constant ω, the above model reduces to a standard parametric IGARCH(1,1) model, and when $\omega = 0$ it is the EWMA. If $m(y) = \delta y^2 + \omega$, then the process is an "explosive" GARCH process, strictly stationary for some range of $\delta \geq 0$. In general we allow the nonparametric function $m(\cdot)$ to take a flexible form, as long as it satisfies some regularity condition including smoothness, a nonnegativity constraint ($m(\cdot) \geq 0$), and some additional conditions guaranteeing the strong stationarity of y_t. The nonparametric term is introduced to correct possible misspecification with a quadratic growth function of news impact on volatility. We do not assume that $\mathrm{E} y_t^2 < \infty$, in contrast with Linton and Mammen (2005).

Defining the martingale difference sequence $\eta_t = \sigma_t^2[\varepsilon_t^2 - 1]$, we write the squared returns as $y_t^2 = \sigma_t^2 + \eta_t$. By plugging into (1), we get

$$\begin{aligned} \Delta y_t^2 = y_t^2 - y_{t-1}^2 &= m(y_{t-1}) + (1-\beta)\eta_{t-1} + \eta_t - \eta_{t-1} \\ &= m(y_{t-1}) + \eta_t - \beta \eta_{t-1}. \end{aligned} \tag{2}$$

The squared return $\{y_t^2\}$ is an integrated process with a functional drift term, $m(\cdot)$, and moving average error term. If $m(\cdot) \leq c$ (with c not so large), the model is likely to show similar dynamics to the standard IGARCH process.

For our theoretical development, it is important that $\{\Delta y_t^2\}$ satisfies weak stationarity even when the process $\{y_t^2\}$ does not. In the context of linear time series models this is a property that is quite common, but in the current context it is not obviously possible. Harvey, Ruiz, and Shephard (1994) say that (in the case where ε_t is standard normal) "$\{\Delta y_t^2\}$ is stationary and has an ACF like that of an MA(1) process." Their argument seems to be based on the fact that the innovation process η_t appears to be a martingale difference sequence. Although it appears that $\mathrm{E}[\eta_t | \mathcal{F}_{t-1}] = 0$, this definition only makes sense if $\mathrm{E}[|\eta_t|] < \infty$ or $\lim_{t\to\infty} \mathrm{E}[|\eta_t| | \mathcal{F}_0] < \infty$ (Hall and Heyde, 1980). In fact, these conditions do not hold. Therefore, one cannot conclude anything about weak stationarity. In any case, weak stationarity of Δy_t^2 requires both its mean and variance to exist, which would require that $\lim_{t\to\infty} \mathrm{E}[\eta_t^2 | \mathcal{F}_0] < \infty$. In fact, $\{\Delta y_t^2\}$ is not weakly stationary in the Gaussian strong IGARCH. The cause of this counterintuitive (from the point of view of linear processes) behavior is the independent and identically

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

distributed (i.i.d.) innovation. The following example shows that when the innovations are not i.i.d., one can have $\{\Delta y_t^2\}$ weak stationarity (or at least have finite first moment) even when $\{y_t\}$ is not weakly stationary.

Example
Consider a semistrong IGARCH model, i.e., (1) with $m(y) = \omega$ and

$$\varepsilon_t = \text{sign}(z_t)\left\{1 + \frac{u_t^2 - 1}{1 + \sigma_t^2}\right\}^{1/2},$$

where z_t, u_t are i.i.d. and mutually independent random variables with Esign $(z_t) = 0$ and $\text{E}u_t^2 = 1$. It follows that ε_t and $\varepsilon_t^2 - 1$ are martingale difference sequences and thereby are consistent with (1). Furthermore, $\eta_t = (u_t^2 - 1)\left(\sigma_t^2/(1+\sigma_t^2)\right)$, which satisfies $\text{E}(|\eta_t|) < \infty$, and so we have $\text{E}(|\Delta y_t^2|) < \infty$. However, $\text{E}\left(\sigma_t^2\right) = \text{E}\left(y_t^2\right) = \infty$. If z_t is symmetric about zero, then ε_t is also symmetrically distributed both conditionally and unconditionally. Provided u_t has finite fourth moment, then $\text{E}([\Delta y_t^2]^2) < \infty$.

In the sequel we shall assume that the process $\{\Delta y_t^2\}$ is weakly stationary. This has strong and testable empirical implications, and we investigate whether this is a reasonable assumption in some common data sets below.

An interesting feature of our semiparametric IGARCH model is that the nonlinear correction term m can be identified independently of β. Significant estimates of m are related directly to misspecification of IGARCH. Also, by means of nonparametric function m, (1) can nest both GARCH and IGARCH as a special case. This model is related to recent work of Linton and Mammen (2005), who considered the case with $\sigma_t^2 = \beta\sigma_{t-1}^2 + m(y_{t-1})$ and $\text{E}(y_t^2) < \infty$. The estimation strategy there involved solving a Type 2 integral equation and was simpler to analyze. The estimation strategy we develop here can be used in their model, but yields poorer rates of convergence.

3. IDENTIFICATION

Let $f_0(\cdot)$ be the marginal density function of y_t, and denote the joint density function of (y_{t-1}, y_{t-k}) and $(\Delta y_t^2, y_{t-k})$ by $f_k(\cdot,\cdot)$ and $f_k^{\Delta}(\cdot,\cdot)$, respectively, for suitable k. Letting $\nu_t = \eta_t - \beta\eta_{t-1}$, we write (1) as

$$\Delta y_t^2 = m(y_{t-1}) + \nu_t, \tag{3}$$

where $\text{E}(\nu_t|y_{t-1}) \neq 0$ but $\text{E}(\nu_t|y_{t-k}) = 0$ for $k \geq 2$; (3) is an example of nonparametric structural models with an *infinite* number of instruments. By the finite moment condition ($\text{E}(|\Delta y_t^2|) < \infty$) and the law of iterated expectations, we obtain, from taking conditional expectations of (3),

$$\text{E}[m(y_{t-1})|y_{t-k} = w] = \text{E}[\Delta y_t^2|y_{t-k} = w], \quad \text{for all } k \geq 2. \tag{4}$$

This can be viewed as an integral equation of the first kind with solution $m(.)$. It is convenient to multiply both sides of the equation by the marginal density of y_t

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

evaluated at w, which preserves the equation but makes the analysis a bit simpler. Define the linear operator $\mathcal{T}_k : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ by

$$\mathcal{T}_k m(w) = \int m(x) f_k(x, w)\, dx = \mathrm{E}[m(y_{t-1})|y_{t-k} = w] f_0(w),$$

and a reduced form function $h_k(\cdot)$ by

$$h_k(w) = \int z f_k^{\Delta}(z, w)\, dz = \mathrm{E}(\Delta y_t^2 | y_{t-k} = w) f_0(w),$$

where $f_0(\cdot)$ is the marginal density function of y_t. Then, $m(\cdot)$ satisfies the equation

$$\mathcal{T}_k m(w) = h_k(w) \tag{5}$$

for all $k \geq 2$. The solution to the integral equation is unique (if it exists) if and only if, for some $k \geq 2$, $\mathcal{T}_k$ is one-to-one, or equivalently, the conditional distribution of y_{t-1}, given y_{t-k}, is statistically complete in the sense that $\mathrm{E}[m(y_{t-1})|y_{t-k}] = 0$, almost surely (a.s.), only for $m \equiv 0$. A sufficient condition for completeness is that the conditional cumulative distribution function (CDF), $F_{1|k}(\cdot|\cdot)$, is a member of an exponential family satisfying certain regularity conditions given in Newey and Powell (2004). In Blundell, Chen, and Kristensen (2007), an alternative but weaker condition is suggested for *bounded completeness*. Since we implicitly assume that m is uniformly bounded, the latter kind of completeness may be more relevant for our model. However, considering that the process $\{y_t\}$ defined by (1) does not possess a finite second moment, the aforementioned approaches cannot be applied here. Our identification result below is based on the approach of Kim (2003). We could establish identification from (5) for any single k or from a finite set of such moment equations. However, motivated by the work of Linton and Mammen (2005), we combine all the equations (5) into a single equation using a weighting sequence and establish identification for this weighted equation.

Let $f_\lambda(x, w) = \sum_{k=2}^{\infty} \lambda_k f_k(x, w)$ and $h_\lambda(w) = \sum_{k=2}^{\infty} \lambda_k h_k(w)$, for $\lambda_k \geq 0$ with $\sum_{k=2}^{\infty} \lambda_k = 1$. This includes the case where $\lambda_k = 1$ and $\lambda_j = 0$ for all $j \neq k$, and it also includes the case where $\lambda_k = \beta^k$, which turned out to be the optimal equation in Linton and Mammen (2005). Define $\mathcal{T}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ to be a linear operator such that $\mathcal{T}_\lambda = \sum_{k=2}^{\infty} \lambda_k \mathcal{T}_k$; then we have $\mathcal{T}_\lambda m(w) = h_\lambda(w)$. We will define identifiability of m in the context of this equation.

DEFINITION 3.1. *The true function $m_0(\cdot)$ is identifiable if the solution to the integral equation*

$$\mathcal{T}_\lambda m(w) = \int m(x) f_\lambda(x, w)\, dx = h_\lambda(w) \tag{6}$$

is unique or equivalently if $\mathcal{T}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ is one-to-one, for the weighting scheme $\lambda = \{\lambda_k\}_{k=2}^{\infty}$ such that $\sum_{k=2}^{\infty} \lambda_k = 1$.

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

Below, we give a sufficient condition for invertibility of $\mathcal{T}_\lambda$. Given $\{\omega_l\}_{l=1}^L \subset \mathcal{Y}$, we define a marginal discretization (with respect to y_{t-k}) of the joint density function $f_\lambda(\cdot,\cdot)$ by $[f_\lambda(x,\omega_1),\ldots,f_\lambda(x,\omega_L)]^\top$. Let $\text{lin}(\{f_\lambda(\cdot,\omega_l)\}_{l=1}^L)$ be the linear space generated by $\{f_\lambda(\cdot,\omega_l)\}_{l=1}^L$, and $\overline{\text{lin}}\{f_\lambda(\cdot,\omega_l)\}_{l=1}^\infty$ the closure of $\text{lin}\{f_\lambda(\cdot,\omega_l)\}_{l=1}^\infty$ in $L^2(\mathcal{Y})$. Our identification results make use of the condition below.

Condition 3.1. For the given sequence $\lambda = \{\lambda_k\}_{k=2}^\infty$, satisfying $\lambda_k \geq 0$ and $\sum_{k=2}^\infty \lambda_k = 1$, for some sequence $\overline{\mathcal{Y}} = \{\omega_l\}_{l=1}^\infty \subset Y$, $\text{lin}\{f_\lambda(\cdot,\omega_l)\}_{l=1}^\infty$ is dense in $L^2(Y)$, i.e., $\overline{\text{lin}}\{f_\lambda(\cdot,\omega_l)\}_{l=1}^\infty = L^2(\mathcal{Y})$.

The above condition concerns richness of the linear spaces that are generated by a weighted sum of (unnormalized) conditional density function. Condition 3.1 will hold if a complete orthogonal basis of $L^2(\mathcal{Y})$ is generated by linear combinations of $\{f_\lambda(\cdot,\omega_l)\}_{l=1}^\infty$. The following theorem shows that Condition 3.1 is sufficient for $\mathcal{T}_\lambda$ to be one-to-one. The proof is immediate from Kim (2003, Thm. 2.2(i), p. 7).

PROPOSITION 3.1. *If Condition 3.1 holds for some $k \geq 2$, then the integral operator $\mathcal{T}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ is one-to-one, and $m_0(\cdot)$ is identified by $\mathcal{T}_\lambda^{-1}(h_\lambda) \in L^2(\mathcal{Y})$, for $h_\lambda \in \mathcal{R}(\mathcal{T}_\lambda)$.*

The suggested identifying condition seems rather abstract, partly because we do not use any parametric assumptions. Roughly speaking, identifiability depends on the way that the density function of y_{t-1}, conditional on $y_{t-k} = \omega_l$ (or their weighted version), varies over different values of ω_l's. For example, the model is identifiable if some sequence of the conditional density functions, $\{f_{|\lambda}(\cdot|\omega_l)\}_{l=1}^\infty$, includes (or spans) a complete basis of $L^2(\mathcal{Y})$. Condition 3.1 excludes a joint density function of form $f_k(x,w) = \sum_{k=1}^K p_k(x)q_k(w)$ for finite K.

Remark 3.1. Since $\text{E}(\eta_t|\mathcal{F}_{t-k}) = \text{E}(\eta_{t-1}|\mathcal{F}_{t-k}) = 0$, for any $k \geq 2$, one may think of using a (non)linear function of $(y_{t-k}, y_{t-k-1},\ldots,)$ as an instrument—a possibility that is not covered by the consideration above. Because of the curse of dimensionality arising from high dimensional conditioning variables, we will work only with moment conditions conditionalized on a single instrument. In this context, an alternative approach will be to use $w_t^* = \sum_{k\geq 2} \lambda_k y_{t-k}$ as an instrument. Note that even when Condition 3.1 holds for no λ, a similar condition may hold for w_t^*.

Finally, we turn to the parametric term β. With m_0 given by Proposition 3.1, the GARCH coefficient β_0 can be identified from the expected Gaussian likelihood function[1]

$$l(\beta) = \text{E}\left[\ln\sigma_t^2(\beta,m_0) + \frac{y_t^2}{\sigma_t^2(\beta,m_0)}\right],$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

that is, $\beta_0 = \arg\min_{\beta \in B} l(\beta)$, where B is a compact subset of $(0, 1)$, while

$$\sigma_t^2(\beta, m_0) = \sum_{j=1}^{T-1} \beta^{j-1}[(1-\beta)y_{t-j}^2 + m(y_{t-j})].$$

4. ESTIMATION

We suppose that the quantities h_k and f_k are unknown but that there is an observed sample $\{y_t\}_{t=1}^T$. We shall assume now that the operator $\mathcal{T}_\lambda$ is invertible.

4.1. Nonparametric Term

Let τ be a truncation parameter satisfying $\tau(T) \to \infty$ as $T \to \infty$, and let $T^* = T - \tau$. We propose to estimate the quantities h_k and f_k from the sample data. Specifically, define

$$\widehat{h}_k(w) = \frac{1}{T^*} \sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w)\Delta y_t^2, \tag{7}$$

$$\widehat{f}_k(x, w) = \frac{1}{T^*} \sum_{t=\tau+1}^{T} K_{g_1}(y_{t-1} - x)K_{g_2}(y_{t-k} - w),$$

$$(\widehat{\mathcal{T}}_k m)(w) = \int \frac{1}{T^*} \sum_{t=\tau+1}^{T} K_{g_1}(y_{t-1} - x)K_{g_2}(y_{t-k} - w)m(x)\,dx, \tag{8}$$

where $K_g(s) = K(s/g)/g$, with $K(\cdot)$ being a symmetric function defined on the real line, while g_1 and g_2 are positive bandwidths.

We next solve the implied random integral equation to give our estimate of m_0. Let $\widehat{h}_\lambda(w) = \sum_{k=2}^{\tau} \lambda_k \widehat{h}_k$ and $\widehat{\mathcal{T}}_\lambda = \sum_{k=2}^{\tau} \lambda_k \widehat{\mathcal{T}}_k$. Then consider the random Fredholm integral equation of the first kind,

$$(\widehat{\mathcal{T}}_\lambda m)(w) = \int_{\mathcal{Y}} m(x)\widehat{f}_\lambda(x, w)\,dx = \widehat{h}_\lambda(w), \tag{9}$$

where $\widehat{f}_\lambda(y, w) = \sum_{k=2}^{\tau} \lambda_k \widehat{f}_k(y, w)$. As is well known in mathematical inverse problems, several difficulties arise in estimating m_0 by inverting $\widehat{h}_\lambda$ through $\widehat{\mathcal{T}}_\lambda$.

Since $\widehat{\mathcal{T}}_\lambda$ is generally of finite rank, it is likely that $\widehat{h}_\lambda \notin \mathcal{R}(\widehat{\mathcal{T}}_\lambda)$, or $\widehat{\mathcal{T}}_\lambda$ is not invertible, i.e., the integral equation in (9) may possess *no* solution or *more than one* solution. One resolves the existence and uniqueness problems easily by using the (Moore-Penrose) generalized inverse of $\widehat{\mathcal{T}}_\lambda$:

$$\widehat{m}^\dagger = \underset{m(\cdot) \in L^2(\mathcal{Y})}{\arg\min} \; ||\widehat{\mathcal{T}}_\lambda m - \widehat{h}_\lambda||_{L^2(\mathcal{Y})}^2, \tag{10}$$

where $\widehat{m}^\dagger$ is the solution of minimum norm, unless the minimum-distance estimator is unique. Consistency of the natural estimator $\widehat{m}^\dagger$, however, is not ensured by consistency of the preliminary estimates $\widehat{h}_\lambda$ and $\widehat{\mathcal{T}}_\lambda$, since $\widehat{\mathcal{T}}_\lambda^\dagger$ is not

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

bounded uniformly in T. Let $||\mathcal{T}||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})}$ denote an operator norm of $\mathcal{T}: L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$, given by $\sup_{m\in L^2(\mathcal{Y}),\, m\neq 0} ||\mathcal{T}m||_{L^2(\mathcal{Y})}/||m||_{L^2(\mathcal{Y})}$. We say that $\widehat{\mathcal{T}}_\lambda : L^2(Y) \to L^2(Y)$ is uniformly consistent for $\mathcal{T}_\lambda$ on M_Y, if and only if $||\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \xrightarrow{p} 0$.

PROPOSITION 4.1. *Suppose that $\mathcal{T}_\lambda$ is invertible. Assume that $\widehat{\mathcal{T}}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ is uniformly consistent for T_λ on $M_Y \subset L^2(Y)$ such that* $\dim(M_Y) = \infty$. *Then,*

$$\text{plim}_{T\to\infty} ||\widehat{\mathcal{T}}_\lambda^\dagger||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} = \infty.$$

The naive estimator lacks stability with respect to the statistical errors in $\widehat{\mathcal{T}}_\lambda$ or $\widehat{h}_\lambda$. Small perturbations of $\widehat{\mathcal{T}}_\lambda$ or $\widehat{h}_\lambda$ may result in unacceptably large errors in $\widehat{m}^\dagger = \widehat{\mathcal{T}}_\lambda^\dagger(\widehat{h}_\lambda)$. Note that the estimation problem in (9) is *statistically ill-posed*, since the underlying mapping from h_λ to m is not continuous. For consistent estimation, some regularization is necessary.

We consider the Tikhonov regularization method, in which we define the adjoint operator

$$(\widehat{\mathcal{T}}_k^* h)(y) = \int \left[\frac{1}{T^*}\sum_{t=\tau+1}^{T} K_{g_1}(y_{t-1} - y) K_{g_2}(y_{t-k} - w)\right] h(w)\, dw.$$

By Fubini's theorem,

$$< \widehat{\mathcal{T}}_\lambda m, h >_{L^2(\mathcal{Y})} = < m, \widehat{\mathcal{T}}_\lambda^* h >_{L^2(\mathcal{Y})} \text{ a.s.},$$

and hence the two random operators $\widehat{\mathcal{T}}_\lambda$ and $\widehat{\mathcal{T}}_\lambda^*$, where $\widehat{\mathcal{T}}_\lambda^* = \sum_{k=2}^{\tau} \lambda_k \widehat{\mathcal{T}}_k^*$, are adjoint to each other. From $\dim(\mathcal{R}(\widehat{\mathcal{T}}_\lambda)) \le T$, it follows that both $\widehat{\mathcal{T}}_\lambda$ and the self-adjoint operator $\widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ are bounded and compact.

We define a kernel instrumental variable (IV) estimator, based on the ordinary Tikhonov regularization, as

$$\widehat{m}_\alpha = (\widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda + \alpha I)^{-1} \widehat{\mathcal{T}}_\lambda^* \widehat{h}_\lambda. \tag{11}$$

From the fact that $\widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda$ is self-adjoint, $(\widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda + \alpha I)^{-1}$ is well defined based on spectral theory for self-adjoint linear operators, since the real-valued function $U_\alpha(\kappa) = (\alpha + \kappa)^{-1}$ is well defined on the spectrum of $\widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda$.

To show the closed form of the kernel estimator, we need the following definitions: For

$$K_T^Y(x) = [K_{g_1}(y_\tau - x), \ldots, K_{g_1}(y_{T-1} - x)]^\top,$$

$$K_T^\lambda(w) = [\sum_{k=2}^{\tau} \lambda_k K_{g_2}(y_{\tau-k+1} - w), \ldots, \sum_{k=2}^{\tau} \lambda_k K_{g_2}(y_{T-k} - w)]^\top,$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

define

$$M_Y = \int_{\mathcal{Y}} K_T^Y(x) K_T^Y(x)^\top dx \quad \text{and} \quad M_\lambda = \int_{\mathcal{Y}} K_T^\lambda(w) K_T^\lambda(w)^\top dw.$$

Using a convolution-kernel function, we can rewrite the (i, j)th element of M_λ, for example, in a more compact way, as

$$\begin{aligned} M_{ij}^\lambda &= \int_{\mathcal{Y}} \sum_{k=2}^{\tau} \lambda_k K_{g_2}(y_{\tau-k+i} - w) \sum_{l=2}^{\tau} \lambda_l K_{g_2}(y_{\tau-l+j} - w)\, dw \\ &= \sum_{k=2}^{\tau} \sum_{l=2}^{\tau} \lambda_k \lambda_l \int_{\mathcal{Y}} K_{g_2}(y_{\tau-k+i} - w) K_{g_2}(y_{\tau-l+j} - w)\, dw \\ &= \sum_{k=2}^{\tau} \sum_{l=2}^{\tau} \lambda_k \lambda_l K_{g_2}^c(y_{\tau-k+i} - y_{\tau-l+j}), \end{aligned}$$

where $K_{g_2}^c(w) = (1/g_2) \int_{\mathcal{Y}} K(w/g_2 - s) K(s)\, ds$. A straightforward calculation shows that M_λ is a $(T^* \times T^*)$ symmetric nonnegative semidefinite matrix, for which the square-root matrix $M_\lambda^{1/2}$, satisfying $M_\lambda = M_\lambda^{1/2} M_\lambda^{1/2}$, is well defined.[2] Also, define $Q_{Y,\lambda} = T^{*-2} M_\lambda^{1/2} M_Y M_\lambda^{1/2}$; $Q_{Y,\lambda}$ is a $(T^* \times T^*)$ symmetric nonnegative semidefinite matrix, whose eigenvalues are all real and positive. We denote, by $\lambda_{\max}(Q_{Y,\lambda})$, the maximum of those eigenvalues. We can show that

$$\begin{aligned} \widehat{m}_\alpha(x) &= \left[(\alpha I + \widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda)^{-1} \widehat{\mathcal{T}}_\lambda^* \widehat{h}_\lambda \right](x) \\ &= T^{*-2} K_T^Y(x)^\top M_\lambda^{1/2} (Q_{Y,\lambda} + \alpha I_T)^{-1} M_\lambda^{1/2} \mathbf{y}, \end{aligned} \tag{12}$$

where $\mathbf{y} = (\Delta y_{\tau+1}^2, \ldots, \Delta y_T^2)^\top$.

By (12), the abstract operator form of the kernel estimator translates into a concrete matrix form. Computations of $\widehat{m}_\alpha$ only involve simple operations on finite-dimensional matrices, when the convolution-kernel weights in M_Y and M_λ are given.

Remark 4.1. (i) Suppose that $K(\cdot)$ is a density function from a stable distribution, say, a Gaussian kernel. Then, a further simplification of the convolution-kernel weight is available:

$$M_{ij}^\lambda = \sum_{k=2}^{\tau} \sum_{l=2}^{\tau} \lambda_k \lambda_l K_{g_2}^c(y_{\tau-k+i} - y_{\tau-l+j}) = \sum_{k=2}^{\tau} \sum_{l=2}^{\tau} \lambda_k \lambda_l K_{\sqrt{2} g_2}(y_{\tau-k+i} - y_{\tau-l+j}),$$

from $K^c(s) = K(s/\sqrt{2})/\sqrt{2}$, since, by the stability assumption, the shape of a convoluted density function is not changed, except that the variance doubles. In that case, all the matrices in (12) are calculated in a straightforward way. In general, when there is no explicit form for the convolution kernel, we can compute $K^c(\cdot)$ by numerical integration.

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

(ii) By the spectral representation results in the proof of Theorem 5.2, the naive minimum-distance estimator in the above has a closed form

$$\widehat{m}^{\dagger}(x) = K_T^Y(x)^{\top} M_\lambda^{1/2} (M_\lambda^{1/2} M_Y M_\lambda^{1/2})^{\dagger} M_\lambda^{1/2} \mathbf{y}.$$

If both $K_T^Y(\cdot)$ and $K_T^\lambda(\cdot)$ are assumed to be linearly independent, then M_λ and M_Y are positive definite, from which we get $\widehat{m}^{\dagger}(x) = K_T^Y(x)^{\top} M_Y^{-1}\mathbf{y}$. From

$$\begin{aligned}(\widehat{\mathcal{T}}_\lambda \widehat{m}^{\dagger})(w) &= \int \widehat{f}_\lambda(x,w)\widehat{m}^{\dagger}(x)\,dx = \frac{1}{T^*} K_T^\lambda(w)^{\top} < K_T^Y(\cdot), K_T^{Y\top}(\cdot) >_{L^2(\mathcal{Y})} M_Y^{-1}\mathbf{y} \\ &= \frac{1}{T^*} K_T^\lambda(w)^{\top}\mathbf{y} = \widehat{h}_\lambda(w),\end{aligned}$$

we can confirm that $\widehat{m}^{\dagger}(\cdot)$ is one of the exact solutions to the integral equation, $\widehat{\mathcal{T}}_\lambda m = \widehat{h}_\lambda$, where $\widehat{\mathcal{T}}_\lambda$ will not be invertible in general. By the definition of the generalized inverse, $\widehat{m}^{\dagger}(\cdot)$ will be the solution of minimum-norm. Instability of $\widehat{m}^{\dagger}$ is obvious from the minimum eigenvalue of M_Y converging to zero, as $T \to \infty$, since a pair of elements in $K_T^Y(\cdot)$ should become arbitrarily close to each other.

4.2. Parametric Term

With a nonparametric estimate of m_0 given by $\widehat{m}_\alpha(\cdot)$ in the previous section, the parametric GARCH coefficient β can be estimated by

$$\widehat{\beta} = \arg\min_{\beta \in B} \widehat{\ell}(\beta), \tag{13}$$

where B is a compact subset of $(0,1)$ containing β_0, while

$$\widehat{\ell}(\beta) = \frac{1}{T}\sum_{t=1}^{T} \ln \sigma_t^2(\beta, \widehat{m}_\alpha) + \frac{y_t^2}{\sigma_t^2(\beta, \widehat{m}_\alpha)},$$

$$\sigma_t^2(\beta, \widehat{m}_\alpha) = \sum_{j=1}^{\min\{t-1, \tau_\dagger\}} \beta^{j-1}[(1-\beta)y_{t-j}^2 + \widehat{m}_\alpha(y_{t-j})].$$

Here, $\tau_\dagger = \tau_\dagger(T) < T$ is another truncation parameter. The estimator can be computed easily by a grid search over B.

5. ASYMPTOTIC PROPERTIES

5.1. Nonparametric Estimates

Here we analyze the asymptotic properties of the kernel estimators proposed in the previous section. Let $\mathcal{F}_b^a$ be the σ-algebra of events generated by $\{y_t\}_a^b$ and $\alpha(k)$ the strong mixing coefficient of $\{y_t\}$ that is defined by

$$\alpha(k) \equiv \sup_{A \in \mathcal{F}_{-\infty}^0,\, B \in \mathcal{F}_k^\infty} |P(A \cap B) - P(A)P(B)|.$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

Condition 5.1. (a) $\{y_t\}_{t=1}^{\infty}$ is strictly stationary and strongly mixing (with a mixing coefficient, $\alpha(k) = \rho^{-\xi k}$, for some $\xi > 0$), and satisfies (1) with m_0 identified by $\mathcal{T}_{\lambda}^{-1} h_{\lambda}$. (b) $\mathrm{E}(|\Delta y_t^2||y_{t-k} = w)$ is bounded uniformly in w, a.s.

Condition 5.2. Let $K(\cdot) \in \mathbb{K}_{p^*}$, where $\mathbb{K}_{p^*}$ is the class of all Borel measurable symmetric real-valued functions $K(s)$ such that (a)

$$\int |K(s)|ds < \infty, \quad \int K(s)ds = 1, \quad \int K^2(s)ds < \infty, \quad \sup|K(s)| < \infty,$$

and (b) $\int s^j K(s)ds = 0$, for $j = 1, \ldots, p^* - 1$, and $\mu_{p^*}(K) = \int s^{p^*} K(s)ds < \infty$, where p^* is an even integer.

Condition 5.3. The joint density functions $f_k(\cdot, \cdot)$ is square-integrable and bounded

$$\sup_{k \geq 1} \int_{\mathcal{Y}} \int_{\mathcal{Y}} f_k^2(y, w)dydw < \infty, \quad \text{and} \quad \sup_{k \geq 1} \sup_{(y,w) \in \mathcal{Y} \times \mathcal{Y}} f_k(y, w) \leq C < \infty.$$

Condition 5.4. $f_k(\cdot, \cdot)$ and $m_0(\cdot)$ have continuous p_0th and p_1th partial derivatives, respectively, that are square-integrable, satisfying

$$\sup_{k \geq 2} \left\| \frac{\partial^{p_0} f_k(y, w)}{\partial y^q \partial w^{p_0 - q}} \right\|_{L^2(\mathcal{Y} \times \mathcal{Y})}^2 \leq C, \quad \text{and} \quad \sup_{k \geq 2} \left\| \frac{d^{p_1} m(y)}{dy^{p_1}} \right\|_{L^2(\mathcal{Y})}^2 \leq C.$$

Condition 5.5. (a) The bandwidth parameters (g_1, g_2) satisfy that $\max(g_1, g_2) \to 0$, $Tg_2 \to \infty$. (b) The regularization parameter α satisfies that $\alpha \to 0$, $Tg_2\alpha \to \infty$, and $g_1^{p_0}/\sqrt{\alpha} \to 0$, as $T \to \infty$.

Condition 5.6. $\{\lambda_k\}_{k=2}^{\infty}$ and $\tau = \tau_T$ are such that $\lambda_k \geq 0$, $\sum_{k=2}^{\infty} \lambda_k = 1$, and $\sum_{k=\tau_T+1}^{\infty} \lambda_k = o(1/\sqrt{T})$.

All the technical conditions in Condition 5.2 through 5.4 are standard in nonparametric kernel estimation. As will be shown later, the L^2-convergence rate of our estimate can be derived under no requirement that the joint density functions have compact support or are bounded away from zero. For uniform convergence results, however, the conditions in 5.3 and 5.4 will be strengthened, being replaced by a compact support assumption, together with the continuity condition. Note that the square-integrability condition in Condition 5.3 entails boundedness of the linear operator $\mathcal{T}_k$. It can be satisfied under compact support with densities bounded away from zero, but it can also be satisfied in Gaussian process cases. See Linton and Mammen (2005) for further discussion. Condition 5.5(b), which is rather stronger than 5.5(a), is necessary for consistency of the regularized kernel estimates. Condition 5.6 gives a convenient condition for controlling the approximation errors due to truncation. Condition 5.6 is satisfied, for example, when $\lambda_k = \lambda^k$ and $\tau_T = T^{-1/m}$, for some positive (relatively large) integer. Let $\widehat{h}_k$ and

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

$\widehat{\mathcal{T}}_k$ be given by (7) and (8), respectively. Our first result concerns sufficiency of the above conditions for derivation of the basic properties of the preliminary estimates, including consistency and the convergence rates.

PROPOSITION 5.1. *Suppose that Conditions 5.1 through 5.5(a) and 5.6 hold. Then*

(i) $||\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} = O_p(1/\sqrt{Tg_2} + g_1^{p_0} + g_2^{p_0})$,

(ii) $||\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} = O_p(1/\sqrt{Tg_1} + g_1^{p_0} + g_2^{p_0})$,

(iii) $||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0||_{L^2(\mathcal{Y})} = O_p(1/\sqrt{Tg_2} + g_1^{\overline{p}})$,

where $\overline{p} = \max(p_0, p_1) \le p^*$.

Noting that $U_\alpha(\kappa) = (\alpha + \kappa)^{-1}$ satisfies the conditions of C.3.1 and C.3.2 in Kim (2003, p. 15), we can show the asymptotic properties of the kernel estimator $\widehat{m}_\alpha$, by applying the general results for statistical regularization (Kim, Thm. 3.3), together with Proposition 5.1. Let $\mathcal{N}(T)$ and $\mathcal{R}(T)$ denote the null space and range, respectively, of the operator T.

THEOREM 5.2. *Assume that Conditions 5.1 through 5.6 hold, with* $p_0 = p_1$. *Then,*

(i) $||\widehat{m}_\alpha - m_0||_{L^2(\mathcal{Y})} \xrightarrow{p} 0$, *as* $T \to \infty$, *for all* $m_0 \in L^2(\mathcal{Y})$.

Assume additionally that $m_0 \in \mathcal{R}(\mathcal{T}_\lambda^* \mathcal{T}_\lambda)$. *Then,*

$$\text{(ii) } ||\widehat{m}_\alpha - m_0||_{L^2(\mathcal{Y})} \le O_p\left(\frac{1}{\sqrt{\alpha}}\left[\frac{1}{\sqrt{Tg_2}} + g_1^{p_0}\right]\right) + O_p\left(\frac{1}{\sqrt{Tg_1}} + g_2^{p_0}\right) + O_p(\alpha).$$

Since $\mathcal{T}_\lambda$ is an integral operator, the condition $m_0 \in \mathcal{R}(\mathcal{T}_\lambda^* \mathcal{T}_\lambda)$ imposes certain smoothness on m_0, which we call an abstract smoothness condition. When $\mathcal{T}_\lambda$ is a compact operator, it means that the generalized Fourier coefficients of m_0 (with respect to the eigenfunctions) decay fast enough relative to the eigenvalues of $\mathcal{T}_\lambda$.[3]

Remark 5.1 (The optimal convergence rate)**.** Let $m0_0$ be any function in $\mathcal{R}(\mathcal{T}_\lambda^* \mathcal{T}_\lambda)$. Suppose a side condition on (g_1, g_2) such that $(Tg_1)^{-1/2} \ 0 \le O(g_1^{2p_0/3})$, and $g_2^{3p_0/2} \le O([Tg_2]^{-1/2})$. Then, the optimal convergence rate of $\widehat{m}_\alpha$ is given by

$$||\widehat{m}_\alpha - m_0||_{L^2(\mathcal{Y})} = O_p(T^{-p_0/(3p_0+1)}),$$

under the optimal choice of smoothing parameters such that $g_{1T}^* = C_0 T^{-1/((4/3)p_0+1)}$, $g_{2T}^* = C_1 T^{-1/(3p_0+1)}$, and $\alpha_T^* = C_2 T^{-p_0/(3p_0+1)}$.

THEOREM 5.3 (Uniform convergence rate of $\widehat{m}_\alpha(x)$). *Assume that Conditions 5.1 through 5.6 hold (*$p_0 = p_1$*) with compactness of the support* $\mathcal{Y}$,

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

and that $m_0 \in \mathcal{R}([\mathcal{T}_\lambda^* \mathcal{T}_\lambda]^\mu)$, *with* $\mu \geq 1$. *If* $\alpha = o(1/(\log T)^c)$, *for any* $c > 0$, *then,*

$$\sup_{y \in \mathcal{Y}} |\widehat{m}_\alpha(y) - m_0(y)| = O_p\left(\frac{\log T}{\sqrt{\alpha}}\left[\frac{1}{\sqrt{T g_2}} + g_1^{p_0}\right]\right) + O_p\left(\log T\left[\frac{1}{\sqrt{T g_1}} + g_2^{p_0}\right]\right) + O_p(\alpha \log T).$$

By using the same argument as Remark 5.1, we can show that the optimal uniform convergence rate of $\widehat{m}_\alpha$ is given by

$$\sup_{y \in \mathcal{Y}} | \widehat{m}_\alpha(y) - m_0(y) | = O_p(T^{-p_0/(3p_0+1)} \log T), \quad \text{for } m_0 \in \mathcal{R}(\mathcal{T}_\lambda^* \mathcal{T}_\lambda),$$

where the optimal choice of smoothing parameters is the same as in Remark 5.1. Note that in this case, the twice-differentiability condition of m and f_k implies that

$$\sup_{y \in \mathcal{Y}} | \widehat{m}_\alpha(y) - m_0(y) | = O_p(T^{-2/7} \log T) = o_p(T^{-1/4}).$$

5.2. Parametric Estimate

We here establish the consistency with rate for the parametric part. With the uniform convergence rate of $\widehat{m}_\alpha(\cdot)$ faster than $T^{-1/4}$, the derivations are much simpler than those of Theorem 6 and the discussion in Linton and Mammen (2005, Sect. 4.4), since $\widehat{m}_\alpha$ does not depend on $\widehat{\beta}$. First define the sample Gaussian quasi-likelihood (up to constants) assuming that m_0 is known perfectly,

$$\ell(\beta) = \frac{1}{T}\sum_{t=1}^{T} \ln \sigma_t^2(\beta, m_0) + \frac{y_t^2}{\sigma_t^2(\beta, m_0)},$$

and suppose that $\widetilde{\beta}$ is the unique minimizer of $\ell(\beta)$, while β_0 is the unique minimizer of $\mathrm{E}[\ell(\beta)]$. The properties of $\widetilde{\beta}$ have been well studied in the literature; see, for example, Lee and Hansen (1994) and Rahbek and Jensen (2004). In fact, under the conditions that make y_t strictly stationary and geometrically mixing we can expect $\widetilde{\beta}$ to be root-T consistent and asymptotically normal.

THEOREM 5.4. *Assume that Conditions 5.1 through 5.6 hold* ($p_0 = p_1$) *with compactness of the support* $\mathcal{Y}$, *and that* $m_0 \in \mathcal{R}([\mathcal{T}_\lambda^* \mathcal{T}_\lambda]^\mu)$, *with* $\mu \geq 1$. *Suppose that* $\widetilde{\beta}$ *is* $T^{-1/4}$*-consistent. If* $\alpha = o(1/(\log T)^c)$, *for any* $c > 0$, *and if* $\tau_\dagger(T) = c \log T$, *then,*

$$\widehat{\beta} - \beta = o_p(T^{-1/4}).$$

Under some conditions one might be able to obtain asymptotic normality at rate root-T from the arguments of Chen and Shen (1998), but this is not guaranteed; see Kim (2003).

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

6. NUMERICAL RESULTS

For an extensive simulation study of this method applied to a cross-sectional setting, see Kim (2003). We content ourselves with providing some evidence of the empirical relevance of our model and assumptions.

6.1. Empirical Application

The assumption that the process $\{\Delta y_t^2\}$ is weakly stationary while y_t^2 is not can be tested by examining the tail index of the two series. In Practice one has to demean the process y_t first. We investigate a sample of daily returns on the S&P 500 from 1955 to 2002, a total of 11,893 observations. The tail index κ of a series X_t is defined as the value for which

$$1 - \Pr(X_t > x) \simeq 1 - Lx^{-\kappa}$$

as $x \to \infty$, where L is a constant or a slowly varying function of x. We compute the tail index $\widehat{\kappa}$ of an ordered positive series X_t by the Hill (1975) method

$$\frac{1}{\widehat{\kappa}} = \frac{1}{M} \sum_{t=1}^{M} \log \frac{X_t}{X_{M+1}}.$$

The estimated value of κ is generally above one but less than two, implying that $\mathrm{E}[y_t^2] < \infty$ but $\mathrm{E}[|y_t|^3] = \infty$.[4] The corresponding hill plots for $|\Delta y_t^2|$ show slightly lighter tails but not enough to change the implications about the existence of moments. For these data at least, the evidence of integrated process is weak, although it does seem that differencing reduces the weight of the tails a little. Our second application is to a high frequency stock return series, IBM within day returns, with $n = 4{,}626$. In this case, the differencing operation reduces the thickness of the tails considerably. The squared return series shows some evidence of nonexistence of first moment, but the differenced series has much lighter tails. This type of series is more consistent with the theory we have developed here.

7. CONCLUSIONS

We established the consistency and rate of convergence of our estimates. Under some conditions it may be possible to establish the asymptotic normality of $\widehat{m}(x)$ and $\widehat{\beta}$. Assuming this to be the case, one can conduct an inference about the unknown quantities using a bootstrap method based on resampling the errors ε_t. Optimality of the rate we obtain for $\widehat{m}(x)$ is discussed in Chen and Reiss (2010) in an i.i.d. setting.

NOTES

1. It should be noted that the least squares method is not consistent, since the second moment of y_t does not exist.

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

2. Here $a^\top M_\lambda a = \sum_{1\le i,,j\le n} a_i M_{ij}^\lambda a_j = \int_{\mathcal{Y}} \left[\sum_k \lambda_k \sum_{i=1}^n a_i K_g(y_{\tau+i-k} - w)\right]^2 dw \ge 0$, for any $a(\neq 0) \in \mathbb{R}^n$.

3. As in Tautenhahn (1998) and other papers in this literature, we can use the weaker condition that $m_0 \in -\log \mathcal{R}([\mathcal{T}_\lambda^* \mathcal{T}_\lambda]^\eta)$ for some $\eta > 0$ at the cost of a great deal longer proofs.

4. Details of this are available from the second author upon request.

REFERENCES

Blundell, R., X. Chen, & D. Kristensen (2007) Semi-nonparametric IV estimation of shape-invariant Engle curves. *Econometrica* 75, 1613–1669.

Carrasco, M., J.P. Florens, & E. Renault (2006) Linear inverse problems in structural econometrics. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 6, pp. 5633–5751. Elsevier.

Chen, X. & M. Reiss (2010) On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* 27 (this issue).

Chen, X. & X. Shen (1998) Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.

Engle, R.F. & T. Bollerslev (1986) Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1–50.

Hall, P. & J.L. Horowitz (2005) Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* 33, 2904–2909.

Harvey, A., E. Ruiz, & N. Shephard (1994) Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–1264.

Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1174.

Kim, W. (2003) Identification and Estimation of Nonparametric Structural Models by Instrumental Variables Method. Manuscript, Humboldt University.

Lee, S. & B. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Linton, O.B. & E. Mammen (2005) Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* 73, 771–836.

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

Meitz, M. & P. Saikkonen (2004) Ergodicity, Mixing, and Existence of Moments for a Class of Markov Models with Applications to GARCH and ACD Models. Manuscript, Stockholm School of Economics.

Nelson, D. (1990) Stationarity and persistence in the GARCH(1,1) models. *Econometric Theory* 6, 318–334.

Newey W.K. & J.L. Powell (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71(5), 1565–1578.

Rahbek, A. & S.T. Jensen (2004) Asymptotic normality for nonstationary, explosive GARCH. *Econometric Theory* 20, 1203–1226.

Tautenhahn, U. (1998) Optimality for ill-posed problems under general source conditions. *Numerical Functional Analyses and Optimizations* 19, 377–398.

APPENDIX

Proof of Proposition 4.1. The integral operator $\widehat{\mathcal{T}}_\lambda$ has a degenerate kernel, i.e., $\widehat{f}_\lambda(\cdot,\cdot)$ is a finite sum of products of kernel weights on each observation. Noting that $\widehat{\mathcal{T}}_\lambda$ is of a finite rank, the proof is direct from Kim (2003, Prop. 3.1), since $\widehat{\mathcal{T}}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ is uniformly consistent for $\mathcal{T}_\lambda$ on $\mathcal{M}_Y \subset L^2(\mathcal{Y})$ s.t. $\dim(\mathcal{M}_Y) = \infty$. ■

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

LEMMA A.1. *Assume that Conditions 5.1 through 5.5(a) hold. Then, it holds that*

(i) $\sup_{k\geq 2}||\widehat{\mathcal{T}}_k-\mathcal{T}_k||^2=O_p(g_1^{2p_0}+g_2^{2p_0})+O_p\Big(\frac{1}{Tg_2}\Big)$,

(ii) $\sup_{k\geq 2}||\widehat{\mathcal{T}}_k^*-\mathcal{T}_k^*||^2=O_p(g_1^{2p_0}+g_2^{2p_0})+O_p\Big(\frac{1}{Tg_1}\Big)$,

(iii) $\sup_{k\geq 2}||\widehat{h}_k-\widehat{\mathcal{T}}_k m_0||^2_{L^2(\mathcal{Y})}=O_p\Big(\frac{1}{Tg_2}\Big)+O_p(g_1^{2\overline{p}}+g_2^{2p_0})$,

where $\overline{p}=\max(p_0,p_1)$.

Proof. (i) With $a*b$ denoting convolution of a and b, we define

$$m_{c(g_1)}(y)\equiv(K_{g_1}*m)(y)=\int K_{g_1}(y-s)m(s)\,ds, \tag{A.1}$$

and

$$\begin{aligned}f_k^{c(g)}(y,w)&\equiv(K_{(g_1,g_2)}*f_k)(y,w)=\int_{\mathcal{Y}}\int_{\mathcal{Y}}K_{g_1}(s_1-y)K_{g_2}(s_2-w)f_k(s_1,s_2)\,ds_1\,ds_2\\&=\mathrm{E}[K_{g_1}(y_{t-1}-y)K_{g_2}(y_{t-k}-w)].\end{aligned}$$

By adding and subtracting $\int f_k^{c(g)}(y,w)m(y)\,dy$, the estimation errors of $\widehat{\mathcal{T}}_k$ are decomposed into

$$\begin{aligned}(\widehat{\mathcal{T}}_k m-\mathcal{T}_k m)(w)&=\frac{1}{T^*}\sum_{t=\tau+1}^{T}\int[K_{g_1}(y_{t-1}-y)K_{g_2}(y_{t-k}-w)-f_k^{c(g)}(y,w)]m(y)\,dy\\&\quad+\int[f_k^{c(g)}(y,w)-f_k(y,w)]m(y)\,dy\\&\equiv \mathrm{s}_T(w)+\mathrm{B}_T(w),\end{aligned}$$

from which we obtain the mean integrated square error (MISE) of $\widehat{\mathcal{T}}_k m$, given by

$$\mathrm{E}\int_{\mathcal{Y}}\Big[(\widehat{\mathcal{T}}_k-\mathcal{T}_k)m\Big]^2(w)dw=\int_{\mathcal{Y}}\Big\{\mathrm{Var}[\mathrm{s}_T(w)]+\mathrm{E}^2[\mathrm{B}_T(w)]\Big\}dw.$$

From

$$\mathrm{s}_T(w)=\frac{1}{T^*}\sum_{t=\tau+1}^{T}\big\{K_{g_2}(y_{t-k}-w)m_{c(g_1)}(y_{t-1})-\mathrm{E}[K_{g_2}(y_{t-k}-w)m_{c(g_1)}(y_{t-1})]\big\},$$

a standard calculation of the variance term for kernel estimates (under the mixing condition in conditions 5.1) yields

$$\begin{aligned}\mathrm{Var}[s_T(w)]&=\frac{1}{T^*}\mathrm{Var}\big[K_{g_2}(y_{t-k}-w)m_{c(g_1)}(y_{t-1})\big]+o\Big(\frac{1}{Tg_2}\Big)\\&=\frac{1}{Tg_2}||K||_2^2\int_{\mathcal{Y}}m^2_{c(g_1)}(y)f_k(y,w)\,dy+O(T^{-1})+o\Big(\frac{1}{Tg_2}\Big),\end{aligned}$$

where the last equality is due to the dominated convergence theorem. Let $f_0(\cdot)$ be the marginal density function of y_t. Since $f_0(\cdot)$ is assumed to be uniformly bounded (Condition 5.3), it holds that

$$\begin{aligned}\int_{\mathcal{Y}}\mathrm{Var}[s_T(w)]\,dw&\simeq\frac{1}{Tg_2}||K||_2^2\int_{\mathcal{Y}}m^2(y)f_0(y)\,dy\\&\leq C\frac{1}{Tg_2}||K||_2^2||m||^2_{L^2(\mathcal{Y})}.\end{aligned} \tag{A.2}$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

To calculate the bias term, we observe that, by Cauchy-Schwarz inequality,

$$\int_{\mathcal{Y}} \mathrm{B}_T^2(w)\,dw = \int_{\mathcal{Y}} [\int_{\mathcal{Y}} \{f_k^c(y,w) - f_k(y,w)\} m(y)\,dy]^2\,dw$$
$$\leq ||f_k^c - f_k||^2_{L^2(\mathcal{Y}\times\mathcal{Y})} ||m||^2_{L^2(\mathcal{Y})}.$$

Under Condition 5.2 (i.e., $\int |K(s)|ds < \infty$ and $\sup|K(s)| < \infty$), the convolution errors $(||f_k^c - f_k||^2_{L^2(\mathcal{Y}\times\mathcal{Y})})$ converge to zero, as g_1 and g_2 go to zero, for any square integrable $f_k(\cdot,\cdot)$. When there exist p_0th partial derivatives of $f_k(\cdot,\cdot)$ that are continuous and square integrable (i.e., Condition 5.4 holds), we have, by application of the standard Taylor expansion, that

$$\int_{\mathcal{Y}} \mathrm{B}_T^2(w)\,dw \leq C\{g_1^{2q} g_2^{2(p_0-q)} \sum_{q=0}^{p_0} ||\frac{\partial^{p_0} f_k(y,w)}{\partial y^q \partial w^{p_0-q}}||^2_{L^2(\mathcal{Y}\times\mathcal{Y})}\} ||m||^2_{L^2(\mathcal{Y})}. \tag{A.3}$$

Combining (A.2) and (A.3), we obtain

$$\mathrm{E}\int_{\mathcal{Y}} [(\widehat{\mathcal{T}}_k m - \mathcal{T}_k m)(w)]^2 dw$$
$$\leq C||m||^2_{L^2(\mathcal{Y})} \left(g_1^{2q} g_2^{2(p_0-q)} \sum_{q=0}^{p_0} \left\| \frac{\partial^{p_0} f_k(y,w)}{\partial y^q \partial w^{p_0-q}} \right\|^2_{L^2(\mathcal{Y}\times\mathcal{Y})} + \frac{1}{Tg_2} ||K||_2^2 \right),$$

implying

$$\mathrm{E}||\widehat{\mathcal{T}}_k - \mathcal{T}_k||^2_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})}$$
$$= \sup_{m(\neq 0)\in L^2(\mathcal{Y})} \frac{\mathrm{E}\int_{\mathcal{Y}} \left[(\widehat{\mathcal{T}}_k - \mathcal{T}_k)m\right]^2 (w)dw}{||m||^2_{L^2(\mathcal{Y})}}$$
$$\leq C \left(\sum_{q=0}^{p_0} g_1^{2q} g_2^{2(p_0-q)} \left\| |\frac{\partial^{p_0} f_k(y,w)}{\partial y^q \partial w^{p_0-q}} \right\|^2_{L^2(\mathcal{Y}\times\mathcal{Y})} + \frac{1}{Tg_2} ||K||_2^2 \right).$$

Noting that the variance term in (A.2) does not depend on k and $\sup_{k\geq 2} ||\frac{\partial^{p_0} f_k(y,w)}{\partial y^q \partial w^{p_0-q}}||^2_{L^2(\mathcal{Y}\times\mathcal{Y})} \leq C$ (Condition 5.4), we finally get

$$\sup_{k\geq 2} \mathrm{E}||(\widehat{\mathcal{T}}_k - \mathcal{T}_k)m||^2_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} = O(g_1^{2p_0} + g_2^{2p_0}) + O\left(\frac{1}{Tg_2}\right).$$

(ii) By symmetry of the above arguments, a similar result holds for $\widehat{\mathcal{T}}_k^*$.

(iii) Let $r(y_{t-1}) = m_0(y_{t-1}) - m_{0,c(g_1)}(y_{t-1})$, where $m_{0,c(g_1)}(\cdot)$ is defined by (A.1). From $(\widehat{\mathcal{T}}_k m)(w) = T^{*-1}\sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w) m_{c(g_1)}(y_{t-1})$, we get

$$(\widehat{h}_k - \widehat{\mathcal{T}}_k m_0)(w) = \frac{1}{T^*}\sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w)\eta_t + \frac{1}{T^*}\sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w) r(y_{t-1})$$
$$= \frac{1}{T^*}\sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w)\eta_t + \frac{1}{T^*}\sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w)(v_{t-1} - v_{t-1}^c)$$
$$+ \frac{1}{T^*}\sum_{t=\tau+1}^{T} K_{g_2}(y_{t-k} - w)\mathrm{E}(r(y_{t-1})|y_{t-k})$$
$$\equiv s_{1,T}(w) + s_{2,T}(w) + \mathrm{B}_T(w),$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

where $v_{t-1} = m_0(y_{t-1}) - \mathrm{E}(m_0(y_{t-1})|y_{t-k})$, and $v^c_{t-1} = m_{0,c(g_1)}(y_{t-1}) - \mathrm{E}(m_{0,c(g_1)}(y_{t-1})|y_{t-k})$.

As a consequence,

$$\mathrm{E}\int_{\mathcal{Y}}(\widehat{h}_k - \widehat{\mathcal{T}}_k m_0)^2(w)\,dw = \int_{\mathcal{Y}}\left\{\mathrm{Var}\left[s_{1,T}(w)+s_{2,T}(w)\right] + \mathrm{E}^2[\mathrm{B}_T(w)]\right\}dw.$$

By the standard argument in kernel regression, the variance of the main stochastic term is calculated in a straightforward way:

$$\mathrm{Var}\left[s_{1,T}(w)\right] = \frac{1}{Tg_2}||K||_2^2\mathrm{E}(\eta_t^2|y_{t-k}=w)f_0(w)(1+o(1)),$$

leading to

$$\begin{aligned}\int_{\mathcal{Y}}\mathrm{Var}\left[s_{1,T}(w)\right]dw &= \frac{1}{Tg_2}||K||_2^2\left[\int_{\mathcal{Y}}\mathrm{E}(\eta_t^2|y_{t-k}=w)f_0(w)\,dw\right](1+o(1))\\ &= \frac{1}{Tg_2}||K||_2^2\sigma_\eta^2(1+o(1)) = O\left(\frac{1}{Tg_2}\right), \qquad \textbf{(A.4)}\end{aligned}$$

where $\sigma_\eta^2 = \mathrm{E}(\eta_t^2)$. From $\mathrm{E}[v_{t-1}|y_{t-k}] = \mathrm{E}[v^c_{t-1}|y_{t-k}] = 0$, it follows that

$$\mathrm{E}[(v_{t-1}-v^c_{t-1})^2|y_{t-k}] = \mathrm{Var}[r(y_{t-1})|y_{t-k}] \le \mathrm{E}[r^2(y_{t-1})|y_{t-k}],$$

implying, by the mixing assumption and the law of iterated expectation, that

$$\begin{aligned}\mathrm{Var}\left[s_{2,T}(w)\right] &\simeq \frac{1}{T^*}\mathrm{E}\{[K_{g_2}(y_{t-k}-w)(v_{t-1}-v^c_{t-1})]^2\}\\ &= \frac{1}{T^*}\mathrm{E}\{[K_{g_2}(y_{t-k}-w)]^2\mathrm{E}[r^2(y_{t-1})|y_{t-k}]\}\\ &= \frac{1}{T^*}\mathrm{E}\{[K_{g_2}(y_{t-k}-w)r(y_{t-1})]^2\}\\ &= \frac{1}{Tg_2}||K||_2^2\int_{\mathcal{Y}}r^2(y)f_k(y,w)\,dy(1+o(1)).\end{aligned}$$

From boundedness of $f_0(\cdot)$,

$$\begin{aligned}\int_{\mathcal{Y}}\mathrm{Var}\left[s_{2,T}(w)\right]dw &\simeq \frac{1}{Tg_2}||K||_2^2\int_{\mathcal{Y}}r^2(y)f_0(y)\,dy\\ &\le \frac{C}{Tg_2}||K||_2^2||m(\cdot)-m_{c(g_1)}(\cdot)||^2_{L^2(\mathcal{Y})} = o\left(\frac{1}{Tg_2}\right),\end{aligned}$$

since the convolution error, $||m(\cdot)-m_{c(g_1)}(\cdot)||_{L^2(\mathcal{Y})}$, converges to zero, as $g_1 \to 0$. To calculate the bias term, we note, by the dominated convergence theorem, that

$$\begin{aligned}\mathrm{E}[\mathrm{B}_T(w)] &= \mathrm{E}[K_{g_2}(y_{t-k}-w)r(y_{t-1})]\\ &= \int_{\mathcal{Y}}[m(y)-m_{c(g_1)}(y)]f_k(y,w)\,dy(1+o(1)). \qquad \textbf{(A.5)}\end{aligned}$$

Letting $f_k^{c(g_1)}(u,w) = \int_{\mathcal{Y}}K_{g_1}(y-u)f_k(y,w)\,dy$, we obtain an alternative form of the bias such that

$$\mathrm{E}[\mathrm{B}_T(w)] = \{\int_{\mathcal{Y}}\left[f_k(u,w) - f_k^{c(g_1)}(u,w)\right]m(u)\,du\}(1+o(1)), \qquad \textbf{(A.6)}$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

since $\int_{\mathcal{Y}} m_{c(g_1)}(z) f_k(z,w)dz = \int_{\mathcal{Y}} f_k^{c(g_1)}(u,w)m(u)du$, by Fubini's theorem. By Cauchy-Schwarz inequality, it follows from (A.5) and (A.6), together with square-integrability of $f_k(\cdot,\cdot)$ and $m(\cdot)$, that

$$\int_{\mathcal{Y}} \mathrm{E}^2[\mathrm{B}_T(w)]dw \le C\min\{||m(\cdot) - m_{c(g_1)}(\cdot)||^2, ||f_k(\cdot) - f_k^{c(g_1)}(\cdot)||^2\},$$

which, by the standard method of Taylor expansion (under Condition 5.4), gives

$$\int_{\mathcal{Y}} \mathrm{E}^2[\mathrm{B}_T(w)]dw = O(g_1^{\max\{2p_0,2p_1\}} + g_2^{2p_0}).$$

Since the variance term in (A.4) does not depend on k and $\sup_{k\ge 2} ||\frac{d^{p_1}m(y)}{dy^{p_1}}||^2_{L^2(\mathcal{Y})} \le C$ (Condition 5.4), it holds that

$$\sup_{k\ge 2} \mathrm{E}\int_{\mathcal{Y}} (\widehat{h}_k - \widehat{\mathcal{T}}_k m_0)^2(w)\,dw = O\left(\frac{1}{Tg_2}\right) + O\left(g_1^{2\overline{p}} + g_2^{2p_0}\right). \qquad \blacksquare$$

Proof of Proposition 5.1. (i) By the triangle inequality,

$$\begin{aligned} ||\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} &\le \sum_{k=2}^{\tau} \lambda_k ||\widehat{\mathcal{T}}_k - \mathcal{T}_k||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} + \sum_{k=\tau+1}^{\infty} \lambda_k ||\mathcal{T}_k||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \\ &\le C\{\sup_{k\ge 2} ||\widehat{\mathcal{T}}_k - \mathcal{T}_k||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} + o(1/\sqrt{T})\}, \end{aligned}$$

where the second inequality follows from $\sup_{k\ge 2} ||\mathcal{T}_k|| \le \sup_{k\ge 2} ||f_k|| \le C$ (Condition 5.3) and $\sum_{k=\tau+1}^{\infty} \lambda_k = o(1/\sqrt{T})$ (Condition 5.6). Now, the proof is immediate from Lemma A.1(i).

The proof of (ii) can be shown in the same way.

(iii) Note that

$$||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0|| \le \sum_{k=2}^{\tau} \lambda_k ||\widehat{h}_k - \widehat{\mathcal{T}}_k m_0|| \le \sup_{k\ge 2} ||\widehat{h}_k - \widehat{\mathcal{T}}_k m_0||,$$

since $\sum_{k=2}^{\tau} \lambda_k \le \sum_{k=2}^{\infty} \lambda_k = 1$. The proof follows, if we apply Lemma A.1(iii). $\blacksquare$

Before proving the main results, we need to introduce some useful lemmas that are borrowed from Kim (2003). First, note that $U_\alpha(\kappa) = (\alpha + \kappa)^{-1}$ satisfies the conditions below.

Condition A.1. Let $\overline{\kappa} \equiv \sup_{n\ge n_0} ||\widehat{\mathcal{T}}_\lambda^* \widehat{\mathcal{T}}_\lambda||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})}$. A parameter dependent family of continuous functions, $\{U_\alpha(\cdot)\}_{\alpha>0}$, defined on $(0,\overline{\kappa}]$, satisfy that (i) $\sup_{\kappa\in(0,\overline{\kappa}]} |U_\alpha(\kappa)\kappa| \le C < \infty$, for $\alpha > 0$, (ii) $\lim_{\alpha\to 0^+} U_\alpha(\kappa) = \frac{1}{\kappa}$, for all $\kappa \in (0,\overline{\kappa}]$, and (iii) $\sup_{\kappa\in(0,\overline{\kappa}]} |U_\alpha(\kappa)| = O(\frac{1}{\alpha})$, as $\alpha \to 0^+$.

Condition A.2. Given $U_\alpha : (0,\overline{\kappa}] \to \mathbb{R}$, it holds for any $\mu \in (0,1]$ that $\sup_{\kappa\in(0,\overline{\kappa}]} \kappa^\mu |U_\alpha(\kappa)\kappa - 1| \le C\alpha^\mu$, for any $\alpha \in (0,\alpha_0)$, where $\alpha_0 > 0$.

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

LEMMA A.2. *If $\alpha = \alpha(T) \to 0$ as $T \to \infty$, then*

(i) $||U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} = O_{a.s}(\alpha^{-1})$,

(ii) $||U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} = O_{a.s}(\sqrt{\alpha}^{-1})$,

Assume additionally that $\widehat{\mathcal{T}}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ converges pointwise, in probability, to $\mathcal{T}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$, which is bounded and one-to-one. Then,

(iii) $||[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I]m||_{L^2(\mathcal{Y})} = o_p(1)$, *for all $m \in L^2(\mathcal{Y})$.*

Proof. Since $U_\alpha(\cdot)$ satisfies Condition A.1 and $\widehat{\mathcal{T}}_\lambda$ has a finite rank, the results are immediate from Kim (2003, Lem. 3.2, p. 41). ■

LEMMA A.3. *Let $G : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ be a linear bounded operator and $G^* : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ be adjoint to G. Then, for all $m \in L^2(\mathcal{Y})$,*

(i) $||[U_\alpha(G^*G)G^*G - I](G^*G)^\mu||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \le C\alpha^{\min(\mu,1)}$, *for $\mu > 0$.*

(ii) $||[U_\alpha(G^*G)G^*G - I]G^*||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \le C\alpha^{1/2}$, *for $\mu > 0$.*

Proof. Since U_α satisfies Condition A.2, the results follow from Kim (2003, Lem. 3.3, p. 42). See also Carrasco et al. (2006). ■

For $\mathcal{M}_T(\mathcal{Y}) = \{m \in \mathcal{M}(\mathcal{Y}) : m_T(x) = T^{*-2}K_T^Y(x)^\top b_T$, for $b_T \in \mathbb{R}^{T^*}\}$, we define $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T} : \mathcal{M}_T \to C(\mathcal{Y})$ to be the restriction (into $\mathcal{M}_T$) of $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega : L^2(\mathcal{Y}) \to C(\mathcal{Y})$, where $C(\mathcal{Y})$ is a space of continuous functions defined on $\mathcal{Y}$.

LEMMA A.4. *Assume that $\widehat{f}_k(\cdot,\cdot)$ is a uniformly consistent estimate of $f_k(\cdot,\cdot)$ that is continuous with a compact support. Then, $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T} : \mathcal{M}_T \to C(\mathcal{Y})$ is uniformly bounded in the sense that*

$$||(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T}||_{L^2(\mathcal{Y})\to C(\mathcal{Y})} = \sup_{m\in\mathcal{M}_T}\sup_{x\in\mathcal{Y}}\frac{|(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega m(x)|}{||m||_{L^2(\mathcal{Y})}} = O_p(1),$$

where $0 < \omega \le 1$.

Proof. By applying the uniform boundedness principle (or the Banach-Steinhaus principle), the assertion follows if we show that, for each $m_T \in \mathcal{M}_T \subset L^2(\mathcal{Y})$ with $||m||_{L^2(\mathcal{Y})} = 1$,

$$\sup_{x\in\mathcal{Y}}|(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega m_T(x)| = O_p(1).$$

Observing that $Q_{Y,\lambda}$ is symmetric nonnegative semidefinite, we can show that, for $m_T(\cdot) \in \mathcal{M}_T(\mathcal{Y})$,

$$\begin{aligned}\sup_{x\in\mathcal{Y}}|(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T} m_T(x)| &= \sup_{x\in\mathcal{Y}}|T^{*-2}K_T^Y(x)^\top M_\lambda^{1/2}Q_{Y,\lambda}^\omega M_\lambda^{-1/2}b_T| \\ &\le C\sup_{x\in\mathcal{Y}}|T^{*-2}K_T^Y(x)^\top M_\lambda^{1/2}Q_{Y,\lambda}M_\lambda^{-1/2}b_T| \\ &= C\sup_{x\in\mathcal{Y}}|(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)_{|\mathcal{M}_T} m_T(x)|, \quad \text{for all } T.\end{aligned}$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

By the Cauchy-Schwarz (CS) inequality and boundedness of $f_k(\cdot,\cdot)$, it holds that

$$\begin{aligned}
&\sup_x |\,(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda^*\mathcal{T}_\lambda)m_T(x)| \\
&\quad = \sup_x |\int_{\mathcal{Y}}\int_{\mathcal{Y}} \sum_{k=2}^{\tau}\sum_{l=2}^{\tau} \lambda_k\lambda_l[\widehat{f}_k(x,w)\widehat{f}_l(z,w) - f_k(x,w)f_l(z,w)]m_T(z)\,dw\,dz| \\
&\qquad + \sum_{k=\tau+1}^{\infty}\sum_{l=\tau+1}^{\infty} \lambda_k\lambda_l \int_{\mathcal{Y}}\int_{\mathcal{Y}} |f_k(x,w)f_l(z,w)m_T(z)|\,dw\,dz \\
&\quad \le \sup_x \sum_{k=2}^{\tau}\sum_{l=2}^{\tau} \lambda_k\lambda_l ||\widehat{f}_k(x,\cdot)\widehat{f}_l(\cdot,\cdot) - f_k(x,\cdot)f_l(\cdot,\cdot)||_{L^2(\mathcal{Y})\times L^2(\mathcal{Y})} ||m_T||_{L^2(\mathcal{Y})} \\
&\qquad + C\sum_{k=\tau+1}^{\infty}\sum_{l=\tau+1}^{\infty} \lambda_k\lambda_l \\
&\quad \le C \sup_{x,z,w} \sum_{k=2}^{\tau}\sum_{l=2}^{\tau} \lambda_k\lambda_l |\widehat{f}_k(x,w)\widehat{f}_l(z,w) - f_k(x,w)f_l(z,w)| + o(1) = o_p(1),
\end{aligned}$$

where the last inequality comes from the compact support assumption, uniform consistency of the kernel density estimate $\widehat{f}_k(\cdot,\cdot)$, and Condition 5.6. In a similar way, we can show, by the CS inequality and boundedness of $f_k(\cdot,\cdot)$, that, for any $m \in L^2(\mathcal{Y})$,

$$\begin{aligned}
\sup_x |\mathcal{T}_\lambda^*\mathcal{T}_\lambda m(x)| &\le \sup_x \sum_{k=2}^{\infty}\sum_{l=2}^{\infty} \lambda_k\lambda_l ||f_k(x,\cdot)f_l(\cdot,\cdot)||_{L^2(\mathcal{Y})\times L^2(\mathcal{Y})} ||m_T||_{L^2(\mathcal{Y})} \\
&\le C \sup_{x,z,w} \sum_{k=2}^{\infty}\sum_{l=2}^{\infty} \lambda_k\lambda_l |f_k(x,w)f_l(z,w)| < \infty.
\end{aligned}$$

Hence, it follows that

$$\sup_x |\,\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda m_T(x)| \le C \sup_x |\,\mathcal{T}_\lambda^*\mathcal{T}_\lambda m_T(x)| < \infty,$$

implying that

$$\sup_{x\in\mathcal{Y}} |(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega m_T(x)| < \infty,$$

as required. ■

Proof of Theorem 5.2. (i) We use the error decomposition

$$\widehat{m}_\alpha - m_0 = U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*(\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0) + [U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I]m_0. \tag{A.7}$$

By Proposition 5.1(i), $\widehat{\mathcal{T}}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$ is a consistent estimator for the true operator $\mathcal{T}_\lambda : L^2(\mathcal{X}) \to L^2(\mathcal{W})$ that is bounded and one-to-one. Also, under the bandwidth conditions in Condition 5.5, it holds by Proposition 5.1(iii) that $||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0||_{L^2(\mathcal{Y})}/\sqrt{\alpha} \xrightarrow{p} 0$, as $T \to \infty$. Now, the assertion follows, by Lemma A.2(ii) and (iii) to (A.7).

(ii) Let $m_1 = (\mathcal{T}_\lambda^*\mathcal{T}_\lambda)^{-1}m_0$. The error decomposition in this case takes the form of

$$\begin{aligned}
&U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*(\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0) + [(U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda]m_1 \\
&-[(U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I)\widehat{\mathcal{T}}_\lambda^*](\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda)m_1 - (U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I)(\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*)\mathcal{T}_\lambda m_1.
\end{aligned}$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

By Lemma A.2(ii), L^2-norm of the first term is bounded by $\frac{C_1}{\sqrt{\alpha}}||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0||_{L^2(\mathcal{W})}$, almost surely. By applying Lemma A.3(i) and (ii), we get:

$$||(U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda|| \le C\alpha, \quad \text{a.s.}$$
$$||(U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I)\widehat{\mathcal{T}}_\lambda^*|| \le C\alpha^{1/2}, \quad \text{a.s.}$$

Also, by Lemma A.2.(iii), it holds that $(U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I) \le C$, which, together with the results of Proposition 5.1, completes the proof. ■

Proof of Theorem 5.3. Let $m_1 = (\mathcal{T}_\lambda^*\mathcal{T}_\lambda)^{-1}m_0$. From the proof of Theorem 5.2(ii),

$$\begin{aligned}
&U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*(\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0) + [U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I]\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda m_1 \\
&\quad -[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I]\widehat{\mathcal{T}}_\lambda^*(\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda)m_1 \\
&\quad -U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda(\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*)\mathcal{T}_\lambda m_1 + (\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*)\mathcal{T}_\lambda m_1 \\
&\quad \equiv \xi_{1T} + \xi_{2T} + \xi_{3T} + \xi_{4T} + \xi_{5T}.
\end{aligned}$$

Let $\omega > 0$ be any fixed real number. In the proof of Theorem 5.2 (step ii), we showed the spectral representation of $r(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)$ from which follows

$$(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega m(x) = T^{*-2}K_T^Y(x)^\top M_\lambda^{1/2}Q_{Y,\lambda}^{\omega-1}M_\lambda^{1/2} < K_T^Y, m >_{L^2(\mathcal{Y})}.$$

Since, for each T, $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega m$ is a continuous function for any $m \in L^2(\mathcal{Y})$, it holds that

$$\mathcal{R}[(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega] \subset C(\mathcal{Y}), \quad \text{for each } T,$$

i.e., $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega : L^2(\mathcal{Y}) \to C(\mathcal{Y})$ is well defined. Let $\mathcal{M}_T(\mathcal{Y}) = \{m \in L^2(\mathcal{Y}) : m_T(x) = T^{-2}K_T^Y(x)^\top b_T$, for $b_T \in \mathbb{R}^T\}$. Define $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T} : \mathcal{M}_T \to C(\mathcal{Y})$. Note that, under the given conditions, $\widehat{f}_k(\cdot,\cdot)$ is the uniformly consistent estimate of $f_k(\cdot,\cdot)$. Consequently, by Lemma A.4, $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T} : \mathcal{M}_T \to C(\mathcal{Y})$ is uniformly bounded in the sense that

$$||(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T}||_{L^2(\mathcal{Y})\to C(\mathcal{Y})} = \sup_{m\in\mathcal{M}_T}\sup_{x\in\mathcal{Y}}\frac{|\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda m(x)|}{||m||_{L^2(\mathcal{Y})}} = O_p(1).$$

From the definition of the operator norm and $\mathcal{R}\{U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{-\omega}\widehat{\mathcal{T}}_\lambda^*\} \subset \mathcal{M}_T$, we obtain

$$\begin{aligned}
\sup_{x\in\mathcal{Y}}|\xi_{1T}(x)| &= \sup_{x\in\mathcal{Y}}|U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*(\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0)(x)| \\
&= \sup_{x\in\mathcal{Y}}|(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{-\omega}\widehat{\mathcal{T}}_\lambda^*(\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0)](x)| \\
&\le ||(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T}||_{L^2(\mathcal{Y})\to C(\mathcal{Y})}||U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{-\omega}\widehat{\mathcal{T}}_\lambda^*||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \\
&\quad \times ||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0||_{L^2(\mathcal{Y})} \\
&\le C\alpha^{-1/2-2\omega}||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0||_{L^2(\mathcal{Y})} \\
&= O_p(\alpha^{-1/2-2\omega}[1/\sqrt{Tg_2} + g_1^{p_0}]), \quad \text{for any } \omega > 0,
\end{aligned}$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

where the last equality comes from Proposition 5.1. Since we assume that $\alpha = o(1/(\log T)^c)$, for any $c > 0$, we get, by letting $c = 1/2\omega$,

$$\sup_{x\in\mathcal{Y}} |\xi_{1T}(x)| = O_p(\alpha^{-1/2}[1/\sqrt{Tg_2} + g_1^{p_0}]\log T).$$

By the same arguments, we have

$$\begin{aligned}
\sup_{x\in\mathcal{Y}} |\xi_{4T}(x)| &= \sup_{x\in\mathcal{Y}} |(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{1-\omega}(\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*)\mathcal{T}_\lambda m_1](x)| \\
&\le ||(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T}||_{L^2(\mathcal{Y})\to C(\mathcal{Y})}||U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{1-\omega}||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \\
&\quad \times ||(\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*)\mathcal{T}_\lambda m_1||_{L^2(\mathcal{Y})} \\
&= O_p(\alpha^{-\omega}[1/\sqrt{Tg_1} + g_1^{p_0} + g_2^{p_0}]) = O_p([1/\sqrt{Tg_1} + g_1^{p_0} + g_2^{p_0}]\log T),
\end{aligned}$$

$$\begin{aligned}
\sup_{x\in\mathcal{Y}} |\xi_{2T}(x)| &= \sup_{x\in\mathcal{Y}} |(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega\{[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I](\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{1-\omega} m_1\}(x)| \\
&\le ||(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T}||_{L^2(\mathcal{Y})\to C(\mathcal{Y})}||[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I](\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{1-\omega} \\
&\quad \times ||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})}||m_1||_{L^2(\mathcal{Y})} \\
&\le C\alpha^{1-2\omega}||m_1||_{L^2(\mathcal{Y})} = O_p(\alpha\log T),
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{x\in\mathcal{Y}} |\xi_{3T}(x)| \\
&\quad = \sup_{x\in\mathcal{Y}} |(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega\{[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I](\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{-\omega}\widehat{\mathcal{T}}_\lambda^*(\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda)m_1\}(x)| \\
&\quad \le ||(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega_{|\mathcal{M}_T}||_{L^2(\mathcal{Y})\to C(\mathcal{Y})}||[U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I](\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{-\omega}\widehat{\mathcal{T}}_\lambda^*||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} \\
&\qquad \times ||(\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda)m_1||_{L^2(\mathcal{Y})} \\
&\quad \le C\alpha^{1/2-2\omega}||(\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda)m_1||_{L^2(\mathcal{Y})} = O_p(\alpha^{1/2}[1/\sqrt{Tg_2} + g_1^{p_0} + g_2^{p_0}]\log T).
\end{aligned}$$

For the uniform convergence rate of $\xi_{4T}(x) = (\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*)h_1$, we note that $\widehat{\mathcal{T}}_\lambda^* h_1$ is equivalent (up to some bias term) to the standard one-dimensional kernel estimate, where $h_1 = \mathcal{T}_\lambda m_1$. Hence, its uniform convergence rate follows from application of Masry (1996):

$$\sup_{x\in\mathcal{Y}} |\xi_{3T}(x)| = O_p([1/\sqrt{Tg_1} + g_1^{p_0} + g_2^{p_0}]\log T). \qquad \blacksquare$$

Proof of Theorem 5.4. First, we show that $\widehat{\beta}$ is consistent, which follows from the consistency of $\widetilde{\beta}$ and the uniform approximation of $\widehat{\ell}(\beta)$ by $\ell(\beta)$ over B. Then, we establish that $\widehat{\ell}(\beta) - \ell(\beta) = o_p(T^{-1/4})$ uniformly over a shrinking neighborhood of β_0. The argument is fairly standard so we just sketch the second part here. We have

$$|\widehat{\ell}(\beta) - \ell(\beta)| \le \frac{1}{T}\sum_{t=1}^T \left|\ln \frac{\sigma_t^2(\beta, \widehat{m}_\alpha)}{\sigma_t^2(\beta, m_0)}\right| + \frac{1}{T}\sum_{t=1}^T y_t^2 \frac{\left|\sigma_t^2(\beta, \widehat{m}_\alpha) - \sigma_t^2(\beta, m_0)\right|}{\sigma_t^2(\beta, m_0)\sigma_t^2(\beta, \widehat{m}_\alpha)}, \tag{A.8}$$

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press

where

$$\begin{aligned}
&\max_{\tau_\dagger \le t \le T} \left| \sigma_t^2(\beta, \widehat{m}_\alpha) - \sigma_t^2(\beta, m_0) \right| \\
&\quad \le \sum_{j=1}^{\infty} \beta^{j-1} \sup_{y \in \mathcal{Y}} \left| \widehat{m}_\alpha(y) - m_0(y) \right| \\
&\qquad + \sum_{j=\tau_\dagger+1}^{\infty} \beta^{j-1} [(1-\beta) \max_{\tau_\dagger \le t \le T} y_{t-j}^2 + \max_{\tau_\dagger \le t \le T} m(y_{t-j})] \\
&\quad = o_p(T^{-1/4}),
\end{aligned}$$

since y_s and $m(y_s)$ are bounded by assumption. Using $\ln(1+\delta) \le |\delta|$ for $|\delta| < 1$, we can bound the first term in (A.8) with probability tending to one by

$$\left| \frac{1}{T} \sum_{t=1}^{T} \ln \frac{\sigma_t^2(\beta, \widehat{m}_\alpha)}{\sigma_t^2(\beta, m_0)} \right| \le \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\sigma_t^2(\beta, \widehat{m}_\alpha) - \sigma_t^2(\beta, m_0)}{\sigma_t^2(\beta, m_0)} \right| = o_p(T^{-1/4}),$$

since $\min_{\tau_\dagger \le t \le T} \sigma_t^2(\beta, m_0)$ is bounded away from zero. Likewise the second term in (A.8) is $o_p(T^{-1/4})$. ■

https://doi.org/10.1017/S0266466610000435 Published online by Cambridge University Press