ESTIMATION OF A SEMIPARAMETRIC IGARCH(1,1) MODEL*

Woocheol Kim† and Oliver Linton‡

February 4, 2005

Abstract

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.

Key words: Inverse Problem; Instrumental Variable; IGARCH; Kernel Estimation; Nonparametric regression.

Journal of Economic Literature Classification: C14

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 GARCH or IGARCH process which imposes this restriction, see for example Engle and Bollerslev (1986). 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 variance can be strongly stationary, see Nelson (1990). In fact, IGARCH processes can also be strongly mixing, see Meitz and Saikonnen (2004).

We propose a semiparametric extension of the IGARCH model. Our model nests the JP Morgan model and the standard IGARCH(1,1) model, but it allows more flexibility in functional form. It

*We would like to thank Joel Horowitz for interesting discussions. We thank the ESRC for financial support.

†Korea Institute of Public Finance, 79-6 Garak-Dong, Songpa-Gu, Seoul, Republic of Korea, 138-774; email: wkim@kipf.re.kr

‡Department of Economics, The London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: lintono@lse.ac.uk.

extends the recent model of Linton and Mammen (2003) 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. We establish the rate of uniform convergence of the nonparametric part of our model and the consistency of the parametric part.

2 Model

Let $\{y_t\}$ satisfy

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

where ε_t is a martingale difference sequence with conditional mean zero and conditional variance one, 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 JP Morgan model. 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 a nonnegativity constraint $(m(\cdot) \geq 0)$ and some additional conditions guaranteeing strong stationarity of y_t. The nonparametric term is introduced to correct possible misspecification with a quadratic growth function of news impact on volatility.

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

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

The squared returns $\{y_t^2\}$ 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 a similar dynamics to the standard IGARCH.

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 is a martingale difference sequence. While it is true that $E[\eta_t | \mathcal{F}_{t-1}] = 0$, this does not imply that $\lim_{t\to\infty} E[|\eta_t| | \mathcal{F}_0] < \infty$ and so 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 $\lim_{t\to\infty} E[\eta_t^2|\mathcal{F}_0] < \infty$. 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) behaviour is due to the 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 semi-strong 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 mutually independent random variables with mean zero and variance one. It follows that $\eta_t = (u_t^2 - 1)\,(\sigma_t^2/(1+\sigma_t^2))$ which satisfies $E(|\eta_t|) < \infty$, and so we have $E(|\Delta y_t^2|) < \infty$. However, $E(\sigma_t^2) = E(y_t^2) = \infty$. 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 $E([\Delta y_t^2]^2) < \infty$.

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 datasets 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 (2003) who considered the case with $\sigma_t^2 = \beta\sigma_{t-1}^2 + m(y_{t-1})$ and $E(y_t^2) < \infty$. 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 $f(\cdot)$ 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, where $k \geq 2$. 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 $E(\nu_t|y_{t-1}) \neq 0$ but $E(\nu_t|y_{t-k}) = 0$ for $k \geq 2$; (3) is an example of nonparametric structural models with *infinite* number of instruments. By the finite moment condition ($E(|\Delta y_t^2|) < \infty$) and the law of iterated expectations, we obtain, from taking conditional expectations of (3),

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

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 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

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

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

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

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 $E[m(y_{t-1})|y_{t-k}] = 0$, a.s., only for $m \equiv 0$. A sufficient condition for completeness is that the conditional CDF, $F_{1|k}(\cdot|\cdot)$, is a member of exponential family, satisfying certain regularity conditions given in Newey and Powell (2003). In Blundell, Chen, and Kristensen (2003), 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 parametric approaches cannot be applied here. In Kim (2003), a more general identifying condition is suggested that does not rely on a parametric distributional assumption. Our identification result below is based on the approach of Kim (2003). By combining the equations (5) we can obtain a family of integral equations. 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 $\sum_{k=2}^{\infty} \lambda_k = 1$. 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 family of equations.

Definition 3.1 *The true function $m_0(\cdot)$ is identifiable if the solution to the following 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 some weighting scheme $\lambda = \{\lambda_k\}_{k=2}^{\infty}$ such that $\sum_{k=2}^{\infty} \lambda_k = 1$.

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 following condition.

A.1. *There exists $\lambda = \{\lambda_k\}_{k=2}^\infty$, satisfying $\lambda_k \geq 0$ and $\sum_{k=2}^\infty \lambda_k = 1$, such that, for some sequence $\overline{\mathcal{Y}} = \{\omega_l\}_{l=1}^\infty \subset Y$, $\text{lin}\{f_k(\cdot,\omega_l)\}_{l=1}^\infty$ is dense in $L^2(Y)$, i.e., $\overline{\text{lin}}\{f_k(\cdot,\omega_l)\}_{l=1}^\infty = L^2(\mathcal{Y})$.*

The above condition concerns richness of the linear spaces that are generated by a weighted sum of (unnormalized) conditional density function. A.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 A.1 is sufficient for $\mathcal{T}_\lambda$ to be one-to-one.

Proposition 3.1 (i) *If A.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 proof is immediate from Kim (2003, Theorem 2.2 (i), p.7).

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 $L^2(\mathcal{Y})$. A.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 1. Since $E(\eta_t|\mathcal{F}_{t-k}) = E(\eta_{t-1}|\mathcal{F}_{t-k}) = 0$, 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 A.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 1, the GARCH coefficient β_0 can be identified by

$$\arg\min_{\beta\in B} \ \mathrm{E}[\ln \sigma_t^2(\beta, m_0) + \frac{y_t^2}{\sigma_t^2(\beta, m_0)}],$$

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

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

Here, the least squares method is not consistent, since the second moment of y_t does not exist.

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

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

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

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

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* solutions. One resolve the existence and uniqueness problems easily by using the (Moore-Penrose) generalized inverse of $\widehat{\mathcal{T}}_\lambda$:

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

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 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 $\mathcal{T}_\lambda$ 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,*

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

Proof in the appendix.

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.

4.1.1 Tikhonov Regularization Method

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,

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

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 $\dim(\mathcal{R}(\widehat{\mathcal{T}}_\lambda)) \leq T$, 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 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

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

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

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 semi-definite 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.[1] 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 semi-definite matrix, whose eigenvalues are all real and positive. We denote, by $\lambda_{\max}(Q_{Y,\lambda})$, the maximum of those eigenvalues.

Theorem 4.2 *For some $T_0 \geq 1$, it holds that, for any $T \geq T_0$,*

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

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

Proof in the appendix.

By Theorem 4.2, the abstract operator-form of the kernel estimator translates into a concrete matrix-form. Computations of $\widehat{m}_\alpha$ only involve simple operation of 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.

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

(ii) By the spectral representation results in the proof of Theorem 4.2, 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

$$\begin{aligned}\widehat{\ell}(\beta) &= \frac{1}{T}\sum_{t=1}^{T} \ln \sigma_t^2(\beta, \widehat{m}_\alpha) + \frac{y_t^2}{\sigma_t^2(\beta, \widehat{m}_\alpha)} \\ \sigma_t^2(\beta, \widehat{m}_\alpha) &= \sum_{j=1}^{\min\{t-1,\tau\}} \beta^{j-1}[(1-\beta)y_{t-j}^2 + \widehat{m}_\alpha(y_{t-j})].\end{aligned}$$

Here, $\tau = \tau(T) < T$ is a 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 $\{y_t\}$

which is defined by

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

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

C.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.*

C.3 *The joint density functions* $f_k(\cdot,\cdot)$ *is square-integrable and bounded;*

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

C.4 $f_k(\cdot,\cdot)$ *and* $m_0(\cdot)$ *have continuous* p_0*-th and* p_1*-th partial derivatives, respectively, that are square-integrable, satisfying*

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

C.5 (a) *The bandwidth parameters* $(g_1,\ g_2)$ *satisfy that* $\max(g_1,\ g_2) \to 0$*,* $Tg_2 \to \infty$*. (a) The regularization parameter* α *satisfies that* $\alpha \to 0$*,* $Tg_2\alpha \to \infty$*, and* $g_1^{p_0}/\sqrt{\alpha} \to 0$*, as* $T \to \infty$*.*

C.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 C.2 through C.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 a compact support or be bounded away from zero. For uniform convergence results, however, the conditions in C.3 and C.4 will be strengthened, being replaced by a compact support assumption, together with the continuity condition. Note that the square-integrability condition in C.3 entails boundedness of the linear operator $\mathcal{T}_k$. C.5(b), which is rather stronger than C.5(a), is necessary for consistency of the regularized kernel estimates. C.6 gives a convenient condition for controlling the approximation errors due to truncation. C.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 $\widehat{\mathcal{T}}_k$ 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 C.1 through C.5(a) and C.6 hold. Then*

$$\begin{aligned}
\text{(i)}\ ||\widehat{\mathcal{T}}_\lambda - \mathcal{T}_\lambda||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} &= O_p(1/\sqrt{Tg_2} + g_1^{p_0} + g_2^{p_0}),\\
\text{(ii)}\ ||\widehat{\mathcal{T}}_\lambda^* - \mathcal{T}_\lambda^*||_{L^2(\mathcal{Y})\to L^2(\mathcal{Y})} &= O_p(1/\sqrt{Tg_1} + g_1^{p_0} + g_2^{p_0}),\ \textit{and}\\
\text{(iii)}\ ||\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0||_{L^2(\mathcal{Y})} &= O_p(1/\sqrt{Tg_2} + g_1^{\overline{p}}),
\end{aligned}$$

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

Proof in the appendix.

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, 2003, Theorem 3.3), together with Proposition 5.1.

Theorem 5.2 *Assume that C.1 through C.6 hold, with* $p_0 = p_1$. *Then,*

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

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

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

Needs a discussion about $m_0 \in \mathcal{R}(\mathcal{T}_\lambda^* \mathcal{T}_\lambda)$. Since $\mathcal{T}_\lambda$ is an integral operator, this 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 $\mathcal{T}_\lambda$.

Remark 5.1. (the Optimal Convergence Rate) Let m_0 be any function in $\mathcal{R}(\mathcal{T}_\lambda^* \mathcal{T}_\lambda)$. Suppose a side condition on (g_1, g_2) such that $(Tg_1)^{-1/2} \leq O(g_1^{2p_0/3})$, and $g_2^{3p_0/2} \leq O([Tg_2]^{-1/2})$. Then, the optimal convergence rate of $\widehat{m}_\alpha$ is given by

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

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

Theorem 5.3 (uniform convergence rate of $\widehat{m}_\alpha(x)$) *Assume that C.1 through C.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 $\mu \geq 1$. If $\alpha = o(1/(\log T)^c)$, for any $c > 0$, then,*

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

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

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

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

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

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 Section 4.4 in Linton and Mammen (2003), since $\widehat{m}_\alpha$ does not depend on $\widehat{\beta}$. First define

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

and suppose that $\widetilde{\beta}$ is the unique minimizer of $\ell(\beta)$, while β_0 is the unique minimizer of $E[\ell(\beta)]$. The properties of $\widetilde{\beta}$ have been well studied in the literature, see for example Lee and Hansen (1994) and Rahbek and Jensen (2004).

Theorem 5.4. *Assume that C.1 through C.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 $\tau(T) = c \log T$, then,*

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

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

6 Numerical Results

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&P500 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

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

In Figure 1 we give the Hill plot with 95% confidence interval for y_t^2.

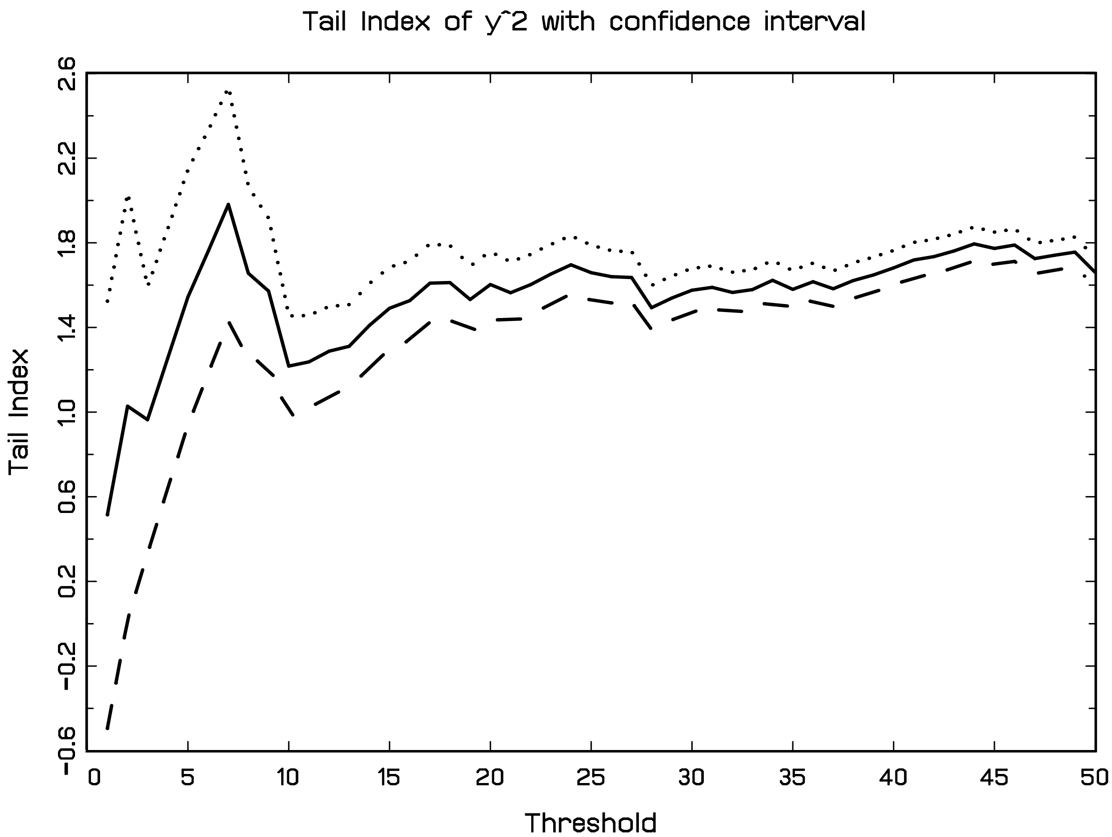

Figure 1. Shows the value of $\widehat{\kappa}$ against threshold size M for the series y_t^2.

It is generally above one but less than two implying that $E[y_t^2] < \infty$ but $E[y_t^4] = \infty$. The corresponding hill plots for $|\Delta y_t^2|$ show slightly lighter tails: we just show the ratio of $\widehat{\kappa}$ for $|\Delta y_t^2|$ to $\widehat{\kappa}$ for y_t^2, which is generally above one.

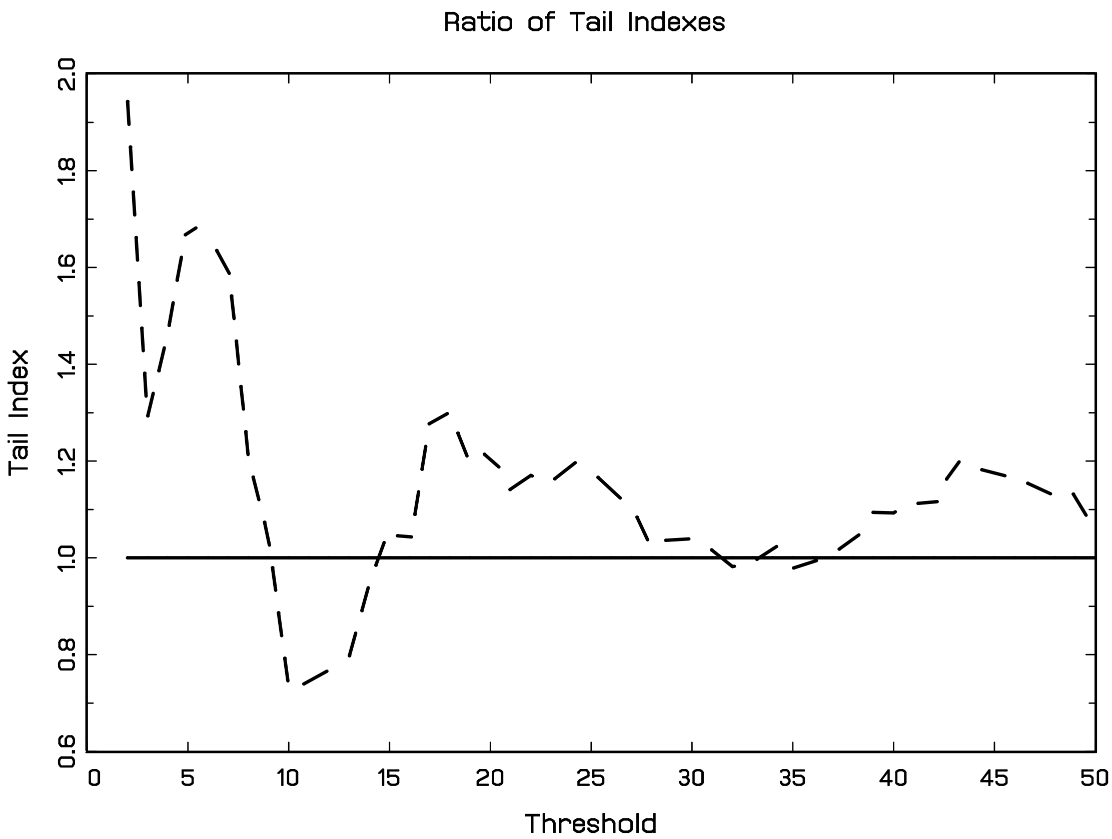

Figure 2. Shows the ratio of $\widehat{\kappa}$ for the series $|\Delta y_t^2|$ to $\widehat{\kappa}$ for the series y_t^2 against threshold size M.

For these data at least the evidence of integrated process is weak although it does seem that differencing reduces the weight of the tails significantly. Our second application is to a high frequency stock return series with n =4626.

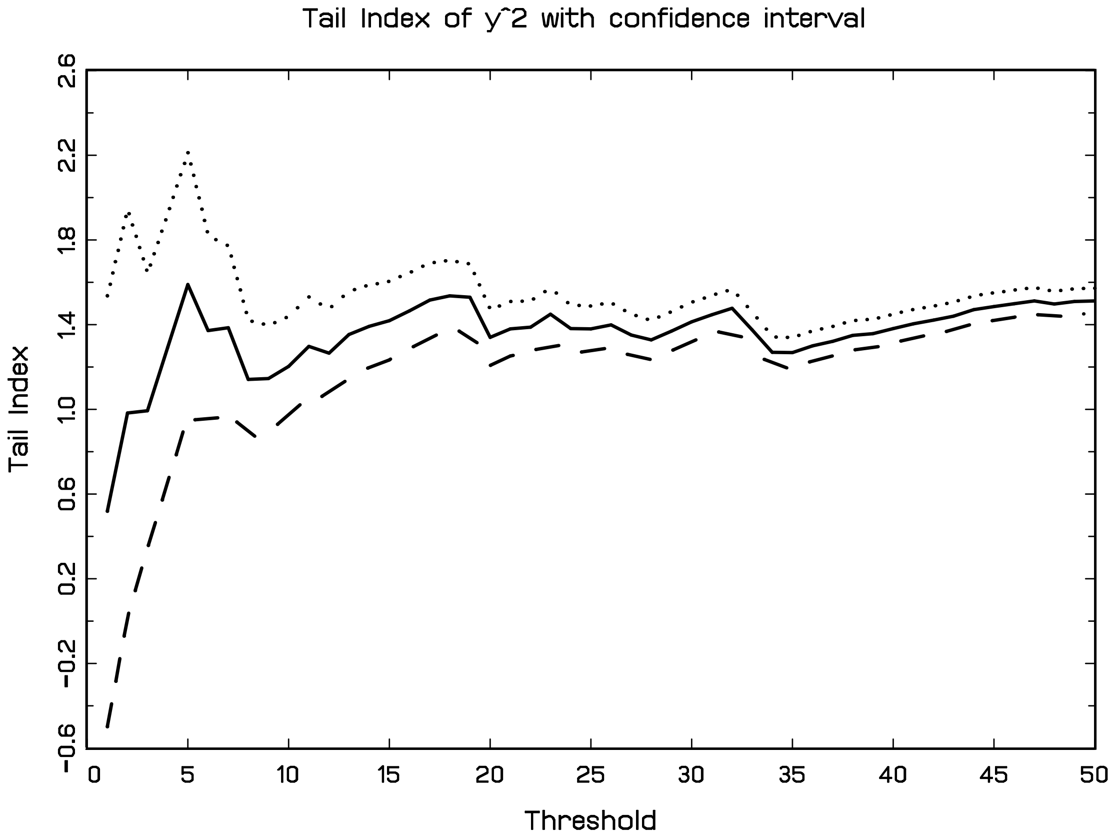

Figure 3. Shows the value of $\widehat{\kappa}$ against threshold size M for the series y_t^2. High frequency stock return series.

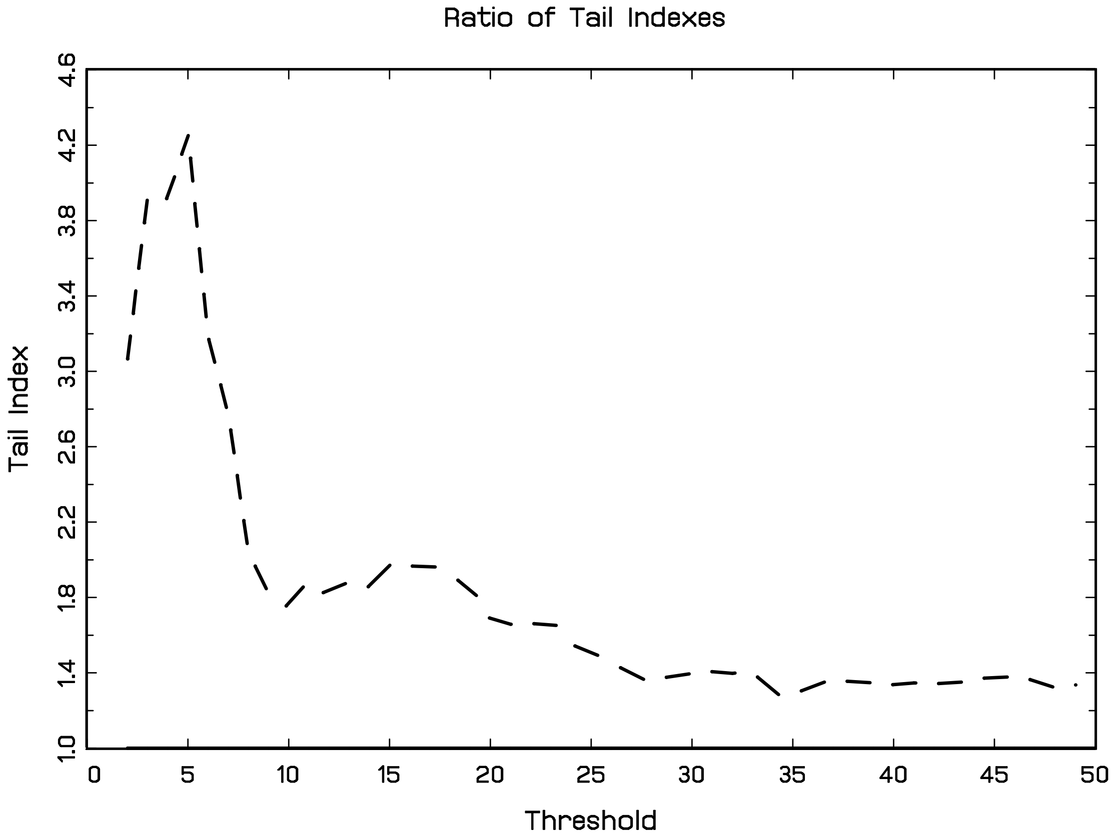

Figure 4. Shows the ratio of $\widehat{\kappa}$ for the series $|\Delta y_t^2|$ to $\widehat{\kappa}$ for the series y_t^2 against threshold size M. High frequency stock return series.

In this case, the differencing operation reduces the thickness of the tails considerably. The squared return series shows some evidence of non existence of first moment, but the differenced series has much lighter tails.

7 Conclusions

We established the consistency and rate of convergence of our estimates. Unfortunately, the pointwise distribution theory for $\widehat{m}$ appears to be very difficult and we have not anything to offer on this. Likewise our theory for $\widehat{\beta}$ falls short of root-T consistency and asymptotic normality. However, simulation evidence suggests that our estimators work well and that $\widehat{\beta}$ converges at rate root-T and is asymptotically normal. This will be presented in a future version of this paper.

8 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 Proposition 3.1 in Kim (2003), 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. $\dim(\mathcal{M}_Y) = \infty$. ■

Proof Theorem 4.2 Step I (a matrix form of $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$): Let $\widehat{g}_\lambda(\cdot,\cdot)$ be the kernel of the self-adjoint operator $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda : L^2(\mathcal{Y}) \to L^2(\mathcal{Y})$, i.e.,

$$\widehat{g}_\lambda(x,u) = \int_{\mathcal{Y}} \widehat{f}_\lambda(x,w)\widehat{f}_\lambda(u,w)dw,$$

where $\widehat{f}_\lambda(y,w) = T^{*-1}\sum_{t=\tau+1}^{T} K_{g_1}(y_{t-1}-y)\left[\sum_{k=2}^{\tau}\lambda_k K_{g_2}(y_{t-k}-w)\right]$. Through a straightforward calculation, $\widehat{g}_\lambda(\cdot,\cdot)$ is written, in a matrix form, as

$$\widehat{g}_\lambda(x,u) = T^{*-2}K_T^Y(x)^\top M_\lambda K_T^Y(u).$$

Plugging in $\widehat{g}_\lambda(\cdot,\cdot)$ into the operator $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$ yields

$$\begin{aligned}(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda m)(x) &= \int \widehat{g}_\lambda(x,u)m(u)du = T^{*-2}K_T^Y(x)^\top M_\lambda \int_{\mathcal{Y}} K_T^Y(u)m(u)du \\ &= T^{*-2}K_T^Y(x)^\top M_\lambda < K_T^Y(\cdot),\ m(\cdot) >_{L^2(\mathcal{Y})}. \end{aligned} \tag{14}$$

Step II (the spectral representation of $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$): Let $\{(\kappa_s,e_s)\}_{s=1}^{T'}$ denote all the nonzero eigenvalues and the corresponding eigenvectors of $Q_{Y,\lambda} = T^{*-2}M_\lambda^{1/2}M_Y M_\lambda^{1/2}$, where $T' = \text{rank}(Q_{Y,\lambda}) \le T^*$. Define

$$v_s(x) = K_T^Y(x)^\top M_\lambda^{1/2}e_s. \tag{15}$$

We claim that the spectral representation of the compact self-adjoint operator $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$ is given by

$$\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda = \sum_{s=1}^{T'}\kappa_s P_{v_s},$$

where P_{v_s} denotes the orthogonal projection on the subspace generated by the function v_s. To prove the claim, it suffices to show that all the nonzero eigenvalues and the corresponding eigenfunctions of $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$ are given by $\{(\kappa_s,v_s)\}_{s=1}^{T'}$. From the definition of (κ_s,e_s), it follows that

$$\begin{aligned}(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda v_s)(x) &= T^{*-2}K_T^Y(x)^\top M_\lambda < K_T^Y,\ v_s >_{L^2(\mathcal{Y})} \\ &= T^{*-2}K_T^Y(x)^\top M_\lambda < K_T^Y,\ K_T^{Y\prime} >_{L^2(\mathcal{Y})} M_\lambda^{1/2}e_s \\ &= K_T^Y(x)^\top M_\lambda^{1/2}(T^{*-2}M_\lambda^{1/2}M_Y M_\lambda^{1/2})e_s \\ &= K_T^Y(x)^\top M_\lambda^{1/2}(\kappa_s e_s) = \kappa_s v_s,\end{aligned}$$

implying that $\{(\kappa_s, v_s)\}_{s=1}^{T'}$ is a subset of the eigensystem of $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$ corresponding to the nonzero eigenvalues. From $\dim(\mathcal{R}(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)) = \min[\dim(\text{lin}\{K_T^Y(\cdot)\}), \dim(\text{lin}\{K_T^\lambda(\cdot)\})] = \text{rank}(Q_{Y,\lambda})$, the number of nonzero eigenvalues of $\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda$ is equal to T', completing the proof for the claim.

Step III (the spectral representation of $r(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)$): From the theorem on spectral calculus-see, Taylor and Lay (1980, p.368, for example), we obtain the spectral representation of $r(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)$

$$r(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)(\cdot) = \sum_{s=1}^{T'} r(\kappa_s) P_{v_s}(\cdot) = \sum_{s=1}^{T'} r(\kappa_s) v_s (< v_s, v_s >_{L^2(\mathcal{Y})})^{-1} < v_s, \cdot >_{L^2(\mathcal{Y})} .$$

By plugging in (15) into the above equation,

$$[r(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)m](x) = \sum_{s=1}^{T'} r(\kappa_s) K_T^Y(x)^\top M_\lambda^{1/2} e_s (e_s^\top M_\lambda^{1/2} M_Y M_\lambda^{1/2} e_s)^{-1} e_s^\top M_\lambda^{1/2} < K_T^Y,\ m >_{L^2(\mathcal{Y})} ,$$

which, by definition of e_s, reduces to

$$\begin{aligned} & T^{*-2} K_T^Y(x)^\top M_\lambda^{1/2} \left[\sum_{s=1}^{T'} r(\kappa_s) \kappa_s^{-1} P_{e_s} \right] M_\lambda^{1/2} < K_T^Y,\ m >_{L^2(\mathcal{Y})} \\ & = T^{*-2} K_T^Y(x)^\top M_\lambda^{1/2} r(Q_{Y,\lambda}) Q_{Y,\lambda}^{-1} M_\lambda^{1/2} < K_T^Y,\ m >_{L^2(\mathcal{Y})} . \end{aligned}$$

Step IV (Closed form of $\widehat{m}_\alpha = (\alpha I + \widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^{-1}\widehat{\mathcal{T}}_\lambda^*\widehat{h}_\lambda$): From $\widehat{h}_\lambda(w) = T^{*-1}K_T^\lambda(w)\mathbf{y}$,

$$\begin{aligned} (\widehat{\mathcal{T}}_\lambda^*\widehat{h}_\lambda)(x) &= \int_\mathcal{Y} \widehat{h}_\lambda(w) \widehat{f}_\lambda(x,w) dw = T^{*-2} K_T^Y(x)^\top \left[\int_\mathcal{Y} K_T^\lambda(w) K_T^\lambda(w)^\top dw \right] \mathbf{y} \\ &= T^{*-2} K_T^Y(x)^\top M_\lambda \mathbf{y}. \end{aligned}$$

Therefore,

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

■

Lemma A.1 *Assume that C.1 through C.5(a) hold. Then, it holds that*

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

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

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

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

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

$$m_{c(g_1)}(y) \equiv (K_{g_1} * m)(y) = \int K_{g_1}(y-s)m(s)ds, \tag{16}$$

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 MISE of $\widehat{\mathcal{T}}_k m$, given by

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

From

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

a standard calculation of the variance term for kernel estimates (under the mixing condition in C.1) yields

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

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 C.3), it holds

$$\begin{aligned} \int_{\mathcal{Y}}\mathrm{Var}\left[s_T(w)\right]dw &\simeq \frac{1}{Tg_2}||K||_2^2\int_{\mathcal{Y}} m^2(y)f_0(y)dy \\ &\leq C\frac{1}{Tg_2}||K||_2^2||m||^2_{L^2(\mathcal{Y})}. \end{aligned} \tag{17}$$

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

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

Under C.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_0-th partial derivatives of $f_k(\cdot,\cdot)$ that are continuous and square integrable (i.e., condition C.4 holds), we have, by application of the standard Taylor expansion, that

$$\int_{\mathcal{Y}} \mathrm{B}_T^2(w)dw \leq C\{g_1^{2q} g_2^{2(p_0-q)} \sum_{q=0}^{p_0} ||\frac{\partial^{p_0} f_k(y,w)}{\partial y^q \partial w^{p_0-q}}||^2_{L^2(\mathcal{Y}\times\mathcal{Y})}\} ||m||^2_{L^2(\mathcal{Y})}. \quad (18)$$

Combining (17) and (18), 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

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

Noting that the variance term in (17) dose 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 C.4), we finally get

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

(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 (16). 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

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

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

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

$$\begin{aligned}\int_{\mathcal{Y}}\mathrm{Var}\left[s_{1,T}(w)\right]dw &= \frac{1}{Tg_2}||K||_2^2\left[\int_{\mathcal{Y}}\mathrm{E}(\eta_t^2|y_{t-k} = w)f_0(w)dw\right](1 + o(1)) \\ &= \frac{1}{Tg_2}||K||_2^2\sigma_\eta^2(1 + o(1)) = O(\frac{1}{Tg_2}), \end{aligned} \qquad (19)$$

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

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

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

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

From boundedness of $f_0(\cdot)$,

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

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

$$\begin{aligned}\mathrm{E}[\mathrm{B}_T(w)] &= \mathrm{E}[K_{g_2}(y_{t-k} - w)r(y_{t-1})] \\ &= \int_{\mathcal{Y}}[m(y) - m_{c(g_1)}(y)]f_k(y,w)dy(1 + o(1)). \end{aligned} \qquad (20)$$

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

$$\mathrm{E}[\mathrm{B}_T(w)] = \{\int_{\mathcal{Y}} \left[f_k(u,w) - f_k^{c(g_1)}(u,w) \right] m(u) du\}(1+o(1)), \tag{21}$$

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 (20) and (21), together with square-integrability of $f_k(\cdot,\cdot)$ and $m(\cdot)$, that

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

which, by the standard method of Taylor expansion (under C.4), gives

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

Since the variance term in (19) dose 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 C.4), it holds that

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

■

Proof of Proposition 5.1 (i) By Triangle inequality,

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

where the second inequality follows from $\sup_{k\ge 2} \; ||\mathcal{T}_k|| \; \le \sup_{k\ge 2} \; ||f_k|| \; \le C$ (condition C.4) and $\sum_{k=\tau+1}^{\infty} \lambda_k = o(1/\sqrt{T})$ (condition C.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). ■

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 following conditions:

B.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 $\alpha \to 0^+$.

B.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 $\alpha_0 > 0$.

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 B.1 and $\widehat{\mathcal{T}}_\lambda$ has a finite rank, the results are immediate from Lemma 3.2 (Kim 2004, 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 B.2, the results follow from Lemma 3.3 (Kim 2004, p.42). ■

For $\mathcal{M}_T(\mathcal{Y}) = \{m \in \mathcal{M}(\mathcal{Y}) : m_T(x) = T^{*-2} K_T^Y(x)^\top \beta_T$, for $\beta_T \in \mathbb{R}^{T^*}\}$, 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 $f_k(\cdot,\cdot)$ which 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*

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

where $0 < \omega \le 1$.

Proof By applying 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 semi-definite, we can show that, for $m_T(\cdot) \in \mathcal{M}_T(\mathcal{Y})$,

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

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

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

where the last inequality comes from the compact support assumption, uniform consistency of the kernel density estimate $\widehat{f}_k(\cdot,\cdot)$, and condition C.6. In a similar way, we can show, by the CS inequality and boundedness of $f_k(\cdot,\cdot)$, that, for any $m \in L^2(\mathcal{Y})$,

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

Hence, it follows

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

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 following error decomposition

$$\widehat{m}_\alpha - m_0 = U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*(\widehat{h}_\lambda - \widehat{\mathcal{T}}_\lambda m_0) + [U_\alpha(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda - I] m_0. \tag{22}$$

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 $\mathcal{T}_\lambda : L^2(\mathcal{X}) \to L^2(\mathcal{W})$ which is bounded and one-to-one. Also, under the bandwidth conditions in C.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 (22).

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

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

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:

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

Also, by 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),

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

Let $\omega > 0$ be any fixed real number. In the proof of Theorem 4.2 (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

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

i.e., $(\widehat{\mathcal{T}}_\lambda^*\widehat{\mathcal{T}}_\lambda)^\omega : L^2(\mathcal{Y}) \to C(\mathcal{Y})$ is well defined. Let $\mathcal{M}_T(\mathcal{Y}) = \{m \in L^2(\mathcal{Y}) : m_T(x) = T^{-2} K_T^Y(x)^\top \beta_T$, for $\beta_T \in \mathbb{R}^T\}$. 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

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

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

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

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

and

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

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 (1997)[2];

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

[2] More rigorously, we may follow the same line of the proof of Proposition 3.1(ii) to apply the uniform convergence results in Masry (1997).

■

Proof of Theorem 5.4. We have

$$\widehat{\ell}(\beta) - \ell(\beta) = \frac{1}{T} \sum_{t=1}^{T} \ln \frac{\sigma_t^2(\beta, \widehat{m}_\alpha)}{\sigma_t^2(\beta, m_0)} + y_t^2 \frac{[\sigma_t^2(\beta, \widehat{m}_\alpha) - \sigma_t^2(\beta, m_0)]}{\sigma_t^2(\beta, m_0)\sigma_t^2(\beta, \widehat{m}_\alpha)}.$$

Given the properties of $\sigma_t^2(\beta, m_0)$, $\widehat{\ell}(\beta) - \ell(\beta)$ just depends on $\max_t |\sigma_t^2(\beta, \widehat{m}_\alpha) - \sigma_t^2(\beta, m_0)|$, which in turn can be bounded by $\tau(T) \times \sup_{y \in \mathcal{Y}} \mid \widehat{m}_\alpha(y) - m_0(y) \mid = o_p(T^{-1/4})$ as required. Specifically,

$$\sup_{\beta \in B} \left| \widehat{\ell}(\beta) - \ell(\beta) \right| = o_p(T^{-1/4}),$$

which implies that $\widehat{\beta} - \widetilde{\beta} = o_p(T^{-1/4})$. Therefore, $\widehat{\beta}$ is $T^{-1/4}$ consistent as required. ■

References

[1] Ai, C. and X. Chen (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica* 71(6), 1795-1844.

[2] Birman, M. and M. Solomjak (1980): *Quantitative Analysis in Sobolev Imbedding Theorems and Applications to Spectral Theory.* American Math. Soc. Tran., series 2, Vol 114.

[3] Blundell, R., X. Chen, and D. Kristensen (2004): "Semiparametric Estimation of Shape Invariant Engle Curves under Endogeneity" CEMMAP working paper WP15/03

[4] Blundell, R. and J. L. Powell (2001a): "Endogeneity in nonparametric and semiparametric regression models," forthcoming in *Advances in Econometrics, Proceedings of the World Meetings, 2000*, ed. by L. Hansen, North Holland.

[5] Blundell, R. and J. L. Powell (2001b): "Endogeneity in semiparametric binary response models," Center for Microdata Methods and Practice, Working Paper, 05/01.

[6] Chen, X. and X. Shen (1998). Sieve extremum estimates for weakly dependent data. Econometrica 66, 289-314.

[7] Darolles, S., J-P. Florens, and E. Renault (2001): "Nonparametric instrumental regression, Working Paper, GREMAQ, University of Toulouse.

[8] Das, M. (1999): "Instrumental variable estimation of models with discrete endogenous regressors, " presented at 2000 World Congress of the Econometric Society.

[9] Engl, H. (1987): "On the choice of the regularization parameter for iterated Tikhonov regularization of ill-posed problems," *Journal of Approximation Theory 49, 55-63.*

[10] Engl, H. and H. Gfrerer (1988): "A posteriori parameter choice for general regularization methods for solving linear ill-posed problems," *Applied Numerical Mathematics 4, 395-417.*

[11] Engle, R.F., and T. Bollerslev (1986): "Modelling the persistence of Conditional Variances," *Econometric Reviews* 5, 1-50.

[12] Engle, H.W., M. Hanke, and A. Neubauer (2000): *Regularization of Inverse Problems.* Dordrecht: Kluwer Academic Press.

[13] Fan, J. (1991): "Global behavior of deconcovolution kernel estimates," *Statistica Sinica* 1, 541-551.

[14] Groetsch, C.W. (1993): *Inverse Problems in the Mathematical Sciences.* Braunschweig: Vieweg.

[15] Groetsch, C.W. (1983): "On the Asymptotic Order of Accuracy of Tikhonov Regularization," *Journal of Optimization Theory and Applications 41, 293-298.*

[16] Groetsch, C.W. (1977): *Generalized Inverses of Linear Operators: Representation and Approximation.* New York: Dekker.

[17] Hall P. and J. L. Horowitz (2003): "Nonparametric methods for inference in the presence of instrumental variables," Working Paper, Center for Microdata Methods and Practice.

[18] Hansen L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica* 50, 1029-1054.

[19] Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies* 61, 247-64.

[20] Hutson, V. and J. S. Pym (1980): *Applications of Functional Analysis and Operator Theory.* London: Academic Press.

[21] Imbens, G. W. and W. K. Newey (2001): "Identification and estimation of triangular simultaneous equations models without additivity, Working Paper.

[22] Ivanov, V.K., V.V. Vasin, and V.P. Tanana (1978): *Theory of Linear Ill-Posed Problems.* Moscow: Nauka.

[23] Kim, W. (2003): "Identification and Estimation of Nonparametric Structural models by Instrumental Variables Method," Manuscript, Humboldt University.

[24] King, J.T. and D. Chillingworth (1979): "Approximation of generalized inverses by iterated regularization," *Numerical Functional Analyses and Optimizations 1, 499*-513.

[25] Kirsch, A. (1996): *An Introduction to the Mathematical Theory of Inverse Problems.* New York: Springer Verlag.

[26] Kress, R. (1989): *Linear Integral Equations.* Berlin: Springer Verlag.

[27] Lee, S., and Hansen, B. (1994): "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29-52.

[28] Linton, O. B. and E. Mammen (2003): "Estimating semiparametric ARCH(∞) models by kernel smoothing methods," Forthcoming in *Econometrica.*

[29] Mammen, E., O. B. Linton, and J. Nielsen (1999): "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions," *The Annals of Statistics,* 27(5), 1443-1490.

[30] Meitz, M. and 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.

[31] Nashed, M.Z. (1976): *Generalized Inverses and Applications.* New York: Academic Press.

[32] Nashed, M.Z. and G. Wahba (1974): "Convergence rates of approximate least squares solutions of linear integral and operator equations of the first kind," *Mathematics of Computation* 28(125), 69-80.

[33] Nelson, D. (1990). Stationarity and persistence in the GARCH(1,1) models. Econometric Theory 6, 318-334.

[34] Newey W. K. and J. L. Powell (1988): "Nonparametric instrumental variables estimation," MIT Working Paper.

[35] Newey W. K. and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica* 71(5), 1565-1578.

[36] Newey W. K., J. L. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica* 67, 565-603.

[37] Nychka, D. and D. Cox (1989): "Convergence rates for regularized solutions of integral equations from discrete noisy data," *The Annals of Statistics,* 17(5), 556-572.

[38] O'Sullivan, F. (1986): "Ill-posed inverse problems (with Discussion)," *Statistical Science,* 4, 503-527.

[39] Phillips, D.L. (1962): "A technique for the numerical solution of certain integral equations of the first kind," *Journal of the Association for Computing Machinery* 9, 84-97.

[40] Plato, R. and G. Vainikko (1990): "On the regularization of projection methods for solving ill-posed problems," *Numerische Mathematik* 57, 63-79.

[41] Rahbek, A. and S.T. Jensen (2004): "Asymptotic normality for Nonstationary, Explosive GARCH ," Econometric Theory 20, 1203-1226.

[42] Roehrig, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 55, 875-891.

[43] Sargan, J. D. (1958): "The estimation of economic relationships using instrumental variables," *Econometrica* 26, 393-415.

[44] Schock, E. (1985): "Approximate solution of ill-posed equations: Arbitrarily slow convergence vs superconvergence, in *Constructive Methods for the Practical Treatment of Integral Equations* (edited by G. Hämmerlin and K.H. Hoffmann), pp. 234-243, Basel: Birkhäuser.

[45] Showalter D. (1967): "Representation and computation of the pseudoinverse," *Proc. Amer. Math. Soc. 18*, 584-586.

[46] Silverman, B.W. (1978): "Weak and strong uniform consistency of the kernel estimate of a density and its derivatives," *Annals of Statistics* Vol.6, No.1, 177-184.

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

[48] Theil, H. (1953): Repeated Least Squares Applied to Complete Equation Systems, The Hague; Central Planning Bureau.

[49] Tikhonov, A.N. and V. Arsenin (1977): *Solutions of Ill-Posed Problems.* New York: Wiley.

[50] Tikhonov, A.N (1963): "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics Doklady 4*, 1035-1038.

[51] Vainikko, G.M. and A.Y. Veretennikov (1986): *Iteration Procedures in Ill-Posed problems.* 1st Ed. Moscow: Nauka.

[52] van Rooij, A.C.M., and F.H. Ruymgaart (1999): "On inverse estimation," in *Asymptotics, Nonparametrics, and Time Series*, ed. by S. Ghosh. New York: Marcel Dekker, 579-613.