

Department of Economics

Working Paper Series

Local GMM Estimation of Time Series Models with Conditional Moment Restrictions

08-010 | Nikolay Gospodinov
Concordia University
Taisuke Otsu
Yale University



UNIVERSITÉ

Concordia

UNIVERSITY

LOCAL GMM ESTIMATION OF TIME SERIES MODELS WITH CONDITIONAL MOMENT RESTRICTIONS*

Nikolay Gospodinov[†]
Concordia University and CIREQ

Taisuke Otsu[‡]
Yale University

First Version: September 2008

Revised: December 2009

Abstract

This paper investigates statistical properties of the local generalized method of moments (LGMM) estimator for some time series models defined by conditional moment restrictions. First, we consider Markov processes with possible conditional heteroskedasticity of unknown forms and establish the consistency, asymptotic normality, and semi-parametric efficiency of the LGMM estimator. Second, we undertake a higher-order asymptotic expansion and demonstrate that the LGMM estimator possesses some appealing bias reduction properties for positively autocorrelated processes. Our analysis of the asymptotic expansion of the LGMM estimator reveals an interesting contrast with the OLS estimator that helps to shed light on the nature of the bias correction performed by the LGMM estimator. The practical importance of these findings is evaluated in terms of a bond and option pricing exercise based on a diffusion model for spot interest rate.

KEYWORDS: Conditional moment restriction; Local GMM; Higher-order expansion; Conditional heteroskedasticity.

JEL CLASSIFICATION: C13; C22; G12.

*We would like to thank the editors, a co-editor, three anonymous referees, conference participants at the 2007 CIREQ conference on Generalized Method of Moments, the 2007 meeting of the Canadian Econometrics Study Group and the 2009 Joint Statistical Meetings for useful comments and suggestions. Financial support from FQRSC and SSHRC (Gospodinov) and National Science Foundation under SES-0720961 (Otsu) is gratefully acknowledged.

[†]Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada, email: nikolay.gospodinov@concordia.ca.

[‡]Cowles Foundation and Department of Economics, Yale University, P.O. Box 208281, New Haven, CT, 06520-8281 USA, email: taisuke.otsu@yale.edu.

1 Introduction

While modern economic theory typically implies a set of conditional moment restrictions, estimation of the model parameters is often performed using a framework based on unconditional moment restrictions such as the standard generalized method of moments (GMM). Despite its computational attractiveness, this GMM-based approach may result in efficiency losses and inconsistency that arises from possible nonidentifiability of the parameters of interest by the unconditional moment restrictions even when the conditional moment restrictions identify the parameters (Dominguez and Lobato, 2004). Time series regression models, for example, are usually defined in terms of a sequence of disturbances whose expectation conditional on a few recent lags of the data is assumed to be zero. This specification allows for conditional heteroskedasticity which is a stylized feature of many economic and financial time series data such as interest rates, exchange rates, asset returns, etc.

This paper studies estimation of conditional moment restriction models in a time series context. In particular, we focus on the local (conditional or smoothed) GMM (LGMM, hereafter) estimator that belongs to the class of localized versions of the generalized empirical likelihood estimator (Newey and Smith, 2004) introduced and developed by Kitamura, Tripathi and Ahn (2004), Smith (2007), and Antoine, Bonnal and Renault (2007) (see, e.g., Newey, 1990, 1993; Carrasco and Florens, 2000; and Donald, Imbens and Newey, 2003, for some alternative methods to estimate conditional moment restriction models). While all these papers investigate the properties of the local estimators for *iid* data, we study the first- and higher-order asymptotic properties of the LGMM estimator for strictly stationary and geometrically ergodic Markov processes.

First, we study the first-order asymptotic behavior of the LGMM estimator and show that the LGMM estimator is consistent and asymptotically normal, and attains the semi-parametric efficiency bound for conditional moment restrictions with a martingale difference structure derived in Carrasco and Florens (2004). We should note that semi-parametrically efficient moment-based estimators have been proposed by Carrasco, Chernov, Florens and Ghysels (2007), Kuersteiner (2001, 2002) and West, Wong and Anatolyev (2009) for more general time series processes. Although our Markov setup is more restrictive than the setups of these papers, our emphasis is on the bias property of the LGMM estimator in small samples and the Markov setup helps us to simplify the higher-order analysis of the estimator. Indeed, our simulation results for AR models with conditional heteroskedasticity suggest that the LGMM estimator is characterized by a smaller bias and mean squared error than these alternative estimators and, in some cases, even the infeasible GLS estimator.

Next, in order to explain the bias reduction property of the LGMM estimator, we consider

explicitly the AR(1) model with *iid* errors and undertake a higher-order expansion for the LGMM estimator. In particular, while the bias of the OLS estimator is given by a single negative term of order $O(T^{-1})$, the higher-order expansion of the LGMM estimator reveals the presence of two $O(T^{-1})$ bias terms. We compare these terms with those of the OLS estimator and provide an approximation formula for one of them. Our numerical results show that for positively autocorrelated processes the two leading bias terms in the LGMM estimator tend to have opposite signs and similar magnitudes that offset each other. Also, this bias reduction is achieved without inflating the variance of the LGMM estimator and then the LGMM can potentially dominate the GLS and other efficient estimators in terms of mean squared errors. The economic significance of the bias reduction property of the LGMM estimator is assessed for derivative interest rate products whose prices exhibit strong sensitivity to the parameter that governs the persistence of the process. Interestingly, the LGMM estimator seems to outperform in terms of mean square error the infeasible optimal GMM as the LGMM tends to remove a substantial portion of the bias that arises from the highly persistent dynamics of the interest rate data.

Our paper complements the existing econometric and statistical literature which is concerned with extending the local estimators for moment condition models to dependent data. For example, Kitamura (1997) considers unconditional moment restrictions with weakly dependent data and proposes a non-local empirical likelihood estimator by using blocked (or local average) moment restrictions. Our paper has three important differences with Kitamura (1997): (i) while Kitamura (1997) focuses on unconditional moments, we consider conditional moment restrictions that imply an infinite number of unconditional moments, (ii) while Kitamura (1997) requires local averaging for moments to account for the long-run variance of the unconditional sample moments, we employ local averaging to nonparametrically approximate the conditional moment restrictions, and (iii) while Kitamura (1997) allows for general weakly dependent data, our analysis is restricted to Markov processes.

In a different but related strand of literature, Chen, Härdle and Li (2003) propose a goodness-of-fit test statistic based on the local empirical likelihood function to check the validity of conditional moment restrictions. Furthermore, Su and White (2003) adopt the local empirical likelihood framework to test conditional independence restrictions which imply sequences of conditional moment restrictions. Both papers focus on hypothesis testing with possibly dependent data and derive first-order asymptotic properties of the local empirical likelihood-based test statistics. In contrast, we focus on point estimation of parameters in conditional moment restriction models and derive some first- and higher-order properties of the LGMM estimator. Finally, Gagliardini, Gourieroux and Renault (2007) extend the local estimation approach to estimate conditional moments of interest (derivative prices in their example) with dependent data when the conditional moment restrictions

(say, $E[u_1(y_{t+1}, \theta_0)|x_t] = 0$) and local moment restrictions (say, $E[u_2(y_{t+1}, \theta_0)|x_t = \bar{x}] = 0$ for some given \bar{x}) coexist. While Gagliardini, Gourieroux and Renault (2007) use the LGMM approach to evaluate the local moments $E[u_2(y_{t+1}, \theta_0)|x_t = \bar{x}] = 0$, they adopt the optimal instrumental variable approach of Chamberlain (1987) and Newey (1990, 1993) to accommodate the conditional moments $E[u_1(y_{t+1}, \theta_0)|x_t] = 0$. Our setup differs from Gagliardini, Gourieroux and Renault (2007) since we do not consider the local moments $E[u_2(y_{t+1}, \theta_0)|x_t = \bar{x}] = 0$ but utilize the LGMM framework to estimate the parameters θ_0 in the conditional moments $E[u_1(y_{t+1}, \theta_0)|x_t] = 0$. Also, the general theoretical arguments in Gagliardini, Gourieroux and Renault (2007) are mostly concerned with deriving the efficiency bounds in their setup. On the other hand, we are interested in establishing the first- and higher-order asymptotic properties of the LGMM estimator.

The rest of the paper is organized as follows. The next section describes the model and estimation procedure. Section 3 develops the first-order asymptotic theory for the LGMM estimator and reports some numerical properties of the estimator in finite samples. Section 4 derives the higher-order asymptotic expansion of the LGMM estimator in an AR(1) model with *iid* errors and analyzes the bias properties of the estimator. The bias terms of the LGMM and OLS estimators are evaluated by simulation. Section 5 assesses the economic significance of the bias reduction and the efficiency of the LGMM in the context of a bond and derivative pricing exercise. Section 6 concludes. All proofs are contained in the appendix.

2 Model and Estimation Procedure

Suppose that the univariate process of interest $\{r_t\}_{t=-\infty}^{\infty}$ on \mathbb{R} is strictly stationary and geometrically ergodic¹ and denote the conditional moment restrictions imposed by some economic theory as

$$E[u(r_{t+1}, r_t, \dots, r_{t-p+1}, \theta_0) | r_t, \dots, r_{t-p+1}] = 0, \quad (1)$$

for each $t \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, where $u : \mathbb{R}^{p+1} \times \Theta \rightarrow \mathbb{R}^l$ is a known function up to a vector of unknown parameters $\theta_0 \in \Theta \subseteq \mathbb{R}^k$. Technical conditions that restrict the dependence structure of the process and ensure the validity of the estimation procedure are discussed in the next section.

One popular example of this framework is the AR(1) model with martingale difference errors

$$r_{t+1} = \gamma_0 + \gamma_1 r_t + u_{t+1}, \quad E[u_{t+1} | r_t] = 0, \quad (2)$$

¹Let $\{X_t\}_{t=-\infty}^{\infty}$ be a time-homogeneous Markov process with state space $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, where $\mathcal{B}(\mathbb{R}^p)$ is a Borel σ -algebra of \mathbb{R}^p , and m -step transition probability $P^m(x, A) = P(X_m \in A | X_0 = x)$ for $x \in A$ and $A \in \mathcal{B}(\mathbb{R}^p)$. The process $\{X_t\}_{t=-\infty}^{\infty}$ is geometrically ergodic if there exists a probability measure π on $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, a constant $0 < \rho < 1$, and a π -integrable non-negative measurable function $C(\cdot)$ such that $\|P^m(x, \cdot) - \pi(\cdot)\|_{\tau} \leq \rho^m C(x)$ for all $m \in \mathbb{N}$ and $x \in \mathbb{R}^p$, where $\|\cdot\|_{\tau}$ denotes the total variation norm (Carrasco and Chen, 2002; Meyn and Tweedie, 1993).

for each $t \in \mathbb{Z}$. In this case, the moment function is specified as $u(r_{t+1}, r_t, \theta_0) = r_{t+1} - \gamma_0 - \gamma_1 r_t$ with $\theta_0 = (\gamma_0, \gamma_1)$. For $\gamma_1 > 0$, this model can be regarded as a discrete-time specification of the diffusion process

$$dr_t = \kappa(\mu - r_t) dt + \sigma(r_t) dW_t, \quad (3)$$

for each $t \in [0, \infty)$, where $\{W_t\}_{t \geq 0}$ is the standard Brownian motion, $\kappa(\mu - r_t)$ is the drift component, and $\sigma(\cdot)$ is the diffusion function. In this linear parametrization of the drift function, μ is the long-run unconditional mean of the process and κ is the speed of mean reversion. The diffusion process (3) is often used to model the dynamics of spot interest rate whose parameters are inferred from the estimates of the discrete-time model (2) by setting $\gamma_0 = \mu(1 - e^{-\kappa})$ and $\gamma_1 = (1 - e^{-\kappa})$. In Section 5, we employ this model for a bond and derivative pricing exercise.

In order to simplify the notation, let $x_t = (r_t, \dots, r_{t-p+1})'$ and $y_{t+1} = (r_{t+1}, x_t')'$.² The conditional moment restriction model (1) is typically estimated by the GMM estimator based on the unconditional moment restrictions $E[g(y_{t+1}, \theta_0)] = E[\mathcal{A}(x_t, \theta_0) u(y_{t+1}, \theta_0)] = 0$ with a matrix of instruments $\mathcal{A}(x_t, \theta_0)$, which is implied from the original model (1). For example, the (continuously updated) GMM estimator (Hansen, Heaton and Yaron, 1996) is defined as

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} \left(\frac{1}{T-p} \sum_{t=p}^{T-1} g(y_{t+1}, \theta) \right)' W_T(\theta)^{-1} \left(\frac{1}{T-p} \sum_{t=p}^{T-1} g(y_{t+1}, \theta) \right), \quad (4)$$

where $W_T(\theta) = \frac{1}{T-p} \sum_{t=p}^{T-1} g(y_{t+1}, \theta) g(y_{t+1}, \theta)'$ is an optimal weight matrix to estimate the parameters from the unconditional moment restrictions $E[g(y_{t+1}, \theta_0)] = 0$.

In this paper, we pursue an alternative approach and use a localized version of the GMM estimator that operates directly on the conditional moment restriction (1). Let $w_{tj} = \mathbb{K}\left(\frac{x_j - x_t}{h}\right) / \sum_{j=p}^{T-1} \mathbb{K}\left(\frac{x_j - x_t}{h}\right)$ denote kernel weights, where $\mathbb{K} : \mathbb{R}^p \rightarrow \mathbb{R}$ is a kernel function and h is a bandwidth parameter. Let $\mathbb{I}_{tT} = I\{|x_t| \leq c_T\}$ be a trimming term to deal with some bias problems of kernel estimators, where $I\{\cdot\}$ is the indicator function and c_T is a sequence satisfying $c_T \propto T^\xi$ for some $\xi > 0$.³ The kernel estimator for the conditional moment $E[u(y_{t+1}, \theta) | x_t]$ is defined as $u_T(x_t, \theta) = \sum_{j=p}^{T-1} w_{tj} u(y_{j+1}, \theta)$ and the LGMM estimator minimizes its quadratic form, i.e.,

$$\hat{\theta}_{LGMM} = \arg \min_{\theta \in \Theta} \sum_{t=p}^{T-1} \mathbb{I}_{tT} u_T(x_t, \theta)' V_T(x_t, \theta)^{-1} u_T(x_t, \theta), \quad (5)$$

where $V_T(x_t, \theta) = \sum_{j=p}^{T-1} w_{tj} u(y_{j+1}, \theta) u(y_{j+1}, \theta)'$ is an optimal weight matrix to estimate the parameters from the conditional moment restrictions.

²If we have exogenous regressors z_t , the vectors x_t and y_t can be defined as $x_t = (r_t, \dots, r_{t-p+1}, z_t)'$ and $y_{t+1} = (r_{t+1}, x_t', z_t)'$.

³Alternatively, similar to Kitamura, Tripathi and Ahn (2004), the trimming term may be defined as $I\left\{\frac{1}{T-p} \sum_{j=p}^{T-1} \mathbb{K}\left(\frac{x_j - x_t}{h}\right) \geq c'_T\right\}$ with some $c'_T \rightarrow 0$, which trims observations that have small kernel density estimates. Although we employ the trimming term \mathbb{I}_{tT} to simplify our technical argument, a similar but more lengthy argument for the alternative trimming term defined above will yield analogous results to ours.

Smith (2007) shows that the LGMM estimator (5) belongs to the class of local Cressie-Read minimum distance estimators which also includes the local (conditional or smoothed) empirical likelihood estimator proposed by Kitamura, Tripathi and Ahn (2004). See Smith (2007) and Antoine, Bonnal and Renault (2007) for a detailed discussion and interpretation of these local estimators.⁴ In this study, we adopt the LGMM estimator (5) instead of the local empirical likelihood estimator due to technical (for higher-order analysis) and computational reasons (see, Antoine, Bonnal and Renault, 2007). However, our preliminary numerical experiments with the local empirical likelihood reveal only negligible differences for those estimators.

3 First-Order Asymptotic Theory

3.1 Asymptotic Properties of LGMM Estimator

In this section, we establish the consistency, asymptotic normality, and semi-parametric efficiency of the LGMM estimator for stationary, possibly nonlinear, and conditionally heteroskedastic p -th order Markov processes.

ASSUMPTION A1. The process $\{r_t\}_{t=-\infty}^{\infty}$ is a strictly stationary, absolutely regular, p -th order Markov process in \mathbb{R} (i.e., $P(r_{t+1}|r_t, r_{t-1}, \dots) = P(r_{t+1}|r_t, \dots, r_{t-p+1})$ for each $t \in \mathbb{Z}$) with mixing coefficients of order $O(\rho^m)$ for $m \in \mathbb{N}$ and some $0 < \rho < 1$.

Assumption A1 requires that the process $\{r_t\}_{t=-\infty}^{\infty}$ is strictly stationary and absolutely regular (β -mixing) with exponentially decaying mixing coefficients. This assumption is used to invoke the central limit theorem for U -statistics with weakly dependent data (Fan and Li, 1999). Also, recall that if the process is absolutely regular, it is also strong (α -) mixing. For example, Masry and Tjøstheim (1995) establish the conditions for geometric α -mixing of the conditionally heteroskedastic first-order Markov process

$$r_{t+1} = \phi(r_t) + u_{t+1}, \quad u_{t+1} = \sigma(r_t) \varepsilon_{t+1}, \quad (6)$$

for each $t \in \mathbb{Z}$, where $\phi : \mathbb{R} \rightarrow \mathbb{R}$ denotes the conditional mean function $E[r_{t+1}|r_t]$, $\sigma^2 : \mathbb{R} \rightarrow (0, \infty)$ denotes the conditional variance function $Var(r_{t+1}|r_t)$, and ε_{t+1} is *iid* with $E[\varepsilon_{t+1}] = 0$ and $E[\varepsilon_{t+1}^2] = 1$ for $t \in \mathbb{Z}$. The model in (6) allows for various types of conditional heteroskedasticity but it should be stressed that the LGMM estimator does not require any knowledge of the explicit form of the skedastic function $\sigma^2(\cdot)$ to estimate the parameters in the conditional mean function $\phi(\cdot)$.

Furthermore, the process generated by (6) is geometrically ergodic if $\limsup_{|r| \rightarrow \infty} \frac{\phi^2(r) + \sigma^2(r)}{r^2} < 1$ (Maercker, 1995; Meyn and Tweedie, 1993). If $E[r_{t+1}|r_t] = \theta_0 r_t$ and $\sigma^2(r_t) = \omega_0 + \beta_0 r_t^2$, then

⁴Other papers that adopt a similar type of local weighting include Kuersteiner (2006), Lavergne and Patilea (2008), and Lewbel (2007).

this condition becomes $\theta_0^2 + \beta_0 < 1$; if $E[r_{t+1}|r_t] = \theta_0 r_t$ and $\sigma^2(r_t) = \sigma_0^2 r_t^{\gamma_0}$, then this condition becomes $|\theta_0| < 1$ for $\gamma_0 \in [0, 2)$ and $\theta_0^2 + \sigma_0^2 < 1$ for $\gamma_0 = 2$. For the case of $E[r_{t+1}|r_t] = \theta_0 r_t$ and $\sigma^2(r_t) = \omega_0 + \beta_0 r_t^2$, Borkovec (2001) and Borkovec and Klüppelberg (2001) prove the geometric ergodicity and strict stationarity of $\{r_t\}_{t=-\infty}^{\infty}$ under the weaker conditions of symmetry of ε_{t+1} and $E[\log|\theta_0 + \varepsilon_{t+1}\sqrt{\beta_0}|]$ which allows for $|\theta_0| \geq 1$ and $E[u_{t+1}^2] = \infty$. Moreover, Ling (2004) establishes the asymptotic normality of the quasi maximum likelihood estimator of θ_0 that holds even for $|\theta_0| \geq 1$ and $E[u_{t+1}^2] = \infty$. Carrasco and Chen (2002) provide some sufficient conditions for β -mixing of various GARCH and stochastic volatility models. Finally, Guégan and Diebolt (1994) discuss some challenges in verifying the probabilistic properties of higher-order Markov processes.

Let $|A| = \sqrt{\text{trace}(A'A)}$ be the Euclidean norm for a scalar, vector, or matrix A , and $\min \text{eig}(A)$ and $\max \text{eig}(A)$ be the minimum and maximum eigenvalues for a matrix A , respectively. Let $f(\cdot)$ be the density function of X_t , $V(x, \theta) = E[u(y_{t+1}, \theta) u(y_{t+1}, \theta)' | x_t = x]$, and $D(x, \theta) = E[\partial u(y_{t+1}, \theta) / \partial \theta' | x_t = x]$. The following definition is useful to summarize the boundedness and continuity conditions for the moment function u and its derivatives.

DEFINITION (\mathcal{D} -bounded function). A function $a : \mathbb{R}^{p+1} \times A \rightarrow \mathbb{R}^q$ is called \mathcal{D} -bounded on A with order s if

- (i) $a(y, \theta)$ is almost surely differentiable at each $\theta \in A$,
- (ii) $\sup_{\theta \in A} E[|a(y_{t+1}, \theta)|^s] < \infty$ and $\sup_{\theta \in A} E[|\partial a(y_{t+1}, \theta) / \partial \theta'|^s] < \infty$ for each $t = p, \dots, T-1$,
- (iii) there exist constants $C_1, C_2 \in (0, \infty)$ such that $\sup_{x \in \mathbb{R}^{p+1}} \sup_{\theta \in A} E[|a(y_{t+1}, \theta)| | x_t = x] f(x) \leq C_1$ and $\sup_{x \in \mathbb{R}^{p+1}} \sup_{\theta \in A} E[|\partial a(y_{t+1}, \theta) / \partial \theta'| | x_t = x] f(x) \leq C_2$ for each $t = p, \dots, T-1$.

\mathcal{D} -boundedness assumes boundedness of the conditional and unconditional (higher-order) moments of the function a and its derivative. These boundedness requirements guarantee sufficient regularity in order to apply a general uniform convergence theorem of Kristensen (2009); see Lemma A in Appendix A. Based on the above definition, Assumption A2 states the regularity conditions on the moment function u .

ASSUMPTION A2. Assume that

- (i) $\Theta \subset \mathbb{R}^k$ is compact and $\theta_0 \in \text{int}(\Theta)$.
- (ii) $E[u(y_{t+1}, \theta_0) | x_t] = 0$ almost surely for each $t = p, \dots, T-1$, and for each $\theta \in \Theta \setminus \{\theta_0\}$ and $t = p, \dots, T-1$, there exists a set $\mathbb{X}_\theta \subseteq \mathbb{R}^p$ such that $\Pr\{x_t \in \mathbb{X}_\theta\} > 0$ and $E[u(y_{t+1}, \theta) | x_t = x] \neq 0$ for all $x \in \mathbb{X}_\theta$.

- (iii) $E|x_t|^{1+s_0} < \infty$ for some $s_0 > 0$. $u(y, \theta)$ and $u(y, \theta)u(y, \theta)'$ are \mathcal{D} -bounded on Θ with order $s_1, s_2 > 2$, respectively. $\inf_{x \in \mathbb{R}^p} \inf_{\theta \in \Theta} \min \text{eig}(V(x, \theta)) > 0$ and $\sup_{x \in \mathbb{R}^p} \sup_{\theta \in \Theta} \max \text{eig}(V(x, \theta)) < \infty$. For each $\theta \in \Theta$ and $t = p, \dots, T-1$, the derivatives of $f(x)$, $V(x, \theta)f(x)$, and $E[u(y_{t+1}, \theta)|x_t = x]f(x)$ with respect to x are uniformly continuous and bounded on \mathbb{R}^p .
- (iv) There exists a neighborhood \mathcal{N} around θ_0 such that $\frac{\partial u}{\partial \theta_{k_1}}$ and $\frac{\partial^2 u}{\partial \theta_{k_1} \partial \theta_{k_2}}$ are \mathcal{D} -bounded on \mathcal{N} with order $s_3, s_4 > 2$ for each $k_1, k_2 = 1, \dots, k$, respectively, and the derivatives of $E\left[\frac{\partial u(y_{t+1}, \theta)}{\partial \theta_{k_1}} \Big| x_t = x\right]f(x)$ and $E\left[\frac{\partial^2 u(y_{t+1}, \theta)}{\partial \theta_{k_1} \partial \theta_{k_2}} \Big| x_t = x\right]f(x)$ with respect to x are uniformly continuous and bounded on \mathbb{R}^p for each $\theta \in \mathcal{N}$, $t = p, \dots, T-1$, and $k_1, k_2 = 1, \dots, k$. For some $\omega \in \mathbb{N}$, the ω -th order derivatives⁵ of (each element of) $f(x)$, $V(x, \theta_0)f(x)$, and $D(x, \theta_0)f(x)$ with respect to x are uniformly continuous and bounded on \mathbb{R}^p . For each $t = p, \dots, T-1$, the derivative of $E\left[\frac{\partial u(y_{t+1}, \theta_0)}{\partial \theta'} u(y_{t+1}, \theta_0)' \Big| x_t = x\right]f(x)$ with respect to x is uniformly continuous and bounded on \mathbb{R}^p .

Assumption A2 (i) is standard. Assumption A2 (ii) provides the identification condition for the true parameter θ_0 . Assumption A2 (iii) lists regularity conditions for the consistency of the LGMM estimator. Assumption A2 (iv) contains additional conditions to derive the asymptotic normality.

Let $a^\mu = \prod_{j=1}^p a_j^{\mu_j}$ for p -vectors a and μ of real values and nonnegative integers, respectively. Define $\delta_T = \inf_{|x| \leq c_T} f(x)$. The conditions for the kernel function and bandwidth are summarized in the following assumption.

ASSUMPTION A3. Assume that

- (i) \mathbb{K} is uniformly bounded and satisfies $\int \mathbb{K}(a) da = 1$, $\int a^\mu \mathbb{K}(a) da = 0$ for each μ with $1 \leq \sum_{j=1}^p \mu_j \leq \omega - 1$, and $\int |a^\mu \mathbb{K}(a)| da < \infty$ for each μ with $\sum_{j=1}^p \mu_j = \omega$. Also, there exist $C_K, L_K \in (0, \infty)$ such that either (a) $\mathbb{K}(a) = 0$ for all $|a| > L_K$ and $|\mathbb{K}(a) - \mathbb{K}(a')| \leq C_K |a - a'|$ for all a and a' , or (b) $\mathbb{K}(a)$ has a uniformly bounded derivative and $|\partial \mathbb{K}(a) / \partial a| \leq C_K |a|^{-v_K}$ for all $|a| \geq L_K$ and some $v_K > 1$.
- (ii) $h = h_T$ satisfies $\log T / (Th^p) \rightarrow 0$, $\log T / (\delta_T^2 T^{1/2} h^p) \rightarrow 0$, $h^p / \delta_T \rightarrow 0$, and $T^{1/4} h^\omega / \delta_T \rightarrow 0$ as $T \rightarrow \infty$. Also, $c_T \propto T^\xi$ for some $0 < \xi < \infty$.

Assumption A3 (i) imposes restrictions on the shape of the kernel function and is based on Kristensen (2009, Assumption A.6.1). To obtain reasonably fast convergence rates for bias components in kernel estimators, we use a higher-order (ω -th order) kernel. Combined with the condition

⁵Here the ω -th order derivatives of $a(x)$ with respect to $x = (x_1, \dots, x_p)'$ mean all derivatives $\frac{\partial^\omega a(x)}{\partial x_1^{\omega_1} \dots \partial x_p^{\omega_p}}$ with $\omega_1, \dots, \omega_p \in \mathbb{N}$ satisfying $\omega_1 + \dots + \omega_p = \omega$.

$T^{1/4}h^\omega/\delta_T \rightarrow 0$ in Assumption A3 (ii), the bias components of the kernel estimators are typically of order $o(T^{-1/4})$. Assumption A3 (ii) contains conditions for the bandwidth h and trimming constant c_T , which also controls δ_T . The first condition $\log T/(Th^p) \rightarrow 0$ comes from Kristensen (2009, Theorem 1) by setting $\theta = 1$ in his notation, which follows from exponential decay of the mixing coefficients in Assumption A1. The second condition $\log T/(\delta_T^2 T^{1/2} h^p) \rightarrow 0$ is required to control the variance components for kernel estimators. The third and fourth conditions are used to control the bias components of kernel estimators. The last condition is on the trimming constant c_T . This condition is not as weak as it seems for two reasons. First, the weak requirement on ξ is due to the exponentially decaying mixing coefficients in Assumption A1. If we allow polynomial decay in the mixing coefficients, we will have some upper bound for ξ . Second, since the rate of c_T determines the rate of δ_T , we cannot choose arbitrary the rate for c_T .

The first-order asymptotic properties of the LGMM estimator are established in the following theorem.

THEOREM 1.

(a) Under Assumptions A1, A2 (i)-(iii), and A3,

$$\hat{\theta}_{LGMM} \xrightarrow{p} \theta_0.$$

(b) Under Assumptions A1-A3,

$$\sqrt{T} \left(\hat{\theta}_{LGMM} - \theta_0 \right) \xrightarrow{d} N \left(0, \mathcal{I}(\theta_0)^{-1} \right),$$

$$\text{where } \mathcal{I}(\theta_0) = E \left[D(x_t, \theta_0)' V(x_t, \theta_0)^{-1} D(x_t, \theta_0) \right].$$

Theorem 1 demonstrates the consistency and asymptotic normality of the LGMM estimator. Also, the form of the asymptotic variance $\mathcal{I}(\theta_0)^{-1}$ implies that the LGMM estimator attains the semi-parametric efficiency bound for p -th order Markov models derived in Carrasco and Florens (2004). The basic strategy of the proof of Theorem 1 is a modest modification of the arguments in Kitamura, Tripathi and Ahn (2004) and Antoine, Bonnal and Renault (2007) to our time series context using a uniform convergence theorem of Kristensen (2009) and central limit theorem for U -statistics of Fan and Li (1999). Although the result in this theorem is new in the literature, the main focus of the paper is to formally investigate the bias reduction property of the LGMM estimator in a time series model using higher-order analysis.

An alternative approach that achieves semi-parametric efficiency is based on the unconditional GMM estimator (4) with the optimal instruments $\mathcal{A}(x_t, \theta_0) = D(x_t, \theta_0)' V(x_t, \theta_0)^{-1}$ (Chamberlain, 1987, and Newey, 1993, for *iid* data; Carrasco and Florens, 2004, for dependent data). As mentioned

in the introduction, Carrasco, Chernov, Florens and Ghysels (2007), Kuersteiner (2001, 2002) and West, Wong and Anatolyev (2009) have developed efficient estimators in more general setups than ours. Our Markov framework, however, delivers analytical tractability of the estimator that allows us to study the finite-sample behavior of the estimator by higher-order expansions. Also, while the implementation of some efficient estimators could be quite involved, the LGMM estimator is characterized by some appealing properties from a practical point of view such as computational simplicity and speed.

3.2 Finite-Sample Performance of LGMM Estimator

To get some initial idea about the numerical properties of the LGMM estimator, we generate data from the AR(1) model

$$r_{t+1} = \theta_0 r_t + u_{t+1}, \quad u_{t+1} = \sigma_{t+1} \varepsilon_{t+1},$$

for each $t = 0, 1, \dots, T$, where $\varepsilon_{t+1} \sim iidN(0, 1)$ and r_0 is initialized from its stationary distribution. The skedastic function σ_{t+1} is parametrized as an ARCH(1) process $\sigma_{t+1} = \sqrt{(1 - \beta_0) + \beta_0 u_t^2}$ with $\beta_0 = 0, 0.5, \text{ and } 0.9$.⁶ The AR parameter θ_0 is set equal to 0.4, 0.7, and 0.95.

Although the true value of the intercept is zero, we do not impose this restriction in the estimation and the estimated model includes an intercept. The number of Monte Carlo replications is 10,000. In the practical implementation of the LGMM estimation, we standardize the conditioning variable and use the Gaussian kernel, which satisfies Assumption A3 (i). For the bandwidth parameter, we employ the rule of thumb choice of $h = 1.06T^{-1/5}$ but our preliminary simulation results suggest that the estimator is relatively insensitive to a wide range of bandwidth values. The LGMM estimation is performed without any trimming, i.e. $\mathbb{I}_{tT} = 1$ for all t .⁷

In addition to the LGMM estimator, we consider the OLS estimator and the infeasible GLS estimator which uses the true skedastic function as GLS weights. Furthermore, we compare the LGMM estimator to two other efficient estimators: optimal instrumental variables estimator for heteroskedastic AR models (Kuersteiner, 2002; West, Wong and Anatolyev, 2009) and Carrasco and Florens' (2000) efficient estimator for moment condition models. For the estimator of Carrasco and Florens (2000) (see also, Carrasco, Chernov, Florens and Ghysels, 2007, for the extension of the estimator to dependent data), we use a smoothing parameter (α in their notation) of 0.02 and a standard normal integrating density.^{8,9}

⁶In an earlier version of the paper, we also considered different forms of conditional heteroskedasticity. The results are qualitatively very similar and are available from the authors upon request.

⁷Our preliminary experiments showed that the effect of trimming in the simulation setups considered in the paper is negligible.

⁸We would like to thank Marine Carrasco for generously providing the codes for implementing the Carrasco-Florens estimator.

⁹The OLS estimator can be regarded as the simplest example of a non-local GMM estimator based on the un-

The mean bias, median bias, standard deviation, and root mean squared error (RMSE) for all estimators are reported in Tables 1 and 2 for sample sizes $T = 100$ and 500 , respectively. As expected, the OLS estimator is characterized by a substantial downward bias that seems to increase with the degree of conditional heteroskedasticity. The optimal instrumental variable estimator exhibits similar behavior and in most cases its bias even exceeds the OLS bias which is consistent with the simulation results reported in Kuersteiner (2002). While the other three estimators are also biased, the magnitude of their bias is smaller and the bias tends to decrease as the conditional heteroskedasticity becomes stronger. Not surprisingly, the infeasible GLS estimator delivers the smallest standard errors among all estimators. Also, the GLS estimator provides a nontrivial bias correction and in several cases its bias is less than half of the OLS bias.

Compared to the OLS and optimal instrumental variables estimators, the Carrasco-Florens estimator also produces a bias reduction which is most effective when the persistence of the process is small and the conditional heteroskedasticity is strong. In terms of efficiency and RMSE, the Carrasco-Florens estimator substantially dominates the optimal instrumental variables estimator. Interestingly, the LGMM estimator is characterized by the smallest bias across all parametrization. The bias reduction for the LGMM estimator is particularly pronounced as the degree of conditional heteroskedasticity increases and in the ARCH case with $\beta_0 = 0.9$ (the last panels in Tables 1 and 2), this estimator is almost median unbiased. Furthermore, the LGMM estimator enjoys some significant efficiency gains and dominates in several cases the infeasible GLS in terms of RMSE.

One interesting finding that emerges from this simulation experiment is that the LGMM estimator tends to reduce the bias of the slope parameter even in the conditionally homoskedastic case (see the first panels of Tables 1 and 2) where the conditioning of the error term on past information appears unnecessary. In the next section, we further investigate this bias reduction property of the LGMM estimator by higher-order asymptotic analysis.

4 Higher-Order Analysis

4.1 Stochastic Expansion in AR(1) Model with IID Errors

In this section, we undertake a stochastic expansion to study the higher-order properties of the LGMM estimator by specializing the model of interest to an AR(1) process with *iid* errors. There are two main reasons for this simplification. First, it is well documented that this model could cause a large estimation bias as the persistence of the process increases and the bias properties of

conditional moment restriction $E[r_t(r_{t+1} - \theta_0 r_t)] = 0$. The GLS and other semi-parametrically efficient estimators utilize all information from the conditional moment restriction $E[r_{t+1} - \theta_0 r_t | r_t] = 0$ to estimate θ_0 . Thus, these two approaches provide two extreme treatments of the information contained in the conditional moment, and other non-local GMM estimators based on the unconditional moment restriction $E[v(r_t)(r_{t+1} - \theta_0 r_t)] = 0$ with some vector of instruments $v(r_t)$ can be considered as intermediate cases.

the OLS estimator are analytically derived and widely discussed in the literature (Marriott and Pope, 1954; Kendall, 1954; among others). Second, the *iid* assumption on the errors simplifies the LGMM objective function and allows us to focus on a closed form expression for the LGMM estimator. Interestingly, our simulation results reported below suggest that the LGMM estimator enjoys smaller bias (and MSE) than the OLS estimator even when conditioning (or smoothing) is unnecessary due to the *iid* structure of the errors. This result bears some resemblance to the bias reduction of the smoothed generalized empirical likelihood estimator in serially uncorrelated models shown by Anatolyev (2005) although it is derived in a completely different context and framework.

Suppose that the data are generated by a zero-mean AR(1) model

$$r_{t+1} = \theta_0 r_t + u_{t+1}, \quad (7)$$

for each $t = 1, \dots, T$, where $u_t \sim iid(0, 1)$, and the conditional moment restriction (1) can be obtained by defining $u(y_{t+1}, \theta_0) = r_{t+1} - \theta_0 r_t$ with $y_{t+1} = (r_{t+1}, r_t)'$ and $x_t = r_t$.¹⁰ The case with deterministic terms can be analyzed by decomposing the lagged and deterministic regressors as in van Giersbergen (2005) although the possible bias reduction of the LGMM estimator are expected to come only from the localized weighting of the stochastic components.

To highlight the effect of smoothing on the moment functions, we compare the OLS estimator $\hat{\theta}_{OLS} = \arg \min_{\theta \in \Theta} \sum_{t=1}^{T-1} u(y_{t+1}, \theta)^2$ and the LGMM estimator with a constant weight matrix

$$\hat{\theta}_{LGMM1} = \arg \min_{\theta \in \Theta} \sum_{t=1}^{T-1} \mathbb{I}_{tT} u_T(x_t, \theta)^2 = \arg \min_{\theta \in \Theta} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=1}^{T-1} w_{tj} (r_{j+1} - \theta r_j) \right]^2. \quad (8)$$

Note that the difference between $\hat{\theta}_{LGMM1}$ and $\hat{\theta}_{OLS}$ is whether we smooth the moment function $u(y_{t+1}, \theta)$ or not. One convenient feature of $\hat{\theta}_{LGMM1}$ is that it is written by the explicit form

$$\hat{\theta}_{LGMM1} = \frac{\sum_{t=1}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=1}^{T-1} w_{tj} r_j \right] \left[\sum_{j=1}^{T-1} w_{tj} r_{j+1} \right]}{\sum_{t=1}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=1}^{T-1} w_{tj} r_j \right]^2} = \theta_0 + \frac{\sum_{t=1}^{T-1} \mathbb{I}_{tT} \hat{r}_t \hat{u}_{t+1}}{\sum_{t=1}^{T-1} \mathbb{I}_{tT} \hat{r}_t^2}, \quad (9)$$

where $\hat{r}_t = \sum_{j=1}^{T-1} w_{tj} r_j$ and $\hat{u}_{t+1} = \sum_{j=1}^{T-1} w_{tj} u_{j+1}$.

For the OLS estimator, an analogous expression to (9) is obtained as

$$\hat{\theta}_{OLS} = \theta_0 + \frac{\sum_{t=1}^{T-1} r_t u_{t+1}}{\sum_{t=1}^{T-1} r_t^2}.$$

¹⁰The focus of the higher-order analysis here is to gain some insights about the bias properties of the LGMM and OLS estimators in a simple setup and does not target the asymptotically efficient estimator. Thus, we do not augment the conditional moment restriction $E[u(y_{t+1}, \theta_0) | x_t] = 0$ with the homoskedasticity assumption $E[u(y_{t+1}, \theta_0)^2 | x_t] = 1$ when we estimate θ_0 . The homoskedasticity assumption is only used to simplify the higher-order analysis and make the comparison of the LGMM and OLS more intuitive.

Let $s_r^2 = \frac{1}{T} \sum_{t=1}^{T-1} r_t^2$ and $\sigma_r^2 = E[r_{t-1}^2] = \frac{1}{1-\theta_0^2}$. An expansion of $(\hat{\theta}_{OLS} - \theta_0)$ around $s_r^2 = \sigma_r^2$ yields¹¹

$$\hat{\theta}_{OLS} - \theta_0 = A_{OLS} - B_{OLS} + O_p\left(T^{-3/2}\right), \quad (10)$$

where

$$A_{OLS} = \frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} r_t u_{t+1}, \quad B_{OLS} = \frac{s_r^2 - \sigma_r^2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} r_t u_{t+1}.$$

Note that $E[A_{OLS}] = 0$ in model (7). Simple algebraic manipulations (see Davidson, 2000, for example) yield

$$E[B_{OLS}] = \frac{2\theta_0}{T}.$$

Therefore, as derived in Marriott and Pope (1954) and Kendall (1954), the higher-order bias of the OLS estimator under the model (7) can be expressed as

$$E[\hat{\theta}_{OLS}] - \theta_0 = -\frac{2\theta_0}{T} + O\left(T^{-3/2}\right). \quad (11)$$

This expansion suggests that the OLS estimator tends to have a negative finite sample bias when θ_0 is positive.

We now derive a stochastic expansion of the LGMM estimator with the unit weight matrix $\hat{\theta}_{LGMM1}$. Recall that $f(\cdot)$ is the marginal density function of $x_t = r_t$, and denote

$$\begin{aligned} f_t &= f(r_t), \quad \tilde{u}_{t+1} = \frac{1}{Th} \sum_{j=1}^{T-1} \mathbb{K}\left(\frac{r_t - r_j}{h}\right) u_{j+1}, \\ \tilde{f}_t &= \frac{1}{Th} \sum_{j=1}^{T-1} \mathbb{K}\left(\frac{r_t - r_j}{h}\right), \quad \bar{f}_t = \frac{1}{h} \int \mathbb{K}\left(\frac{r_t - r}{h}\right) f(r) dr, \\ \tilde{r}_t &= \frac{1}{Th} \sum_{j=1}^{T-1} \mathbb{K}\left(\frac{r_t - r_j}{h}\right) r_j, \quad \bar{r}_t = \frac{1}{h} \int \mathbb{K}\left(\frac{r_t - r}{h}\right) r f(r) dr, \\ V_{f,t} &= \tilde{f}_t - \bar{f}_t, \quad B_{f,t} = \bar{f}_t - f_t, \quad V_{r,t} = \tilde{r}_t - \bar{r}_t, \quad B_{r,t} = \bar{r}_t - r_t. \end{aligned}$$

Based on this notation, we make the following assumptions.

ASSUMPTION A4. Assume that

- (i) $E|u_t|^4 < \infty$ and $E[f_t^{-8}] < \infty$ for each $t = 1, \dots, T$.
- (ii) There exists a positive sequence $\{a_T\}_{T \geq 1}$ such that $a_T \rightarrow 0$ and $T^{1/2}a_T \rightarrow \infty$ as $T \rightarrow \infty$, and

$$\sup_{\{t: \mathbb{1}_{tT}=1\}} |\tilde{u}_t| = O_p(a_T), \quad \sup_{\{t: \mathbb{1}_{tT}=1\}} |V_{f,t}| = O_p(a_T), \quad \sup_{\{t: \mathbb{1}_{tT}=1\}} |V_{r,t}| = O_p(a_T).$$

¹¹The remainder term $O_p(T^{-3/2})$ follows from $\frac{1}{T} \sum_{t=1}^T r_t u_{t+1} = O_p(T^{-1/2})$ and $s_r^2 - \sigma_r^2 = O_p(T^{-1/2})$ by the central limit theorem.

(iii) There exists a positive sequence $\{b_T\}_{T \geq 1}$ such that $b_T \rightarrow 0$ and $T^{1/2}b_T \rightarrow \infty$ as $T \rightarrow \infty$, and

$$\sup_{\{t: \mathbb{I}_{tT}=1\}} |B_{f,t}| = O(b_T), \quad \sup_{\{t: \mathbb{I}_{tT}=1\}} |B_{r,t}| = O(b_T).$$

Assumption A4 (i) is concerned with the existence of higher-order moments. These assumptions are used to derive the law of large numbers, such as $\frac{1}{T} \sum_{t=1}^{T-1} \left| \frac{r_t^2}{f_t^4} \right| \xrightarrow{p} E \left| \frac{r_t^2}{f_t^4} \right| < \infty$ to guarantee $\frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{r_t^2}{f_t^4} \right| = O_p(1)$. Assumption A4 (ii) and (iii) contain higher level conditions for the variance and bias components of the kernel estimators, respectively. Lemma A in Appendix A provides primitive conditions that guarantee the validity of these assumptions. Let $\hat{s}_r^2 = \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} r_t^2$. Based on these assumptions, we obtain the following stochastic expansion of $\hat{\theta}_{LGMM1}$.

THEOREM 2. *Under the model (7) and Assumption A4,*

$$\hat{\theta}_{LGMM1} - \theta_0 = A_{LGMM} - B_{LGMM} + C_{LGMM} + D_{LGMM} + O_p\left(a_T(a_T + b_T)^2\right), \quad (12)$$

where

$$\begin{aligned} A_{LGMM} &= \frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t \tilde{u}_{t+1}}{f_t}, & B_{LGMM} &= \frac{\hat{s}_r^2 - \sigma_r^2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t \tilde{u}_{t+1}}{f_t}, \\ C_{LGMM} &= \frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{(V_{r,t} + B_{r,t}) \tilde{u}_{t+1}}{f_t^2} - \frac{2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t (V_{r,t} + B_{r,t})}{f_t} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t \tilde{u}_{t+1}}{f_t}, \\ D_{LGMM} &= -\frac{2}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{(V_{f,t} + B_{f,t}) r_t \tilde{u}_{t+1}}{f_t^2} + \frac{2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t^2 (V_{f,t} + B_{f,t})}{f_t} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t \tilde{u}_{t+1}}{f_t}. \end{aligned}$$

We first check the stochastic orders of these terms. The first term A_{LGMM} corresponds to A_{OLS} in (10). By applying a U -statistic argument as in the proof of Theorem 1 (b), we typically have $A_{LGMM} = O_p(T^{-1/2})$. The second term B_{LGMM} corresponds to B_{OLS} in (10). Since $\hat{s}_r^2 - \sigma_r^2 = O_p(T^{-1/2})$ and $\frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t \tilde{u}_{t+1}}{f_t} = O_p(T^{-1/2})$ by the central limit theorem and a U -statistic argument, respectively, we typically have $B_{LGMM} = O_p(T^{-1})$. The third term C_{LGMM} arises from the correlation among $V_{r,t} + B_{r,t}$, \tilde{u}_{t+1} , and r_t . Under Assumption A4, this term satisfies $C_{LGMM} = O_p(a_T(a_T + b_T))$. Similarly, the fourth term D_{LGMM} arises from the correlation among $V_{f,t} + B_{f,t}$, \tilde{u}_{t+1} , and r_t , and satisfies $D_{LGMM} = O_p(a_T(a_T + b_T))$ under Assumption A4.

Based on the stochastic orders of these terms, we focus on the dominant term A_{LGMM} and compare it with the OLS counterpart A_{OLS} . For an intuitive argument, let us neglect the trimming term \mathbb{I}_{tT} in A_{LGMM} and consider $A_{LGMM}^* = \frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \frac{r_t \tilde{u}_{t+1}}{f_t}$. Provided that $\sup_{1 \leq t \leq T} |\mathbb{I}_{tT} - 1|$ converges to zero sufficiently fast (which is guaranteed if $c_T \propto T^\xi$ and $E|x_t|^\zeta < \infty$ for sufficiently large ξ and ζ), A_{LGMM}^* can serve as a reasonable approximation to A_{LGMM} . The expectation $E[A_{LGMM}^*]$ can be approximated as follows. Let $f_u(\cdot)$ be the density function of u_t .

THEOREM 3. Suppose that the model (7) and Assumptions A2 (iv) and A3 hold. Then,

$$E[A_{LGMM}^*] = \frac{1 - \theta_0^2}{T^2} \sum_{t=1}^{T-1} \left[\frac{1}{(1 - \theta_0)^2} + \frac{\theta_0}{(1 - \theta_0^2)^2} + \cdots + \frac{\theta_0^{t-2}}{(1 - \theta_0^{t-1})^2} \right] + O(T^{-1}h^\omega). \quad (13)$$

Note that in contrast to $E[A_{OLS}] = 0$, the expectation $E[A_{LGMM}^*]$ tends to be positive and is of order $O(T^{-1})$ when $0 \leq \theta_0 < 1$. This is expected since the weighting scheme uses information from neighboring observations and destroys the zero correlation between r_t and u_t . Interestingly, this proves to be advantageous for the LGMM estimator since the positive value of $E[A_{LGMM}^*]$ tends to offset the second bias term $E[B_{LGMM}]$ in the expansion. Also, it is worth noting that despite the fact that the bandwidth parameter has only a second-order effect on the magnitude of $E[A_{LGMM}^*]$, a data-driven (simulation- or bootstrap-based) choice of h can be used to reduce or even completely eliminate the $O(T^{-1})$ terms in the bias expansion.

Although it is difficult to evaluate the second term $E[B_{LGMM}]$ without specifying the distributional form of f_u , the stochastic order $B_{LGMM} = O_p(T^{-1})$ implies $E[B_{LGMM}] = O(T^{-1})$ if B_{LGMM} is uniformly integrable. Thus, as long as $E[B_{LGMM}]$ shows a similar behavior as $E[B_{OLS}]$ in finite samples, the positive bias term $E[A_{LGMM}^*]$ may compensate for the negative bias caused by $E[B_{LGMM}]$. We evaluate this conjecture by simulation in the next subsection.

4.2 Performance of LGMM Estimator with Unit Weight

To assess the finite-sample properties of the LGMM estimator with the unit weight in models with *iid* errors, we conduct a small Monte Carlo experiment. The data are generated from the zero-mean AR(1) model

$$r_{t+1} = \theta_0 r_t + u_{t+1},$$

for each $t = 1, \dots, T$, where $u_{t+1} \sim iidN(0, 1)$ and $\theta_0 = 0.95$. As in Section 3.2, the weights for the LGMM estimator are obtained from the Gaussian kernel and plug-in bandwidth and the results are based on 5,000 Monte Carlo replications. Unlike the experiment in Section 3.2, however, the zero intercept is imposed in the estimation in order to be consistent with the theoretical framework adopted above. The mean bias, standard deviation and RMSE of the three estimators are reported in Table 3 for sample sizes $T = 50, 100, 200, 400, 800, 1600$, and 3200.

To illustrate the source of the bias reduction in the LGMM estimator, we also numerically evaluate the two leading terms in the bias expansion for the OLS and LGMM estimators. In particular, the last two columns of Table 3 report the Monte Carlo averages of $\frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} r_t u_{t+1}$ (to evaluate $E[A_{OLS}]$) and $\frac{s_r^2 - \sigma_r^2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} r_t u_{t+1}$ (to evaluate $E[B_{OLS}]$) regarding the OLS estimator and $\frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} r_t (\tilde{u}_{t+1}/f_t)$ (to evaluate $E[A_{LGMM}]$) and $\frac{\tilde{s}_r^2 - \sigma_r^2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} r_t (\tilde{u}_{t+1}/f_t)$ (to evaluate $E[B_{LGMM}]$) regarding the LGMM estimator.

The numerical results in Table 3 support our theoretical results in the previous subsection. The LGMM enjoys a substantially smaller mean bias which goes to zero much faster than the OLS estimator. The bias reduction property of LGMM along with its efficiency result in a lower RMSE for small sample sizes although the difference in RMSE converges to zero as the sample size increases.¹² As conjectured above, the terms $E[B_{OLS}]$ and $E[B_{LGMM}]$ for the OLS and LGMM estimators are of similar magnitude and the better bias properties of the LGMM estimator arise from the fact that the positive value of $E[A_{LGMM}]$ tends to offset the downward bias effect of $E[B_{LGMM}]$.

In order to visualize the bias of the LGMM estimator over the stationary region of the AR parameter space, Figure 1 plots the bias functions of the OLS and LGMM estimators for $\theta_0 = -0.95, -0.85, \dots, 0.85, 0.95$ for two sample sizes: $T = 100$ and 200 . As our higher-order analysis in Section 4.1 suggests, the bias reduction property of the LGMM estimator is not expected to hold over the whole parameter space. For example, when $-1 < \theta_0 < 0$, the sign of $E[A_{LGMM}^*]$ in (13) cannot be determined unambiguously since the terms inside the square brackets of (13) are of alternating signs. For this reason, it would be interesting to see how the LGMM estimator performs over the negative part of the parameter space even though our main interest lies in positively autocorrelated processes.

To explore the effect of the smoothing parameter choice on the LGMM estimator, we present the results for three bandwidths: $h = T^{-1/5}$ (approximate *iid* bandwidth), $h = 0.6T^{-1/5}$ and $h = 1.2T^{-1/5}$ with a standardized conditioning variable. The mean bias is computed as a Monte Carlo average over 100,000 replications. Figure 1 shows that while the OLS bias increases linearly with the absolute value of θ_0 as expression (11) suggests, the bias function of the LGMM estimator appears to flatten out for $|\theta_0| \geq 0.5$. For $h = T^{-1/5}$ and $h = 1.2T^{-1/5}$, the bias of the LGMM estimator is substantially smaller than the OLS bias for $\theta_0 \geq 0.2$ and $\theta_0 \leq -0.6$ although the LGMM estimators with these bandwidths exhibit positive bias for θ_0 less than 0.2. While the LGMM bias exceeds the OLS bias over a part of the negative region $(-0.6, 0)$, the difference is relatively small. As expected, the bias of the LGMM estimator approaches the OLS bias as the bandwidth gets smaller.

In the positive part of the parameter space, which is of primary interest for our analysis, the LGMM demonstrates convincingly its higher-order advantages and appears practically unbiased for $T = 200$, $\theta_0 \in [0.2, 0.7]$ and $h = 1.2T^{-1/5}$. Finally, Figure 1 clearly suggests that the different levels of persistence seem to require different bandwidths (smaller bandwidths when θ_0 is near zero and

¹²In an earlier version of the paper, we also included the full maximum likelihood estimator (MLE) that incorporates information about the first observation in the likelihood function (Beach and MacKinnon, 1978). The standard deviation of the LGMM estimator was only slightly higher than the full MLE for all sample sizes but the LGMM estimator dominated the MLE in terms of RMSE. These results are available from the authors upon request.

larger bandwidths when θ_0 is near one) and a data-driven procedure for selecting the appropriate smoothing parameter would prove beneficial.

5 Economic Significance of LGMM

To evaluate the economic significance of the statistical properties of the LGMM estimator, we use this estimator for bond and derivative pricing with data generated from the CIR (Cox, Ingersoll and Ross, 1985) model

$$dr_t = \kappa(\mu - r_t)dt + \sigma r_t^{1/2}dW_t, \quad (14)$$

for each $t = 0, 1, \dots, T$. This model is convenient because the transition and marginal densities are known and the bond and call option prices are available in closed form (Cox, Ingersoll and Ross, 1985). 5,000 sample paths for the spot interest rate of length $T = 600$ observations are simulated using the procedure described in Chapman and Pearson (2000). After drawing an initial value from the marginal Gamma density, the interest rate process is constructed recursively by drawing random numbers from the transition non-central chi-square density and using the values for κ , μ and σ and a time step between two consecutive observation equal to $\Delta = 1/52$ that corresponds to weekly data. Carrasco, Chernov, Florens and Ghysels (2007) show that the CIR process (14) is absolutely regular.

We consider one of the parameter configurations that are used in Chapman and Pearson (2000) and are calibrated to the interest rate data in Aït-Sahalia (1996).¹³ In particular, $(\kappa, \mu, \sigma) = (0.21459, 0.085711, 0.0783)$ which implies a highly persistent interest rate process consistent with the observed data. The expressions for the price of a zero-coupon discount bond and a call option on a zero-coupon discount bond have an analytical form and are given in Cox, Ingersoll and Ross (1985). We follow Phillips and Yu (2005) and compute the prices of a three-year zero-coupon discount bond and a one-year European call option on a three-year discount bond with a face value of \$100 and an exercise price of \$87 given an initial interest rate of 5%. The bond and option prices are computed assuming that the market price of risk is equal to zero.

The parameters of model (3) are typically estimated from the discrete-time representation $u_{t+1} = \Delta r_{t+1} - \alpha_0 - \alpha_1 r_t$. The GMM estimator based on the unconditional moment restrictions $E[u_{t+1}] = 0$, $E[u_{t+1}r_t] = 0$ and $E[u_{t+1}^2] = \sigma^2 r_t$ is the GLS estimator used in Ball and Torous (1996) and Chapman and Pearson (2000). In addition to the GLS estimator, we consider the efficient estimator of Carrasco and Florens (2000) with the same smoothing parameter and integrating den-

¹³The results from the second parametrization in Chapman and Person (2000) are qualitatively similar and are not reported to preserve space. These results and some additional simulations for the specifications used in Phillips and Yu (2005) are available from the authors upon request.

sity as in Section 3.2.¹⁴ Note also that because the transition and marginal densities for the CIR model are available in closed form, one could perform exact maximum likelihood estimation as in Phillips and Yu (2005) and avoid the discretization bias. Since Phillips and Yu (2005) found that the effects of the discretization bias are small, we do not consider explicitly the maximum likelihood estimator and focus only on the GMM estimators.¹⁵ Similarly, we do not assume knowledge of the characteristic function (as in Carrasco, Chernov, Florens and Ghysels, 2007) in constructing the moment conditions for the Carrasco and Florens (2000) estimator. As a result, in estimating the parameters α_0 and α_1 (or equivalently, κ and μ), both the LGMM and the Carrasco-Florens estimators are based on the conditional moment restriction $E[u_{t+1}|r_t] = 0$. The unconditional variance σ^2 for LGMM and Carrasco-Florens estimators is then computed as a sample average of the corresponding squared residuals standardized by r_t and adjusted for Δ .

Table 4 reports the median, mean, standard deviation and root mean square error (RMSE) of the parameter estimates, bond and option prices. Due to the strong persistence of the data and the autoregressive structure of the discretized model, the parameter that measures the speed of mean reversion is estimated with a substantial upward bias. Since the option price is very sensitive to this parameter, the bias of the estimate of κ is translated into a downward bias in the option price which is further exacerbated by the highly nonlinear relationship between the option price and κ . While all estimators of κ are biased, the bias of the GLS and Carrasco-Florens estimators is twice as large as the bias of the conditional GMM estimator. Interestingly, the variance of the LGMM estimator of κ is approximately 30% smaller than the variance of the GLS and Carrasco-Florens estimators. This is typically not the case for some popular bias reduction techniques such as simulation-based and jackknife methods where the bias correction is accompanied with a higher variability of the estimates. As a result of the lower bias and increased efficiency, the RMSE of the LGMM, obtained as the square root of the sum of the squared mean bias and variance of the estimator, is more than 30% smaller than the RMSE of the other two estimators.

Similar results emerge from the estimates of the intercept in the model. In this case, the RMSE of the LGMM estimator is 32-36% lower than the RMSE of the other estimators. The performance of the LGMM estimator is even more impressive given the fact that, unlike the GLS, it does not utilize any knowledge of the conditional variance function and should be expected to possess some robustness in this respect.

Despite the excellent properties of the LGMM estimator of the intercept and slope parameters

¹⁴The results from the optimal instrumental variables estimator of Kuersteiner (2002) are not presented because the numerical performance of this estimator turned out to be highly unstable in our highly persistent setup and was dominated by the other estimators.

¹⁵While the applicability of the ML approach is limited by the fact that the likelihood function is available in closed form only for the CIR specification, the GMM estimators can accommodate a much larger class of models. Most importantly, the LGMM estimator does not require any knowledge of the parametric form of the diffusion function.

in the discrete-time AR model, the long-run mean μ in (14) is computed as a ratio of two estimates and exhibits some instability when α_1 (or equivalently κ) is close to 0. Since the LGMM estimator is less biased than the other estimators, it tends to produce more values of α_1 near 0 which leads to the higher variability of this estimator in Table 4. Similar findings have been reported by Ball and Torous (1996) who point out that this instability is particularly pronounced at low levels of the initial interest rate. Finally, the unconditional standard deviation of the error component is estimated with a slight upward bias and the differences in the performance of the estimators for this parameter are negligible.

The results for the bond and call option prices obtained by plugging the estimated parameters from the different methods are reported in the last two columns of Table 4. The bond price computed from the LGMM estimator appears to be the least biased and most efficient with a RMSE which is 20% lower than the RMSEs of the other two estimators. As pointed out above, the price of the call option is more sensitive to the CIR model parameters and the economic significance of the effects of the lower bias of the LGMM estimator is more evident. For instance, for the true option price of \$1.931, the average LGMM-based option price is \$1.652 compared to \$1.110 and \$1.127 for Carrasco-Florens and GLS estimator, respectively. The higher variability of the LGMM option price is due to two reasons. First, it partly arises from the higher variability of the estimate of μ which was caused by the near unit root specification of the interest rate process. More importantly, however, the lower variability of the other two option prices is somewhat artificial and is due to a large number of economically unreasonable option values around zero. Overall, the attractive properties of the LGMM estimator translate into large gains for pricing bonds and call options.

6 Conclusion

This paper studies the properties of the local GMM estimator in Markov models with conditional heteroskedasticity. We derive the conditions for the consistency, asymptotic normality, and semi-parametric efficiency of the local GMM estimator under this time series setup. The paper also undertakes a higher-order analysis based on an asymptotic expansion to study the bias properties of the local GMM estimator. Some interesting findings about the bias structure and the order of magnitudes of the leading terms in the expansion emerge from this analysis. The finite-sample performance of the LGMM estimator is evaluated in the context of bond and derivative pricing with simulated data from a diffusion model of spot interest rate. In summary, the semi-parametric efficiency, smaller bias, and computational simplicity of the local GMM estimator prove to be very desirable and appealing properties from both theoretical and practical perspectives for estimating time series models defined by conditional moment restrictions.

This paper can be considered as a starting point toward the goal of better understanding the

finite-sample properties of the LGMM estimator for general time series models. Several natural extensions of our results on the bias properties of the unit weight LGMM in AR(1) models include higher-order analyses of (i) optimally weighted LGMM estimator, (ii) times series models with more general dependence structure, and (iii) variance of the LGMM estimator. Since these extensions require rather different and advanced technical arguments, we leave the analysis of these interesting issues for future research.

A Appendix: Derivations and Mathematical Proofs

Hereafter, let $\hat{\theta} = \hat{\theta}_{LGMM}$, “ \sup_t ” denote “ $\sup_{t \in \{t: \mathbb{I}_{tT}=1\}}$ ”, and “w.p.a.1” signify “with probability approaching one.” We repeatedly use the following uniform convergence result by Kristensen (2009, Theorem 1).

LEMMA A. *Suppose Assumptions A1, A2 (i), and A3 (i) hold and the function $a : \mathbb{R}^{p+1} \times A$ is \mathcal{D} -bounded on A with order $s > 2$. If $\log T / (Th^p) \rightarrow 0$ as $T \rightarrow \infty$, then for any $0 < \xi < \infty$,*

$$\sup_{|x| \leq c_T} \sup_{\theta \in A} \left| \frac{1}{(T-p)h^p} \sum_{j=p}^{T-1} \mathbb{K} \left(\frac{x_j - x}{h} \right) a(y_{j+1}, \theta) - E \left[\frac{1}{h^p} \mathbb{K} \left(\frac{x_j - x}{h} \right) a(y_{j+1}, \theta) \right] \right| = O_p \left(\sqrt{\frac{\log T}{Th^p}} \right)$$

with $c_T \propto T^\xi$.

A.1 Proof of Theorem 1

Proof of Part (a): The objective function of the LGMM estimator and its population counterpart are written respectively as

$$\begin{aligned} Q_T(\theta) &= \frac{1}{T-p} \sum_{t=p}^{T-1} \mathbb{I}_{tT} u_T(x_t, \theta)' V_T(x_t, \theta)^{-1} u_T(x_t, \theta), \\ Q(\theta) &= E \left[E[u(y_{t+1}, \theta) | x_t]' V(x_t, \theta)^{-1} E[u(y_{t+1}, \theta) | x_t] \right], \end{aligned}$$

where $V(x, \theta)^{-1}$ exists for each $x \in \mathbb{R}^p$ and $\theta \in \Theta$ from Assumption A2 (iii), and $V_T(x_t, \theta)^{-1}$ exists w.p.a.1 for each $t \in \{t : \mathbb{I}_{tT} = 1\}$ and $\theta \in \Theta$ from Assumption A2 (iii) and (16) below. From Newey and McFadden (1994, Theorem 2.1), it is sufficient to show that

- (i) Θ is compact (assumed in Assumption A2 (i)),
- (ii) $Q(\theta)$ is uniquely minimized at θ_0 (implied by Assumption A2 (ii)),
- (iii) $Q(\theta)$ is continuous on Θ (implied by Assumption A2 (iii)),
- (iv) $\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0$.

Thus, it remains to show (iv), which follows from

$$\sup_t \sup_{\theta \in \Theta} |u_T(x_t, \theta) - E[u(y_{t+1}, \theta) | x_t]| \xrightarrow{p} 0, \quad (15)$$

$$\sup_t \sup_{\theta \in \Theta} |V_T(x_t, \theta) - V(x_t, \theta)| \xrightarrow{p} 0, \quad (16)$$

$$\sup_{\theta \in \Theta} \left| \frac{1}{T-p} \sum_{t=p}^{T-1} \mathbb{I}_{tT} E[u(y_{t+1}, \theta) | x_t]' V(x_t, \theta)^{-1} E[u(y_{t+1}, \theta) | x_t] - Q(\theta) \right| \xrightarrow{p} 0. \quad (17)$$

To show (15), define $\hat{f}(x) = \frac{1}{(T-p)h^p} \sum_{j=p}^{T-1} \mathbb{K}\left(\frac{x_j-x}{h}\right)$ and $\hat{u}(x, \theta) = \frac{1}{(T-p)h^p} \sum_{j=p}^{T-1} \mathbb{K}\left(\frac{x_j-x}{h}\right) u(y_{j+1}, \theta)$. From Lemma A with $a(y_{j+1}, \theta) = 1$ and $u(y_{j+1}, \theta)$,

$$\sup_{|x| \leq c_T} \left| \hat{f}(x) - E[\hat{f}(x)] \right| = O_p\left(\sqrt{\frac{\log T}{Th^p}}\right),$$

$$\sup_{|x| \leq c_T} \sup_{\theta \in \Theta} |\hat{u}(x, \theta) - E[\hat{u}(x, \theta)]| = O_p\left(\sqrt{\frac{\log T}{Th^p}}\right).$$

By a change of variables ($a = \frac{x_j-x}{h}$) and an expansion around $a = 0$,

$$\sup_{|x| \leq c_T} \left| E[\hat{f}(x)] - f(x) \right| = h^p \sup_{|x| \leq c_T} \left| \int \mathbb{K}(a) \frac{df(x + \bar{a}h)}{da} da \right| = O(h^p),$$

where \bar{a} is a point on the line joining a and 0. A similar argument with the law of iterated expectations yields

$$\sup_{|x| \leq c_T} \sup_{\theta \in \Theta} |E[\hat{u}(x, \theta)] - E[u(y_{t+1}, \theta) | x_t = x] f(x)| = O(h^p).$$

Combining these results, we obtain

$$\sup_{|x| \leq c_T} \left| \hat{f}(x) - f(x) \right| = O_p\left(\sqrt{\frac{\log T}{Th^p}}\right) + O(h^p),$$

$$\sup_{|x| \leq c_T} \sup_{\theta \in \Theta} |\hat{u}(x, \theta) - E[u(y_{t+1}, \theta) | x_t = x] f(x)| = O_p\left(\sqrt{\frac{\log T}{Th^p}}\right) + O(h^p).$$

Since the denominator and numerator of $u_T(x, \theta) = \frac{\hat{u}(x, \theta)/\hat{f}(x)}{f(x)/f(x)}$ satisfy

$$\sup_{|x| \leq c_T} \left| \frac{\hat{f}(x)}{f(x)} - 1 \right| \leq \frac{\sup_{|x| \leq c_T} |\hat{f}(x) - f(x)|}{\inf_{|x| \leq c_T} f(x)} = O_p\left(\sqrt{\frac{\log T}{\delta_T^2 Th^p}}\right) + O\left(\frac{h^p}{\delta_T}\right),$$

and

$$\begin{aligned} \sup_{|x| \leq c_T} \sup_{\theta \in \Theta} \left| \frac{\hat{u}(x)}{f(x)} - E[u(y_{t+1}, \theta) | x_t = x] \right| &\leq \frac{\sup_{|x| \leq c_T} \sup_{\theta \in \Theta} |E[\hat{u}(x, \theta)] - E[u(y_{t+1}, \theta) | x_t = x] f(x)|}{\inf_{|x| \leq c_T} f(x)} \\ &= O_p\left(\sqrt{\frac{\log T}{\delta_T^2 Th^p}}\right) + O\left(\frac{h^p}{\delta_T}\right), \end{aligned}$$

Assumption A3 (ii) guarantees that

$$\sup_t \sup_{\theta \in \Theta} |u_T(x_t, \theta) - E[u(y_{t+1}, \theta) | x_t]| = O_p\left(\sqrt{\frac{\log T}{\delta_T^2 Th^p}}\right) + O\left(\frac{h^p}{\delta_T}\right) \xrightarrow{p} 0. \quad (18)$$

This delivers the desired result in (15).

By a similar argument, using Lemma A with $a(y_{j+1}, \theta) = u(y_{j+1}, \theta) u(y_{j+1}, \theta)'$, we obtain (16). Also, by the law of large numbers and $\sup_{1 \leq t \leq T} |\mathbb{I}_{tT} - 1| \xrightarrow{p} 0$ (by Kitamura, Tripathi and Ahn (2004, Lemma D.3)), we obtain (17). This completes the proof of part (a) in Theorem 1.

Proof of Part (b). By expanding the first-order condition $\partial Q_T(\hat{\theta})/\partial\theta = 0$ around θ_0 ,

$$\frac{\partial Q_T(\theta_0)}{\partial\theta} + \frac{\partial^2 Q_T(\bar{\theta})}{\partial\theta\partial\theta'} (\hat{\theta} - \theta_0) = 0,$$

where $\bar{\theta}$ is a point on the line joining $\hat{\theta}$ and θ_0 . It is sufficient for the conclusion to show that

$$\frac{\sqrt{T}}{2} \frac{\partial Q_T(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \mathcal{I}(\theta_0)), \quad (19)$$

$$\frac{1}{2} \frac{\partial^2 Q_T(\bar{\theta})}{\partial\theta\partial\theta'} \xrightarrow{p} \mathcal{I}(\theta_0). \quad (20)$$

To show (19), observe that (see, Donald and Newey, 2000)

$$\begin{aligned} \frac{\sqrt{T}}{2} \frac{\partial Q_T(\theta_0)}{\partial\theta} &= \frac{1}{\sqrt{T}} \sum_{t=p}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta_0)}{\partial\theta'} \right]' V_T(x_t, \theta_0)^{-1} u_T(x_t, \theta_0) \\ &\quad + (A_{2,1}(\theta_0), \dots, A_{2,k}(\theta_0))' \\ &= A_1(\theta_0) + A_2(\theta_0), \end{aligned}$$

w.p.a.1, where $A_1(\theta_0)$ and $A_2(\theta_0)$ are implicitly defined and

$$\begin{aligned} A_{2,l}(\theta_0) &= \frac{1}{2\sqrt{T}} \sum_{t=p}^{T-1} \mathbb{I}_{tT} u_T(x_t, \theta_0)' \frac{\partial V_T(x_t, \theta_0)^{-1}}{\partial\theta_l} u_T(x_t, \theta_0), \\ \frac{\partial V_T(x_t, \theta_0)^{-1}}{\partial\theta_l} &= -V_T(x_t, \theta_0)^{-1} S_T(x_t, \theta_0) V_T(x_t, \theta_0)^{-1}, \\ S_T(x_t, \theta_0) &= \sum_{j=p}^{T-1} w_{tj} u(y_{j+1}, \theta_0) \left(\frac{\partial u(y_{j+1}, \theta_0)}{\partial\theta_l} \right)' + \sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta_0)}{\partial\theta_l} u(y_{j+1}, \theta_0)' \end{aligned}$$

for $l = 1, \dots, k$.

We first consider $A_2(\theta_0)$. Using a similar argument as in the derivation of (18) with $E[u(y_{t+1}, \theta_0) | x_t = x] = 0$ (Assumption A2 (ii)),

$$\sup_t |u_T(x_t, \theta_0)| = O_p \left(\sqrt{\frac{\log T}{\delta_T^2 T h^p}} \right) = o_p(T^{-1/4}), \quad (21)$$

where the second equality follows from Assumption A3 (ii).

From (16), we have $\sup_t |V_T(x_t, \theta_0) - V(x_t, \theta_0)| \xrightarrow{p} 0$. Applying a similar argument in deriving (18) with Lemma A for $a(y_{j+1}, \theta_0) = \frac{\partial u(y_{j+1}, \theta_0)}{\partial\theta_l} u(y_{j+1}, \theta_0)$ yields

$$\sup_t \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta_0)}{\partial\theta_l} u(y_{j+1}, \theta_0)' - E \left[\frac{\partial u(y_{t+1}, \theta_0)}{\partial\theta_l} u(y_{t+1}, \theta_0)' \middle| x_t \right] \right| \xrightarrow{p} 0$$

for $l = 1, \dots, k$. Combining these results and $\sup_{1 \leq t \leq T} |\mathbb{I}_{tT}| \leq 1$,

$$|A_{2,l}(\theta_0)| \leq \left(\sup_t |u_T(x_t, \theta_0)| \right)^2 \sup_t \left| \frac{\partial V_T(x_t, \theta_0)^{-1}}{\partial \theta_l} \right| \frac{1}{2\sqrt{T}} \sum_{t=p}^{T-1} \mathbb{I}_{tT} \xrightarrow{p} 0$$

for $l = 1, \dots, k$, i.e., $|A_2(\theta_0)| \xrightarrow{p} 0$.

Now consider $A_1(\theta_0)$. Using again the same argument for deriving (18) with Lemma A for $a(y_{j+j}, \theta_0) = u(y_{j+1}, \theta_0) u(y_{j+1}, \theta_0)'$ and $\frac{\partial u(y_{j+1}, \theta_0)}{\partial \theta'}$, we obtain

$$\begin{aligned} \sup_t |V_T(x_t, \theta_0) - V(x_t, \theta_0)| &= O_p \left(\sqrt{\frac{\log T}{\delta_T^2 T h^p}} \right) + O \left(\frac{h^\omega}{\delta_T} \right) = o_p \left(T^{-1/4} \right), \\ \sup_t \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta_0)}{\partial \theta'} - D(x_t, \theta_0) \right| &= O_p \left(\sqrt{\frac{\log T}{\delta_T^2 T h^p}} \right) + O \left(\frac{h^\omega}{\delta_T} \right) = o_p \left(T^{-1/4} \right). \end{aligned} \quad (22)$$

Thus, from (21), we have

$$\begin{aligned} A_1(\theta_0) &= \frac{1}{\sqrt{T}} \sum_{t=p}^{T-1} \mathbb{I}_{tT} D(x_t, \theta_0)' V(x_t, \theta_0)^{-1} u_T(x_t, \theta_0) + o_p(1) \\ &= \frac{1}{T^{3/2}} \sum_{t=p}^{T-1} \mathbb{I}_{tT} \sum_{j=p}^{T-1} D(x_t, \theta_0)' V(x_t, \theta_0)^{-1} \frac{\mathbb{K} \left(\frac{x_j - x_t}{h} \right) u(y_{j+1}, \theta_0)}{E \left[\mathbb{K} \left(\frac{x_i - x_t}{h} \right) \mid x_t \right]} + o_p(1), \end{aligned}$$

where the second equality follows from Lemma A with $a(y_{j+1}, \theta) = 1$. Therefore, by applying a similar argument as Kitamura, Tripathi and Ahn (2004, pp. 1696-1698) and using the central limit theorem of U -statistics for absolute regular processes (Fan and Li, 1999), we can show $A_1(\theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0))$, which implies (19).

For (20), note that $\frac{1}{2} \frac{\partial^2 Q_T(\bar{\theta})}{\partial \theta \partial \theta'} = \frac{1}{\sqrt{T}} \frac{\partial A_1(\bar{\theta})}{\partial \theta'} + \frac{1}{\sqrt{T}} \frac{\partial A_2(\bar{\theta})}{\partial \theta'}$ and

$$\begin{aligned} \frac{1}{\sqrt{T}} \frac{\partial A_1(\theta)}{\partial \theta_l} &= \frac{1}{T} \sum_{t=p}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=p}^{T-1} w_{tj} \frac{\partial^2 u(y_{j+1}, \theta)}{\partial \theta_l \partial \theta'} \right]' V_T(x_t, \theta)^{-1} u_T(x_t, \theta) \\ &\quad + \frac{1}{T} \sum_{t=p}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta)}{\partial \theta'} \right]' \frac{\partial V_T(x_t, \theta)^{-1}}{\partial \theta_l} u_T(x_t, \theta) \\ &\quad + \frac{1}{T} \sum_{t=p}^{T-1} \mathbb{I}_{tT} \left[\sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta)}{\partial \theta'} \right]' V_T(x_t, \theta)^{-1} \left[\sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta)}{\partial \theta_l} \right] \\ &= A_{11}(\theta) + A_{12}(\theta) + A_{13}(\theta), \end{aligned}$$

for $l = 1, \dots, k$, where $A_{11}(\theta)$, $A_{12}(\theta)$, and $A_{13}(\theta)$ are implicitly defined. Consider $A_{11}(\bar{\theta})$. From

Lemma A with $a(y_{j+1}, \theta) = \frac{\partial^2 u(y_{j+1}, \theta)}{\partial \theta_l \partial \theta'}$ and $\bar{\theta} \xrightarrow{p} \theta_0$ (by the consistency of $\hat{\theta}$), we have

$$\sup_t \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial^2 u(y_{j+1}, \bar{\theta})}{\partial \theta_l \partial \theta'} \right| \leq \sup_t \sup_{\theta \in \mathcal{N}} \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial^2 u(y_{j+1}, \theta)}{\partial \theta_l \partial \theta'} \right| = O_p(1). \quad (23)$$

From (16), we obtain $\sup_t |V_T(x_t, \bar{\theta})^{-1}| = O_p(1)$. Also, an expansion around $\bar{\theta} = \theta_0$ yields

$$\begin{aligned} \sup_t |u_T(x_t, \bar{\theta})| &\leq \sup_t |u_T(x_t, \theta_0)| + \sup_t \left| \frac{\partial u_T(x_t, \tilde{\theta})}{\partial \theta'} \right| |\bar{\theta} - \theta_0| \\ &= o_p(T^{-1/4}) + o_p(1), \end{aligned}$$

where $\tilde{\theta}$ is a point on the line joining $\bar{\theta}$ and θ_0 , and the equality follows from (21), $\bar{\theta} \xrightarrow{p} \theta_0$, and $\sup_t \left| \frac{\partial u_T(x_t, \tilde{\theta})}{\partial \theta'} \right| = O_p(1)$ (by similar arguments as in the derivation of (22)). Combining these results with $\sup_{1 \leq t \leq T} |\mathbb{I}_{tT}| \leq 1$, we obtain $|A_{11}(\bar{\theta})| \xrightarrow{p} 0$. Similarly, we can show that $A_{12}(\bar{\theta}) \xrightarrow{p} 0$.

Finally consider $A_{13}(\bar{\theta})$. By an expansion around $\bar{\theta} = \theta_0$,

$$\begin{aligned} &\sup_t \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \bar{\theta})}{\partial \theta_l} - E \left[\frac{\partial u(y_{j+1}, \theta_0)}{\partial \theta_l} \middle| x_t \right] \right| \\ &\leq \sup_t \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial u(y_{j+1}, \theta_0)}{\partial \theta_l} - E \left[\frac{\partial u(y_{j+1}, \theta_0)}{\partial \theta_l} \middle| x_t \right] \right| + \sup_t \left| \sum_{j=p}^{T-1} w_{tj} \frac{\partial^2 u(y_{j+1}, \tilde{\theta})}{\partial \theta_l \partial \theta'} \right| |\bar{\theta} - \theta_0| \\ &\xrightarrow{p} 0, \end{aligned}$$

where $\tilde{\theta}$ is a point on the line joining $\bar{\theta}$ and θ_0 , and the convergence follows from (22), (23), and $\bar{\theta} \xrightarrow{p} \theta_0$. Similarly, we have $\sup_t |V_T(x_t, \bar{\theta})^{-1} - V(x_t, \theta_0)| \xrightarrow{p} 0$. Combining these results, we have $A_{13}(\bar{\theta}) \xrightarrow{p} E \left[D(x_t, \theta_0)' V(x_t, \theta_0) E \left[\frac{\partial u(y_{t+1}, \theta)}{\partial \theta_l} \middle| x_t \right] \right]$ and thus $\frac{1}{\sqrt{T}} \frac{\partial A_1(\bar{\theta})}{\partial \theta'} \xrightarrow{p} \mathcal{I}(\theta_0)$. Also, a similar but lengthy argument yields $\frac{1}{\sqrt{T}} \frac{\partial A_2(\bar{\theta})}{\partial \theta} \xrightarrow{p} 0$. Therefore, we obtain (20), which combined with the proof of (19) completes the proof of part (b) in Theorem 1.

A.2 Proof of Theorem 2

We first derive asymptotic expansions for \hat{r}_t and \hat{u}_t . An expansion of $\hat{r}_t = \tilde{r}_t / \tilde{f}_t$ around $\tilde{f}_t = f_t$ yields

$$\hat{r}_t = \left[\frac{1}{f_t} - \frac{V_{f,t} + B_{f,t}}{f_t^2} + R_{f,t} \right] [V_{r,t} + B_{r,t} + r_t f_t], \quad (24)$$

for $t = 1, \dots, T-1$, where $R_{f,t} = (V_{f,t} + B_{f,t})^2 / (4\dot{f}_t^3)$ and \dot{f}_t is a point on the line joining \tilde{f}_t and f_t . Similarly, an expansion of $\hat{u}_{t+1} = \tilde{u}_{t+1} / \tilde{f}_t$ around $\tilde{f}_t = f_t$ is obtained as

$$\hat{u}_t = \left[\frac{1}{f_t} - \frac{V_{f,t} + B_{f,t}}{f_t^2} + R_{f,t} \right] \tilde{u}_{t+1}, \quad (25)$$

for $t = 1, \dots, T-1$. From (24), (25), and an expansion of $\hat{\theta}_{LGMM1} - \theta_0 = \frac{T^{-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \hat{r}_t \hat{u}_{t+1}}{T^{-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \hat{r}_t^2}$ around $T^{-1} \sum_{t=1}^T \mathbb{I}_{tT} \hat{r}_t^2 = \sigma_r^2$, we have

$$\hat{\theta}_{LGMM1} - \theta_0 = M_1 - M_2 + M_3,$$

where

$$\begin{aligned} M_1 &= \frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left[\frac{1}{f_t} - \frac{V_{f,t} + B_{f,t}}{f_t^2} + R_{f,t} \right]^2 [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1}, \\ M_2 &= \frac{T^{-1} \sum_{t=1}^T \mathbb{I}_{tT} \hat{r}_t^2 - \sigma_r^2}{\sigma_r^4} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left[\frac{1}{f_t} - \frac{V_{f,t} + B_{f,t}}{f_t^2} + R_{f,t} \right]^2 [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1}, \\ M_3 &= R \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left[\frac{1}{f_t} - \frac{V_{f,t} + B_{f,t}}{f_t^2} + R_{f,t} \right]^2 [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1}, \end{aligned}$$

and R is the remainder term satisfying $R = O_p \left(\left(T^{-1} \sum_{t=1}^T \mathbb{I}_{tT} \hat{r}_t^2 - \sigma_r^2 \right)^2 \right)$. For M_1 , a lengthy calculation, combined with Assumption A4 (see Appendix A.2.1), implies

$$\begin{aligned} M_1 &= \frac{1}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{(V_{r,t} + B_{r,t} + r_t f_t) \tilde{u}_{t+1}}{f_t} - \frac{2}{\sigma_r^2} \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{(V_{f,t} + B_{f,t}) r_t \tilde{u}_{t+1}}{f_t^2} \\ &\quad + O_p \left(a_T (a_T + b_T)^2 \right). \end{aligned} \quad (26)$$

From (24), an expansion for $T^{-1} \sum_{t=1}^T \mathbb{I}_{tT} \hat{r}_t^2 - \sigma_r^2$ is obtained as

$$\begin{aligned} T^{-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \hat{r}_t^2 - \sigma_r^2 &= (\hat{s}_r^2 - \sigma_r^2) + \frac{2}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t (V_{r,t} + B_{r,t})}{f_t} - \frac{2}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t^2 (V_{f,t} + B_{f,t})}{f_t} \\ &\quad + O_p \left((a_T + b_T)^2 \right). \end{aligned} \quad (27)$$

From (27) and a similar argument as in the derivation of (26), an expansion for M_2 is given by

$$\begin{aligned} M_2 &= \frac{1}{\sigma_r^4} \left((\hat{s}_r^2 - \sigma_r^2) + \frac{2}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t (V_{r,t} + B_{r,t})}{f_t} - \frac{2}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t^2 (V_{f,t} + B_{f,t})}{f_t} \right) \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{r_t \tilde{u}_{t+1}}{f_t} \\ &\quad + O_p \left(a_T (a_T + b_T)^2 \right). \end{aligned}$$

Also, under our assumptions, M_3 satisfies

$$M_3 = O_p \left(a_T (a_T + b_T)^2 \right).$$

Combining these results, we obtain the conclusion.

A.2.1 Derivation of (26)

A direct calculation yields

$$\begin{aligned}
M_1 &= \frac{1}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{1}{f_t^2} [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1} \\
&\quad + \frac{1}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{(V_{f,t} + B_{f,t})^2}{f_t^4} [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1} \\
&\quad + \frac{1}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} R_{f,t}^2 [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1} \\
&\quad - \frac{2}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{V_{f,t} + B_{f,t}}{f_t^3} [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1} \\
&\quad + \frac{2}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{R_{f,t}}{f_t} [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1} \\
&\quad - \frac{2}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{(V_{f,t} + B_{f,t}) R_{f,t}}{f_t^2} [V_{r,t} + B_{r,t} + r_t f_t] \tilde{u}_{t+1} \\
&= M_{11} + M_{12} + M_{13} - 2M_{14} + 2M_{15} - 2M_{16}.
\end{aligned}$$

M_{11} is the main term. For M_{12} , our assumptions and the law of large numbers ($\frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{1}{f_t^4} \right| = O_p(1)$) and $\frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{r_t}{f_t^3} \right| = O_p(1)$) imply

$$\begin{aligned}
|M_{12}| &\leq \frac{1}{\sigma_r^2} \left(\sup_t |V_{f,t} + B_{f,t}| \right)^2 \sup_t |\tilde{u}_{t+1}| \\
&\quad \times \left\{ \sup_t |V_{r,t} + B_{r,t}| \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{1}{f_t^4} \right| + \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{r_t}{f_t^3} \right| \right\} \\
&= O_p \left(a_T (a_T + b_T)^2 \right).
\end{aligned}$$

For M_{13} , we have

$$\begin{aligned}
|M_{13}| &\leq \frac{1}{\sigma_r^2} \sup_t |\tilde{u}_{t+1}| \left(\sup_t |V_{f,t} + B_{f,t}| \right)^2 \left\{ \sup_t |V_{r,t} + B_{r,t}| \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{1}{4f_t^3} \right| + \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{r_t f_t}{4f_t^3} \right| \right\} \\
&= O_p \left(a_T (a_T + b_T)^2 \right),
\end{aligned}$$

where the equality follows from Assumption A4, $\sup_t \left| \frac{\hat{f}_t}{f_t} - 1 \right| \xrightarrow{p} 0$, and the law of large numbers.

For M_{14} , we have

$$M_{14} = \frac{1}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{V_{f,t} + B_{f,t}}{f_t^2} r_t \tilde{u}_{t+1} + O_p \left(a_T (a_T + b_T)^2 \right),$$

where the first equality follows from

$$\begin{aligned} & \left| \frac{1}{\sigma_r^2} \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \frac{V_{f,t} + B_{f,t}}{f_t^3} [V_{r,t} + B_{r,t}] \tilde{u}_{t+1} \right| \\ & \leq \frac{1}{\sigma_r^2} \sup_t |V_{f,t} + B_{f,t}| \sup_t |V_{r,t} + B_{r,t}| \sup_t |\tilde{u}_{t+1}| \times \left\{ \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{1}{f_t^3} \right| \right\} = O_p \left(a_T (a_T + b_T)^2 \right). \end{aligned}$$

For M_{15} ,

$$\begin{aligned} |M_{15}| & \leq \frac{1}{\sigma_r^2} \sup_t |\tilde{u}_{t+1}| \left(\sup_t |V_{f,t} + B_{f,t}| \right)^2 \left\{ \sup_t |V_{r,t} + B_{r,t}| \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{1}{4f_t \dot{f}_t^3} \right| + \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{r_t}{4\dot{f}_t^3} \right| \right\} \\ & = O_p \left(a_T (a_T + b_T)^2 \right). \end{aligned}$$

Similarly, for M_{16} ,

$$\begin{aligned} |M_{16}| & \leq \frac{1}{\sigma_r^2} \sup_t |\tilde{u}_{t+1}| \left(\sup_t |V_{f,t} + B_{f,t}| \right)^3 \left\{ \sup_t |V_{r,t} + B_{r,t}| \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{1}{4f_t^2 \dot{f}_t^3} \right| + \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}_{tT} \left| \frac{r_t}{4f_t \dot{f}_t^3} \right| \right\} \\ & = O_p \left(a_T (a_T + b_T)^3 \right). \end{aligned}$$

Combining these results, we obtain (26).

A.3 Proof of Theorem 3

Note that

$$\begin{aligned} E[A_{LGMM}^*] & = \frac{1}{\sigma_r^2} \frac{1}{T^2} \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} E \left[\frac{r_t}{f_t} \frac{1}{h} \mathbb{K} \left(\frac{r_t - r_j}{h} \right) u_{j+1} \right] \\ & = \frac{1}{\sigma_r^2} \frac{1}{T^2} \sum_{t=1}^{T-1} \sum_{j=1}^{t-1} E \left[\frac{r_t}{f_t} \frac{1}{h} \mathbb{K} \left(\frac{r_t - r_j}{h} \right) u_{j+1} \right] \\ & = \frac{1}{\sigma_r^2} \frac{1}{T^2} \sum_{t=1}^{T-1} \sum_{j=1}^{t-1} A_{t,j}, \end{aligned}$$

where the second equality follows from the law of iterated expectation. Let f_u be the density function of u_t . For $A_{t,t-1}$, we have

$$\begin{aligned}
A_{t,t-1} &= E \left[\frac{r_t}{f(r_t)} \frac{1}{h} \mathbb{K} \left(\frac{r_t - r_{t-1}}{h} \right) u_t \right] \\
&= E \left[\frac{\theta_0 r_{t-1} + u_t}{f(\theta_0 r_{t-1} + u_t)} \frac{1}{h} \mathbb{K} \left(\frac{(\theta_0 - 1) r_{t-1} + u_t}{h} \right) u_t \right] \\
&= E \left[\int \frac{ah + r_{t-1}}{f(ah + r_{t-1})} \mathbb{K}(a) (ah + (1 - \theta_0) r_{t-1}) f_u(ah + (1 - \theta_0) r_{t-1}) da \right] \\
&= E \left[\frac{(1 - \theta_0) r_{t-1}^2 f_u((1 - \theta_0) r_{t-1})}{f(r_{t-1})} \right] + O(h^\omega) \\
&= \int \frac{b^2}{(1 - \theta_0)^2} f_u(b) db + O(h^\omega) \\
&= \frac{1}{(1 - \theta_0)^2} + O(h^\omega),
\end{aligned}$$

where the third equality follows from the law of iterated expectations and a change of variables ($a = \frac{(\theta_0 - 1)r_{t-1} + u_t}{h}$), the fourth equality follows from an expansion around $a = 0$ and the property of the ω -th order kernel \mathbb{K} , and the fifth equality follows from a change of variables ($b = (1 - \theta_0) r_{t-1}$). Similarly, for $A_{t,t-l}$ ($l = 2, \dots, t - 1$),

$$\begin{aligned}
A_{t,t-l} &= E \left[\frac{r_t}{f(r_t)} \frac{1}{h} \mathbb{K} \left(\frac{r_t - r_{t-l}}{h} \right) u_{t-l+1} \right] \\
&= E \left[\frac{\frac{\theta_0^l r_{t-l} + u_t + \theta_0 u_{t-1} + \dots + \theta_0^{l-1} u_{t-l+1}}{f(\theta_0^l r_{t-l} + u_t + \theta_0 u_{t-1} + \dots + \theta_0^{l-1} u_{t-l+1})}}{\times \frac{1}{h} \mathbb{K} \left(\frac{(\theta_0^l - 1) r_{t-l} + u_t + \theta_0 u_{t-1} + \dots + \theta_0^{l-1} u_{t-l+1}}{h} \right)} u_{t-l+1} \right] \\
&= E \left[\int \frac{ah + r_{t-l}}{f(ah + r_{t-l})} \mathbb{K}(a) u_{t-l+1} f_u \left(ah + \left(1 - \theta_0^l \right) r_{t-l} - \theta_0 u_{t-1} - \dots - \theta_0^{l-1} u_{t-l+1} \right) da \right] \\
&= E \left[\frac{r_{t-l}}{f(r_{t-l})} u_{t-l+1} f_u \left(\left(1 - \theta_0^l \right) r_{t-l} - \theta_0 u_{t-1} - \dots - \theta_0^{l-1} u_{t-l+1} \right) \right] + O(h^\omega) \\
&= E \left[\left(\int r f_u \left(\left(1 - \theta_0^l \right) r - \theta_0 u_{t-1} - \dots - \theta_0^{l-1} u_{t-l+1} \right) dr \right) u_{t-l+1} \right] + O(h^\omega) \\
&= E \left[\left(\int \frac{b + \theta_0 u_{t-1} + \dots + \theta_0^{l-1} u_{t-l+1}}{(1 - \theta_0^l)^2} f_u(b) db \right) u_{t-l+1} \right] + O(h^\omega) \\
&= \frac{1}{(1 - \theta_0^l)^2} E \left[\left(\theta_0 u_{t-1} + \dots + \theta_0^{l-1} u_{t-l+1} \right) u_{t-l+1} \right] + O(h^\omega) \\
&= \frac{\theta_0^{l-1}}{(1 - \theta_0^l)^2} + O(h^\omega),
\end{aligned}$$

where the second equality follows from model (7), the third equality follows from the law of iterated expectations (with respect to u_t given the information at time $t - 1$) and a change of variables

$(a = \frac{(\theta_0^l - 1)r_{t-l} + u_t + \theta_0 u_{t-1} + \dots + \theta_0^{l-1} u_{t-l+1}}{h})$, the fourth equality follows from an expansion around $a = 0$ and the property of the ω -th order kernel \mathbb{K} , the fifth equality follows from the law of iterated expectations (with respect to r_{t-l} given $(u_{t-1}, \dots, u_{t-l+1})$) and independence between r_{t-l} and $(u_{t-1}, \dots, u_{t-l+1})$, and the sixth equality follows from a change of variable ($b = (1 - \theta_0^l)r - \theta_0 u_{t-1} - \dots - \theta_0^{l-1} u_{t-l+1}$). Combining these results with $\sigma_r^{-2} = 1 - \theta_0^2$, we obtain the conclusion.

References

- [1] Aït-Sahalia, Y., 1996. Testing continuous time models of the spot interest rate. *Review of Financial Studies* 9, 385-426.
- [2] Anatolyev, S., 2005. GMM, GEL, serial correlation, and asymptotic bias. *Econometrica* 73, 983-1002.
- [3] Antoine, B., Bonnal, H., Renault, E., 2007. On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics* 138, 461-487.
- [4] Ball, C.A., Torous, W.N., 1996. Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance* 3, 215-238.
- [5] Beach, C.M., MacKinnon, J.G., 1978. A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica* 46, 51-58.
- [6] Borkovec, M., 2001. Asymptotic behaviour of the sample autocovariance and autocorrelation function of the AR(1) process with ARCH(1) errors. *Bernoulli* 7, 847-872.
- [7] Borkovec, M., Klüppelberg, C., 2001. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Annals of Applied Probability* 11, 1220-1241.
- [8] Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.
- [9] Carrasco, M., Chernov, M., Florens, J.-P., Ghysels, E., 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics* 140, 529-573.
- [10] Carrasco, M., Florens, J.P., 2000. Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16, 797-834.
- [11] Carrasco, M., Florens, J.P., 2004. On the asymptotic efficiency of GMM. Manuscript, University of Rochester.
- [12] Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 395-334.
- [13] Chapman, D., Pearson, N., 2000. Is the short rate drift actually nonlinear? *Journal of Finance* 55, 355-388.
- [14] Chen, S.X., Härdle, W., Li, M., 2003. An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society B* 65, 663-678.
- [15] Cox, J.C., Ingersoll, J.E.Jr., Ross, S.A., 1985. A theory of the term structure of interest rates. *Econometrica* 53, 385-408.
- [16] Davidson, J., 2000. *Econometric Theory*. Oxford: Blackwell Publishers.

- [17] Dominguez, M., Lobato, I., 2004. Consistent estimation of models defined by conditional moment restrictions. *Econometrica* 72, 1601-1615.
- [18] Donald, S.G., Imbens, G.W., Newey, W.K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117, 55-93.
- [19] Donald, S.G., Newey, W.K., 2000. A jackknife interpretation of the continuous updating estimator. *Economics Letters* 67, 239-243.
- [20] Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245-271.
- [21] Gagliardini, P., Gourieroux, C., Renault, E., 2007. Efficient derivative pricing by extended method of moments. Manuscript, University of Lugano.
- [22] Giersbergen, N.P.A.v., 2005. On the effect of deterministic terms on the bias in stable AR models. *Economics Letters* 89, 75-82.
- [23] Guégan, D., Diebolt, J., 1994. Probabilistic properties of the β -ARCH model. *Statistica Sinica* 4, 71-87.
- [24] Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14, 262-280.
- [25] Kendall, M.G., 1954. Note on bias in the estimation of autocorrelation. *Biometrika* 41, 403-404.
- [26] Kitamura, Y., Tripathi, G., Ahn, H., 2004. Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 1667-1714.
- [27] Kristensen, D., 2009. Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory* 25, 1433-1445.
- [28] Kuersteiner, G., 2001. Optimal instrumental variables estimation from ARMA models. *Journal of Econometrics* 104, 359-405.
- [29] Kuersteiner, G., 2002. Efficient estimation for autoregressive models with conditional heteroskedasticity. *Econometric Theory* 18, 547-583.
- [30] Kuersteiner, G., 2006. Moment selection and bias reduction for GMM in conditionally heteroskedastic models. In: D. Corbea, S. Durlauf and B.E. Hansen, eds, *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips*, Cambridge: Cambridge University Press.
- [31] Lavergne, P., Patilea, V., 2008. Efficient smooth GMM and dimension reduction. Manuscript, Simon Fraser University.
- [32] Lewbel, A., 2007. A local generalized method of moments estimator. *Economics Letters* 94, 124-128.
- [33] Ling, S., 2004. Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society B* 66, 63-78.

- [34] Maercker, G., 1995. Efficient estimation in AR models with ARCH errors. Technical report, Institut für Mathematische Stochastik, TU Braunschweig.
- [35] Marriott, F.H.C., Pope, J.A., 1954. Bias in the estimation of autocorrelations. *Biometrika* 41, 390-402.
- [36] Masry, E., Tjøstheim, D., 1995. Nonparametric estimation and identification of nonlinear arch time series. *Econometric Theory* 11, 258-289.
- [37] Meyn, S.P., Tweedie, R.L., 1993. *Markov Chains and Stochastic Stability*. Berlin: Springer-Verlag.
- [38] Newey, W.K., 1990. Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58, 809-837.
- [39] Newey, W.K., 1993. Efficient estimation of models with conditional moment restrictions. In: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol. 11 (Elsevier, Amsterdam) 419-454.
- [40] Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol. 4 (Elsevier, Amsterdam) 2111-2245.
- [41] Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219-255.
- [42] Phillips, P.C.B., Yu, J., 2005. Jackknifing bond option prices. *Review of Financial Studies* 18, 707-742.
- [43] Smith, R.J., 2007. Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics* 138, 430-460.
- [44] Su, L., White, H., 2003. Testing conditional independence via empirical likelihood. UCSD working paper 2003-14.
- [45] West, K.D., Wong, K., Anatolyev, S., 2009. Instrumental variables estimation of heteroskedastic linear models using all lags of instruments. *Econometric Reviews* 28, 441-467.

TABLE 1. Finite-sample properties of OLS, GLS, optimal instrumental variables (OIV), Carrasco-Florens (CF) and LGMM estimators of the slope parameter in an AR(1) model ($T = 100$).

	$\theta_0=0.4$				$\theta_0=0.7$				$\theta_0=0.95$			
	mean	med	sd	rmse	mean	med	sd	rmse	mean	med	sd	rmse
$\beta_0=0$												
OLS	-0.022	-0.019	0.093	0.095	-0.032	-0.025	0.076	0.083	-0.044	-0.034	0.050	0.066
GLS	-0.022	-0.019	0.093	0.095	-0.032	-0.025	0.076	0.083	-0.044	-0.034	0.050	0.066
OIV	-0.022	-0.018	0.094	0.096	-0.032	-0.025	0.078	0.084	-0.048	-0.038	0.051	0.070
CF	-0.020	-0.016	0.094	0.096	-0.027	-0.020	0.078	0.082	-0.039	-0.029	0.050	0.063
LGMM	-0.018	-0.015	0.101	0.103	-0.021	-0.014	0.083	0.086	-0.029	-0.019	0.051	0.058
$\beta_0=0.5$												
OLS	-0.029	-0.024	0.133	0.136	-0.041	-0.030	0.101	0.109	-0.047	-0.036	0.057	0.074
GLS	-0.023	-0.019	0.109	0.112	-0.031	-0.023	0.081	0.087	-0.037	-0.027	0.047	0.060
OIV	-0.037	-0.030	0.131	0.136	-0.048	-0.037	0.100	0.110	-0.053	-0.042	0.060	0.080
CF	-0.017	-0.012	0.124	0.125	-0.027	-0.018	0.095	0.099	-0.039	-0.027	0.054	0.066
LGMM	-0.006	-0.002	0.126	0.127	-0.009	0.002	0.098	0.099	-0.025	-0.014	0.052	0.058
$\beta_0=0.9$												
OLS	-0.040	-0.029	0.181	0.185	-0.056	-0.036	0.139	0.150	-0.056	-0.038	0.076	0.094
GLS	-0.021	-0.017	0.113	0.115	-0.025	-0.018	0.078	0.082	-0.025	-0.017	0.040	0.047
OIV	-0.057	-0.044	0.177	0.186	-0.074	-0.053	0.140	0.158	-0.060	-0.045	0.081	0.101
CF	-0.018	-0.008	0.165	0.166	-0.032	-0.017	0.125	0.129	-0.042	-0.027	0.065	0.077
LGMM	0.005	0.008	0.158	0.158	0.004	0.015	0.119	0.119	-0.019	-0.007	0.055	0.058

Notes: The statistics in the table are computed from 10,000 samples generated from the AR(1) process $r_{t+1} = \theta_0 r_t + u_{t+1}$ with $T = 100$, $\theta_0 = 0.4, 0.7, 0.95$, and $u_{t+1} = \sigma_{t+1} \varepsilon_{t+1}$, where $\varepsilon_{t+1} \sim iidN(0, 1)$ and $\sigma_{t+1} = \sqrt{(1 - \beta_0) + \beta_0 u_t^2}$ for $\beta_0 = 0, 0.5, 0.9$. The estimated model includes an intercept. “mean”, “med”, “sd” and “rmse” denote the mean bias, median bias, standard deviation and root mean square error, respectively.

TABLE 2. Finite-sample properties of OLS, GLS, optimal instrumental variables (OIV), Carrasco-Florens (CF) and LGMM estimators of the slope parameter in an AR(1) model ($T = 500$).

	$\theta_0=0.4$				$\theta_0=0.7$				$\theta_0=0.95$			
	mean	med	sd	rmse	mean	med	sd	rmse	mean	med	sd	rmse
$\beta_0=0$												
OLS	-0.005	-0.004	0.041	0.041	-0.006	-0.005	0.032	0.033	-0.008	-0.006	0.016	0.018
GLS	-0.005	-0.004	0.041	0.041	-0.006	-0.005	0.032	0.033	-0.008	-0.006	0.016	0.018
OIV	-0.005	-0.005	0.041	0.041	-0.006	-0.006	0.032	0.033	-0.008	-0.006	0.016	0.018
CF	-0.004	-0.004	0.042	0.042	-0.005	-0.004	0.033	0.033	-0.006	-0.004	0.016	0.017
LGMM	-0.004	-0.004	0.042	0.042	-0.004	-0.003	0.033	0.034	-0.004	-0.002	0.016	0.017
$\beta_0=0.5$												
OLS	-0.008	-0.007	0.068	0.068	-0.010	-0.008	0.048	0.049	-0.009	-0.007	0.019	0.021
GLS	-0.005	-0.004	0.049	0.049	-0.006	-0.005	0.034	0.035	-0.007	-0.005	0.014	0.016
OIV	-0.010	-0.010	0.066	0.067	-0.011	-0.009	0.044	0.046	-0.009	-0.007	0.018	0.020
CF	-0.003	-0.002	0.055	0.055	-0.004	-0.003	0.041	0.041	-0.005	-0.003	0.018	0.018
LGMM	-0.001	-0.001	0.053	0.053	-0.001	0.001	0.041	0.041	-0.002	0.001	0.017	0.017
$\beta_0=0.9$												
OLS	-0.019	-0.014	0.130	0.132	-0.025	-0.016	0.095	0.098	-0.016	-0.009	0.035	0.039
GLS	-0.004	-0.004	0.050	0.050	-0.005	-0.003	0.031	0.031	-0.004	-0.003	0.011	0.012
OIV	-0.030	-0.023	0.125	0.129	-0.033	-0.021	0.088	0.094	-0.012	-0.008	0.031	0.033
CF	-0.001	0.001	0.086	0.086	-0.003	0.000	0.063	0.063	-0.004	-0.001	0.023	0.023
LGMM	0.002	0.002	0.064	0.064	0.004	0.005	0.050	0.051	0.002	0.004	0.018	0.018

Notes: The statistics in the table are computed from 10,000 samples generated from the AR(1) process $r_{t+1} = \theta_0 r_t + u_{t+1}$ with $T = 500$, $\theta_0 = 0.4, 0.7, 0.95$, and $u_{t+1} = \sigma_{t+1} \varepsilon_{t+1}$, where $\varepsilon_{t+1} \sim iidN(0, 1)$ and $\sigma_{t+1} = \sqrt{(1 - \beta_0) + \beta_0 u_t^2}$ for $\beta_0 = 0, 0.5, 0.9$. The estimated model includes an intercept. “mean”, “med”, “sd” and “rmse” denote the mean bias, median bias, standard deviation and root mean square error, respectively.

TABLE 3. Bias properties of OLS and LGMM estimators of the slope parameter in a zero-mean AR(1) model with *iid* errors.

	mean bias	sd	rmse	$E[A]$	$E[B]$
<i>T</i> = 50					
OLS	-0.0309	0.0646	0.0716	0.0002	0.0303
LGMM	-0.0128	0.0621	0.0634	0.0108	0.0310
<i>T</i> = 100					
OLS	-0.0167	0.0395	0.0429	0.0000	0.0164
LGMM	-0.0040	0.0383	0.0385	0.0082	0.0168
<i>T</i> = 200					
OLS	-0.0093	0.0260	0.0276	-0.0002	0.0090
LGMM	-0.0010	0.0258	0.0259	0.0055	0.0090
<i>T</i> = 400					
OLS	-0.0048	0.0172	0.0178	0.0000	0.0048
LGMM	0.0000	0.0173	0.0173	0.0034	0.0047
<i>T</i> = 800					
OLS	-0.0026	0.0118	0.0121	-0.0001	0.0024
LGMM	0.0001	0.0119	0.0119	0.0018	0.0024
<i>T</i> = 1600					
OLS	-0.0011	0.0080	0.0081	0.0000	0.0012
LGMM	0.0002	0.0081	0.0081	0.0011	0.0012
<i>T</i> = 3200					
OLS	-0.0006	0.0055	0.0055	0.0000	0.0006
LGMM	0.0001	0.0055	0.0055	0.0005	0.0006

Notes: The statistics in the table are computed from 5,000 samples generated from the AR(1) process $r_{t+1} = \theta_0 r_t + u_{t+1}$ with $\theta_0 = 0.95$ and $u_{t+1} \sim iidN(0, 1)$. “sd” and “rmse” denote the standard deviation and root mean square error, respectively. The last two columns of the table are Monte Carlo averages of the terms $E[A_{OLS}]$ and $E[B_{OLS}]$ for the OLS estimator and $E[A_{LGMM}]$ and $E[B_{LGMM}]$ for the LGMM estimator which are defined in Section 4.1.

TABLE 4. Finite-sample properties of GLS, Carrasco-Florens (CF) and LGMM estimators of the parameters of the CIR model and bond and option prices.

	intercept	κ	μ	σ	bond	option
true	0.0184	0.2146	0.0857	0.0783	83.763	1.931
GLS						
median	0.0488	0.5938	0.0825	0.0852	82.030	0.756
mean	0.0587	0.6931	0.0866	0.0926	81.613	1.127
sd	0.0455	0.4538	0.0279	0.0222	3.262	1.190
rmse	0.0608	0.6595	0.0279	0.0264	3.907	1.436
CF						
median	0.0497	0.6011	0.0826	0.0852	82.000	0.711
mean	0.0586	0.6950	0.0864	0.0926	81.587	1.110
sd	0.0407	0.4281	0.0273	0.0222	3.241	1.191
rmse	0.0572	0.6435	0.0273	0.0264	3.904	1.446
LGMM						
median	0.0328	0.4030	0.0833	0.0852	83.022	1.458
mean	0.0403	0.4805	0.0873	0.0926	82.602	1.652
sd	0.0324	0.3638	0.0297	0.0223	2.921	1.384
rmse	0.0391	0.4506	0.0298	0.0265	3.143	1.412

Notes: The statistics in the table are computed from 5,000 samples generated from the CIR model with $\Delta = 1/52$ and $T = 600$. “sd” and “rmse” denote the standard deviation and root mean square error, respectively. The prices of a three-year zero-coupon discount bond and a one-year European call option on a three-year bond with face value of \$100 and a strike price of \$87 are computed analytically as in Cox, Ingersoll and Ross (1985) with an initial interest rate of 5%. The first row “true” reports the true values of the parameters, bond and option prices.

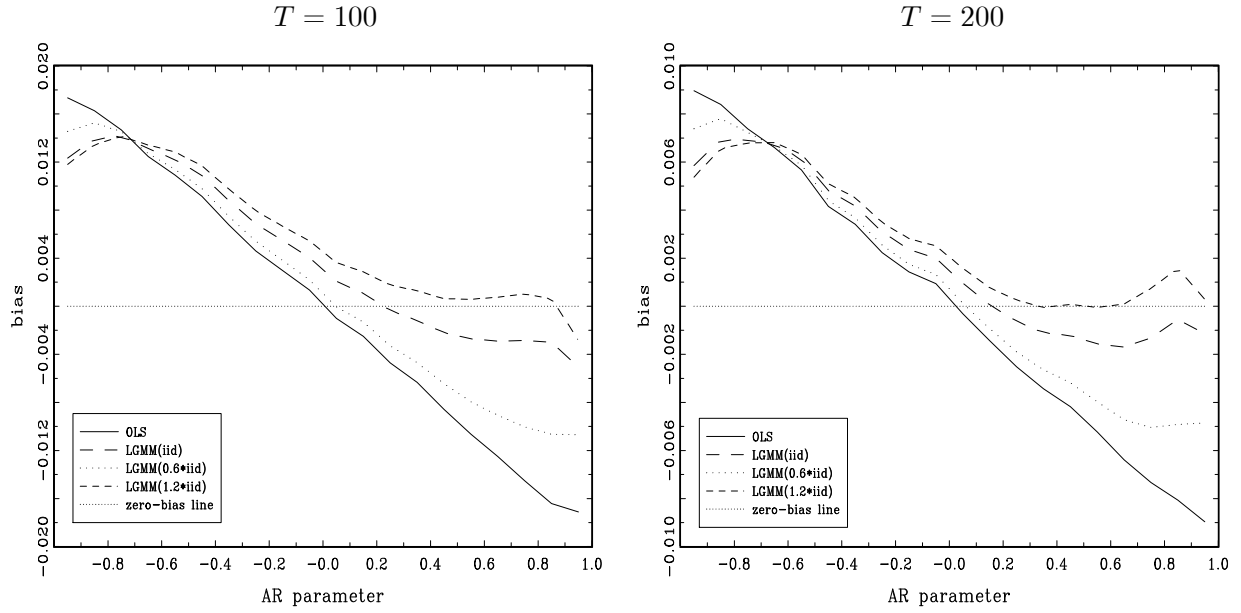


FIGURE 1. Mean bias of OLS and LGMM estimators in a zero-mean AR(1) model with AR parameter $\theta_0 = -0.95, -0.85, \dots, 0.85, 0.95$, and $T = 100$ (left graph) and $T = 200$ (right graph). The LGMM estimators are computed with different bandwidths: (approximate) *iid* bandwidth $T^{-1/5}$, $0.6T^{-1/5}$ and $1.2T^{-1/5}$ with a standardized conditioning variable. The mean bias is computed from 100,000 Monte Carlo replications.