MATHEMATICS OF OPERATIONS RESEARCH
Vol. 37, No. 3, August 2012, pp. 489–500
ISSN 0364-765X (print) | ISSN 1526-5471 (online)

http://dx.doi.org/10.1287/moor.1120.0547

Generalized Look-Ahead Methods for Computing Stationary Densities

R. Anton Braun
Research Department, Federal Reserve Bank of Atlanta, Atlanta, Georgia, r.anton.braun@atl.frb.org

Huiyu Li
Department of Economics, Graduate School, Stanford University, Stanford, California, huiyuli@stanford.edu

John Stachurski
Research School of Economics, Australian National University, Canberra, Australia, john.stachurski@anu.edu.au

The look-ahead estimator is used to compute densities associated with Markov processes via simulation. We study a framework that extends the look-ahead estimator to a broader range of applications. We provide a general asymptotic theory for the estimator, where both L_1 consistency and L_2 asymptotic normality are established. The L_2 asymptotic normality implies $\sqrt{n}$ convergence rates for L_2 deviation.

Key words: densities; numerical methods; simulation
MSC2000 subject classification: Primary: 60J22; secondary: 60J05, 65C05
OR/MS subject classification: Primary: probability/Markov processes; secondary: simulation
History: Received August 1, 2011; revised March 15, 2012. Published online in *Articles in Advance* June 19, 2012.

1. Introduction. Simulation allows researchers to extract probabilities from otherwise intractable models. In some cases the random variables being simulated have distributions that can be represented by densities, and the researcher seeks to construct, via simulation, approximations to these densities. This problem arises frequently in operations research, economics, finance, and statistics. (See, for example, Henderson and Glynn [10], Brandt and Santa-Clara [3], Danielsson [5], or Gelfand and Smith [8].)

In the process of computing densities via Monte Carlo, the two steps are (i) simulate the relevant variables, and (ii) produce density estimates from the simulated observations. Our interest is in the second step. In this step, parametric estimation is problematic, as the parametric classes of these densities are generally unknown or nonexistent. Nonparametric density estimates are more robust, but converge at a slower rate and degrade quickly as the dimension of the state space increases.

An alternative method of constructing densities from simulated observations is to use conditional Monte Carlo. An important example is the look-ahead estimator of Henderson and Glynn [10], which has applications such as computing stationary densities of Markov processes. Similar ideas have appeared in other strands of the literature under different names. For example, Gelfand and Smith [8] contains a related method. Brandt and Santa-Clara [3] and Pedersen [19] independently proposed a method of simulated maximum likelihood based on conditioning ideas. (All of these methods are special cases of Henderson and Glynn's look-ahead estimator.)

In this paper, we study a generalized look-ahead estimator that implements conditional Monte Carlo density estimation with correlated samples. To understand the nature of the generalization, recall that the basic look-ahead estimator can be used to compute distributions of the state variable of a Markov process. The generalization considered here can compute the distributions of random or nonrandom functions of the state variable—in particular, any random variable that can be related to the state variable of the Markov process via a conditional density. For example, if volatility of returns to holding an asset is modeled as Markovian, then the look-ahead estimator can be used to compute the stationary density of volatility. The generalized look-ahead estimator we consider here can be used to compute not only the distribution of volatility, but also that of other variables correlated with volatility, such as returns themselves. (See §§4 and 5 for more discussion of this example.)

In what follows, we provide a relatively complete asymptotic theory of the generalized look-ahead estimator, with a focus on minimal assumptions and global (i.e., norm) deviation between the estimator and the target density. Regarding norm consistency, Henderson and Glynn [10, Theorem 7, Corollary 8] showed that for the geometrically ergodic case, the L_1 deviation between the look-ahead estimator and the target density converges to zero with probability one, provided that the transition density satisfies certain continuity and regularity assumptions. Here we show that the L_1 deviation between the generalized look-ahead estimator (and hence the look-ahead estimator) and the target density converges to zero without any conditions on the transition density. Our ergodicity conditions are also minimal.

Regarding asymptotic normality, Stachurski and Martin [20] showed that under geometric ergodicity, the L_2 deviation between the look-ahead estimator and target density is $O_P(n^{-1/2})$, independent of the dimension of the state space. We extend this result to the generalized look-ahead estimator, proving an L_2 central limit theorem and obtaining the $O_P(n^{-1/2})$ rate as a consequence.

2. Definitions. Let $(\mathbb{Y}, \mathcal{Y}, \mu)$ be a σ-finite measure space, and let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. As usual, a $\mathbb{Y}$-valued random variable is a measurable map Y from $(\Omega, \mathcal{F})$ into $(\mathbb{Y}, \mathcal{Y})$. We use the symbol $\mathcal{L}Y$ to denote the law (i.e., distribution) of Y. We say that Y has density g if $g\colon \mathbb{Y} \to \mathbb{R}$ is a measurable function with

$$\mathcal{L}Y(B) := \mathbf{P}\{Y \in B\} = \int_B g\, d\mu \quad \text{for all } B \in \mathcal{Y}.$$

In most applications we envisage, either $\mathbb{Y} \subset \mathbb{R}^k$, $\mathcal{Y}$ are the Borel sets, and μ is either Lebesgue measure or some density, or $\mathbb{Y}$ is countable, $\mathcal{Y}$ is the set of all subsets, and μ is the counting measure. In the latter setting, g is a probability mass function on $\mathbb{Y}$.

For $p \in [1, \infty]$ we let $L_p(\mu) := L_p(\mathbb{Y}, \mathcal{Y}, \mu)$ be the Banach space of p-integrable real-valued functions on $\mathbb{Y}$. As usual, functions equal μ-almost everywhere are identified. The norm on $L_p(\mu)$ is given by

$$\|g\|_p := \left\{ \int |g|^p \, d\mu \right\}^{1/p} \quad (g \in L_p(\mu)),$$

with $\|g\|_\infty$ being the essential supremum. If $\mathcal{Y}$ is countably generated (i.e., there exists a countable family $\mathcal{A}$ of subsets of $\mathbb{Y}$ such that $\mathcal{A}$ generates $\mathcal{Y}$), then $L_p(\mu)$ is separable whenever $p < \infty$. If $q \in (1, \infty]$ satisfies $1/p + 1/q = 1$, then $L_q(\mu)$ can be identified with the norm dual of $L_p(\mu)$. We define

$$\langle g, h\rangle := \int gh\, d\mu := \int g(x)h(x)\mu(dx) \quad (g \in L_p(\mu), h \in L_q(\mu)).$$

In the sequel, we consider random variables taking values in $L_p(\mu)$, where $p \in \{1, 2\}$. An $L_p(\mu)$-valued random variable F is a measurable map from $(\Omega, \mathcal{F})$ into $L_p(\mu)$ paired with its Borel sets. By the Pettis measurability theorem, if $L_p(\mu)$ is separable, then a sufficient condition for measurability of F is Borel measurability of the scalar random variable $\Omega \ni \omega \mapsto \langle F(\omega), h\rangle \in \mathbb{R}$ for every h in the dual space $L_q(\mu)$. This condition is easily verified in the applications that follow, and hence further discussion of measurability issues is omitted.

The *expectation* $\mathscr{E}F$ of F is defined as the unique element of $L_p(\mu)$ such that

$$\mathbf{E}\langle F, h\rangle = \langle \mathscr{E}F, h\rangle \quad \text{for every } h \in L_q(\mu),$$

where $\mathbf{E}$ is the usual scalar expectation. If $\mathbf{E}\|F\|_p$ is finite, then $\mathscr{E}F$ exists. $\mathscr{E}F$ is also called the Bochner-Pettis integral of F. (For more details, see, e.g., Bosq [1].) An $L_p(\mu)$-valued random variable G is called *centered Gaussian* if, for every $h \in L_q(\mu)$, the real-valued random variable $\langle G, h\rangle$ is centered Gaussian on $\mathbb{R}$.

A *stochastic kernel* (or *Markov kernel*) P on measurable space $(\mathbb{X}, \mathcal{X})$ is a function $P\colon \mathbb{X} \times \mathcal{X} \to [0, 1]$ such that $B \mapsto P(x, B)$ is a probability measure on $\mathcal{X}$ for all $x \in \mathbb{X}$, and $x \mapsto P(x, B)$ is $\mathcal{X}$-measurable for all $B \in \mathcal{X}$. A discrete-time, $\mathbb{X}$-valued stochastic process $(X_t)_{t\geq 0}$ is called *P-Markov* if $P(x, \cdot)$ is the conditional distribution of X_{t+1} given $X_t = x$. The t step transitions are given by P^t, where

$$P^t(x, B) := \int P^{t-1}(x, dx')P(x', B) \qquad \text{and} \qquad P^1 := P.$$

A probability measure ϕ on $\mathcal{X}$ is called *stationary* for P if

$$\phi(B) = \int P(x, B)\phi(dx) \quad \text{for all } B \in \mathcal{X}.$$

Note that if ϕ is stationary for P, the process $(X_t)_{t\geq 0}$ is P-Markov, and if $\mathcal{L}X_0 = \phi$, then $(X_t)_{t\geq 0}$ is itself (strict sense) stationary.

P is called *ergodic* if P has a unique stationary distribution ϕ, and, for every P-Markov process $(X_t)_{t\geq 0}$ and every measurable $h\colon \mathbb{X} \to \mathbb{R}$ with $\int |h|\, d\phi < \infty$, we have

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t) \to \int h\, d\phi \quad \mathbf{P}\text{-almost surely as } n \to \infty. \tag{1}$$

P is called *geometrically ergodic* if, in addition to ergodicity, there exists a measurable function $V\colon \mathbb{X} \mapsto [0, \infty)$ and nonnegative constants $\lambda < 1$ and $L < \infty$ such that

$$\int V\, d\phi < \infty \qquad \text{and} \qquad \sup_{B\in\mathcal{X}} |P^t(x, B) - \phi(B)| \leq \lambda^t LV(x) \quad \text{for all } x \in \mathbb{X},\ t \in \mathbb{N}. \tag{2}$$

For the purposes of this paper, the main additional benefit of geometric ergodicity over ergodicity is that it implies sufficiently fast mixing for the central limit theorem to hold (at least for functions satisfying a certain moment condition). For a general discussion of ergodicity and geometric ergodicity, see Meyn and Tweedie [17].

Ergodicity and geometric ergodicity have been shown to hold in a range of applications in operations research, finance, economics, and time series analysis. For recent examples see Liebscher [14], Meyn and Tweedie [17], Kamihigashi [12], Kristensen [13], or Nishimura and Stachurski [18].

3. Methodology. Let ψ be a density on measure space $(\mathbb{Y}, \mathcal{Y}, \mu)$, where the σ-algebra $\mathcal{Y}$ is countably generated and μ is σ-finite. Here ψ is the target density that we wish to compute. Let $(\mathbb{X}, \mathcal{X})$ be a measurable space, and let ϕ be a distribution (i.e., probability measure) on $(\mathbb{X}, \mathcal{X})$. Let $q = q(\cdot \mid \cdot)$ be a measurable map from $\mathbb{Y} \times \mathbb{X}$ into $\mathbb{R}_+$ such that $y \mapsto q(y \mid x)$ is a density on $(\mathbb{Y}, \mathcal{Y}, \mu)$ for each $x \in \mathbb{X}$. Suppose further that ψ can be decomposed in terms of q and ϕ, in the sense that

$$\psi(y) = \int q(y \mid x)\phi(dx) \quad \text{for all } y \in \mathbb{Y}. \tag{3}$$

To simulate the target density ψ, we assume the existence of a decomposition (3) such that

(i) The conditional density q can be evaluated, at least numerically.
(ii) There exists a stochastic kernel P on $(\mathbb{X}, \mathcal{X})$ such that ϕ is the unique stationary distribution of P.
(iii) We can simulate P-Markov time series $(X_t)_{t=1}^n$ given $X_0 = x_0 \in \mathbb{X}$.

In this setting, we define the generalized look-ahead estimator (GLAE) of ψ as

$$\psi_n(y) = \frac{1}{n}\sum_{t=1}^{n} q(y \mid X_t) \quad \text{where } (X_t)_{t=1}^n \text{ is } P\text{-Markov}. \tag{4}$$

Examples are presented below. The simplest case is where direct IID sampling from ϕ is feasible. (This is obviously a special case, since IID draws are also Markov.) Letting $(X_t)_{t=1}^n$ be such a sample, we can form ψ_n as in (4). The estimator (4) is very natural in this setting because we then have

$$\mathbf{E}q(y \mid X_t) = \int q(y \mid x)\phi(dx) = \psi(y) \quad \text{for all } t \in \{1, \dots, n\}.$$

Assuming finite second moments, this tells us immediately that $\psi_n(y)$ is unbiased and $\sqrt{n}$-consistent for $\psi(y)$.

To understand why incorporating Markov structure on the simulated process (X_t) is important, suppose now that direct IID sampling from ϕ is infeasible. The Markov chain Monte Carlo solution is to construct a kernel P such that ϕ is the stationary distribution of P, and then generate P-Markov time series. Inserting this series into (4) gives an implementation of the GLAE.

3.1. Examples. The GLAE in (4) generalizes the stationary density look-ahead estimator of Henderson and Glynn [10]. To illustrate this point, consider a P-Markov process taking values in the measure space $(\mathbb{Y}, \mathcal{Y}, \mu)$, where P has the density representation $P(x, B) = \int_B q(y \mid x)\mu(dy)$ for some conditional density q: $\mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$. Suppose that a unique stationary distribution exists. In this setting, it is well known that the stationary distribution can be represented by a density ψ on $\mathbb{Y}$, and, moreover, the density ψ satisfies

$$\psi(y) = \int q(y \mid x)\psi(x)\mu(dx) \quad \text{for all } y \in \mathbb{Y}. \tag{5}$$

Suppose that q is tractable but ψ is not, and one wishes to compute ψ. Although there are several techniques for doing this, the stationary density look-ahead estimator of Henderson and Glynn [10] is perhaps the most attractive. The look-ahead estimator of ψ is defined as $\psi_n(y) = n^{-1}\sum_{t=1}^{n} q(y \mid X_t)$, where $(X_t)_{t=1}^n$ is a simulated P-Markov time series. This is a special case of (4). In particular, comparing (3) and (5), we see that the GLAE reduces to the look-ahead estimator when $\phi = \psi$.

Next we consider two examples that demonstrate how our setting extends the look-ahead estimator to new applications. To begin, consider the following reduced-form model from macroeconomic theory. Suppose that capital stock k_{t+1} can be expressed as a function of lagged capital stock k_t and an exogenous correlated productivity shock z_t. In particular, we assume that (k_t, z_t) obeys

$$k_{t+1} = h(k_t)\, z_{t+1},$$
$$z_{t+1} = g(z_t)\, \xi_{t+1},$$

where (ξ_t) is an IID sequence with density f, and all variables are strictly positive. Note that the pair (k_t, z_t) is jointly Markov, with Markov kernel

$$P((k,z),B) = \mathbf{P}\{(k_{t+1}, z_{t+1}) \in B \mid (k_t, z_t) = (k,z)\} = \mathbf{P}\{(h(k)g(z)\xi_{t+1}, g(z)\xi_{t+1}) \in B\}.$$

Suppose that P is ergodic, and that we wish to compute the stationary density of capital stock. The look-ahead estimator of Henderson and Glynn [10] cannot be directly applied to this problem, because the univariate process (k_t) is not Markovian. Moreover, if we try to compute the joint distribution of (k_t, z_t), which is Markovian, we realize that the conditional distribution $P((k,z),\cdot)$ is not absolutely continuous as a probability measure in $\mathbb{R}^2$, and as such it cannot be expressed as a (conditional) density. In other words, a relationship of the form (5) does not exist. (The essence of the problem is that the two-dimensional process (k_t, z_t) is driven by a one-dimensional shock (ξ_t). In this case, the set of possible outcomes for (k_{t+1}, z_{t+1}) given $(k_t, z_t) = (k,z)$ is a parametric curve in $\mathbb{R}^2$, which has zero Lebesgue measure.)

On the other hand, the GLAE can be applied to the computation of the stationary density of the capital stock. Letting ψ be this density, we observe that, from the law of motion $k_{t+1} = h(k_t)g(z_t)\xi_{t+1}$, we have

$$\psi(k') = \int q(k' \mid k, z)\, d\phi(k,z), \tag{6}$$

where ϕ is the stationary density of P, and $q(k' \mid k,z)$ is the conditional density of $k_{t+1} = h(k_t)g(z_t)\xi_{t+1}$ given $(k_t, z_t) = (k,z)$. In particular, by standard manipulations, we have

$$q(k' \mid k, z) = f\left(\frac{k'}{h(k)g(z)}\right)\frac{1}{h(k)g(z)}.$$

Since (6) is a special case of (3), we can apply the GLAE, simulating $(k_1, z_1), \ldots, (k_n, z_n)$ and then calculating

$$\psi_n(k') = \frac{1}{n}\sum_{t=1}^{n} q(k' \mid k_t, z_t).$$

As a second example of how the GLAE extends the look-ahead estimator, take a GARCH(1, 1) process of the form

$$r_t = \sigma_t W_t, \quad \text{where } (W_t) \overset{\text{IID}}{\sim} N(0,1) \qquad \text{and} \qquad \sigma_{t+1}^2 = \alpha_0 + \beta\sigma_t^2 + \alpha_1 r_t^2. \tag{7}$$

Suppose we wish to compute the stationary density ψ of the returns process $(r_t)_{t\geq 0}$ for some application such as density forecasting, value-at-risk calculation, exact likelihood estimation, or model assessment. Let ϕ be the stationary distribution of $X_t := \sigma_t^2$. (We assume that all parameters are strictly positive and $\alpha_1 + \beta < 1$. This is enough to guarantee existence of a unique stationary distribution for (X_t), and hence a stationary density ψ for (r_t). See Lemma 6.3 in §6 for details.) The look-ahead estimator cannot be applied to this problem, for reasons similar to the macroeconomic model discussed above. However, we can use the GLAE as follows: Equation (7) implies that $r_t = \sqrt{X_t}W_t$, and hence the conditional density $q(r \mid x)$ of r_t given $X_t = x$ is centered Gaussian with variance x. For this q we have $\psi(r) = \int q(r \mid x)\phi(dx)$, which is a version of (3). The process $(X_t)_{t\geq 0}$ can be expressed as

$$X_{t+1} = \alpha_0 + \beta X_t + \alpha_1 X_t W_t^2. \tag{8}$$

After simulating a time series $(X_t)_{t=1}^n$ from this process, the GLAE of ψ can be formed as

$$\psi_n(r) = \frac{1}{n}\sum_{t=1}^{n} q(r \mid X_t) = \frac{1}{n}\sum_{t=1}^{n} (2\pi X_t)^{-1/2} \exp\left\{-\frac{r^2}{2X_t}\right\}. \tag{9}$$

4. Results. In this section we provide a general asymptotic theory of the GLAE. To begin, notice that when ψ_n was defined in (4), the distribution of X_0 was not specified. When it is possible to draw X_0 from ϕ, we have the following result, the proof of which is provided in §6:

LEMMA 4.1. *If $\mathcal{L}X_0 = \phi$, then ψ_n is unbiased, in the sense that $\mathscr{E}\psi_n = \psi$.*

(Here $\mathscr{E}$ is the $L_1(\mu)$ Bochner-Pettis expectation, as defined in §2.) In many applications, there is no obvious way to sample directly from ϕ, and Lemma 4.1 cannot be applied. However, when P is ergodic, ψ_n is asymptotically unbiased for large t, and also consistent:

THEOREM 4.1. *If P is ergodic, then ψ_n is strongly globally consistent and asymptotically unbiased, in the sense that*

(i) *$\psi_n \to \psi$ in $L_1(\mu)$ as $n \to \infty$ with probability one*; *and*
(ii) *$\mathscr{E}\psi_n \to \psi$ in $L_1(\mu)$ as $n \to \infty$.*

Notice that Theorem 4.1 requires nothing beyond ergodicity. (In particular, there are no moment conditions, and no continuity or compactness conditions—$\mathbb{X}$ and $\mathbb{Y}$ do not even need topologies or algebraic structure.)

The $L_1(\mu)$ norm used in Theorem 4.1 is perhaps the most natural way to measure deviation between two densities. The deviation is finite and uniformly bounded across the set of densities, and Scheffé's identity and Theorem 4.1 imply that if $\psi_n \to \psi$ in L_1, then the maximum deviation in probabilities over all events converges to zero. For further discussion of the advantages of using the L_1 norm, see Devroye and Lugosi [6].

On the other hand, $L_1(\mu)$ is not a Hilbert space, and, without the Hilbert space property, asymptotic normality is problematic. To prove asymptotic normality, we now shift our analysis into the Hilbert space $L_2(\mu)$. To do so, we add a second moment condition, as well as a stricter form of ergodicity.

We begin with a definition. For each $x \in \mathbb{X}$, let $T(x)$ represent the function $y \mapsto q(y \mid x) - \psi(y)$, and define the linear operator C: $L_2(\mu) \to L_2(\mu)$ by

$$\langle g, Ch\rangle = \mathbf{E}\langle g, T(X_1^*)\rangle\langle h, T(X_1^*)\rangle + \sum_{t\geq 2}^{\infty}\mathbf{E}\langle g, T(X_1^*)\rangle\langle h, T(X_t^*)\rangle + \sum_{t\geq 2}^{\infty}\mathbf{E}\langle h, T(X_1^*)\rangle\langle g, T(X_t^*)\rangle$$

for arbitrary $h, g \in L_2(\mu)$, where $(X_t^*)_{t\geq 0}$ is stationary and P-Markov. (That is, $(X_t^*)_{t\geq 0}$ is P-Markov and X_0^* is drawn from the stationary distribution ϕ.) Under the conditions of the following theorem, C is a well-defined operator from $L_2(\mu)$ to itself:

Theorem 4.2. *Let P be geometrically ergodic. If, in addition, for the function V in (2), there exist positive constants $m_0 < \infty$, $m_1 < \infty$, and $\gamma < 1$ such that*

$$\int q(y \mid x)^2 \mu(dy) \leq m_0 + m_1 V(x)^\gamma \quad \text{for all } x \in \mathbb{X}, \tag{10}$$

then $\sqrt{n}(\psi_n - \psi)$ converges in distribution to a centered Gaussian G on $L_2(\mu)$ with covariance operator C.

Regarding the statement of the theorem, a centered Gaussian G is said to have covariance operator C if $\mathbf{E}\langle g, G\rangle\langle h, G\rangle = \langle Cg, h\rangle$ for every $g, h \in L_2(\mu)$. Also, convergence in distribution is defined in the obvious way: Let $\mathcal{C}$ be the continuous, bounded, real-valued functions on $L_2(\mu)$, where continuity is with respect to the norm topology. Let $(G_n)_{n\geq 0}$ be $L_2(\mu)$-valued random variables. Then $G_n \to G_0$ in distribution if $\mathbf{E}h(G_n) \to \mathbf{E}h(G_0)$ for every $h \in \mathcal{C}$. For more details on Hilbert space valued random variables and Hilbert central limit theorems, see, for example, Merlevède et al. [16], Chen and White [4], or Bosq [1].

One immediate consequence of Theorem 4.2 is that since $h \mapsto \|h\|$ is continuous on $L_2(\mu)$, Theorem 4.2 and the continuous mapping theorem imply that $\|\psi_n - \psi\| = O_P(n^{-1/2})$. In other words, ψ_n is globally $\sqrt{n}$-consistent for ψ when viewed as a sequence of random functions in $L_2(\mu)$.

A second comment on our results is that in Theorems 4.1 and 4.2, we do not assume that the simulated process (X_t) is itself stationary. This is important, because simulating a stationary process would require drawing X_0 such that $\mathcal{L}X_0 = \phi$. In many settings the stationary distribution ϕ is unknown, and generating such a draw is problematic. A more convenient approach is to set X_0 equal to an arbitrary element of the state space. Since we do not assume stationarity of (X_t), our results are valid in this setting.

On the other hand, permitting X_0 to be an arbitrary element of the state space complicates the proofs slightly, since Banach space laws of large numbers and central limit theorems typically assume stationarity of the process. In the proofs below, we adapt a Hilbert space central limit theorem of Merlevède et al. [16] to our nonstationary setting in order to incorporate this case.

As a final remark, note that when IID sampling from ϕ is possible, the conclusions of Theorem 4.1 hold without any conditions on q and ϕ, and Theorem 4.2 holds whenever

$$\left\{\int q(y \mid x)^2 \mu(dy) < \infty \quad \forall x \in \mathbb{X}\right\} \qquad \text{and} \qquad \iint q(y \mid x)^2 \mu(dy)\phi(dx) < \infty.$$

5. Discussion. As an example of how the theory applies, consider again the estimator (9) presented in §3.1. Assume as before that all parameters in (7) are positive and $\alpha_1 + \beta < 1$. Under these assumptions, the process $(X_t)_{t\geq 0}$ defined in (8) is geometrically ergodic on $\mathbb{X} := [\alpha_0/(1-\beta), \infty)$ for V in (2) chosen as $V(x) = x$. The details of the geometric ergodicity argument are given in Lemma 6.3 below. Regarding the moment condition (10) in Theorem 4.2, we have

$$\int q(r \mid x)^2\, dr = \int (2\pi x)^{-1} \exp\left\{-\frac{r^2}{x}\right\} dr = (4\pi x)^{-1/2} \leq \left\{\frac{4\pi\alpha_0}{1-\beta}\right\}^{-1/2} \quad \text{for all } x \in \mathbb{X}.$$

Since the right-hand side of this expression is a finite constant and $V(x) = x$ is nonnegative, the moment condition (10) is certainly satisfied. As a result, both Theorems 4.1 and 4.2 apply.

Before continuing, let us make a brief comparison of (4) with nonparametric kernel density estimation. To define the latter, we must restrict attention to the case where $\mathbb{Y} \subset \mathbb{R}^k$. Assume that one can generate IID samples $Y_1, \ldots, Y_n$ from ψ. The NPKDE f_n is then defined in terms of a kernel (i.e., density) K on $\mathbb{Y}$ and a "bandwidth" parameter δ_n:

$$f_n(y) := \frac{1}{n\delta_n} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{\delta_n}\right). \tag{11}$$

The estimate f_n is known to be consistent, in the sense that $\mathbf{E}\|f_n - \psi\|_1 \to 0$ whenever $\delta_n \to 0$ and $n\delta_n^k \to \infty$ (Devroye and Lugosi [6]). However, rates of convergence are slower than the parametric rate $O_P(n^{-1/2})$. For example, if we fix $y \in \mathbb{Y}$ and take ψ to be twice differentiable, then, for a suitable choice of K, it can be shown that

$$|f_n(y) - \psi(y)| = O_P[(n\delta_n^k)^{-1/2}] \quad \text{when } n\delta_n^k \to \infty \text{ and } (n\delta_n^k)^{1/2}\delta_n^2 \to 0.$$

Thus, even with this smoothness assumption on ψ—which may or may not hold in practice—the convergence rate $O_P[(n\delta_n^k)^{-1/2}]$ of the NPKDE is slower than the rate $O_P(n^{-1/2})$ obtained for ψ_n. Moreover, the rate of convergence slows as the dimension k of $\mathbb{Y}$ increases. (Of course, the slower rate of convergence for the NPKDE is not surprising, as the NPKDE uses no information beyond the sample and some smoothness inherited from the kernel, while the GLAE in (4) makes direct use of the model that generated the sample. The converse of this logic is that the NPKDE can be applied in statistical settings where the underlying model is unknown.)

5.1. Simulation results. Consider the GARCH application in §3.1. For the exercise, we set $\alpha_0 = \alpha_1 = 0.05$ and $\beta = 0.9$, which are reasonable benchmarks for GARCH models of asset price data such as stock indices. To investigate small sample properties, we set $n = 500$. The fast convergence of ψ_n implied by Theorem 4.2 is illustrated in Figure 1. The left panel of the figure contains the true density ψ, drawn in bold, as well as 50 replications of a NPKDE, drawn in grey. Each NPKDE replication uses a simulated time series $(r_t)_{t=1}^n$, combined with standard default settings—a Gaussian kernel and bandwidth calculated according to Silverman's rule. (The density marked as "true" in the figure is in fact an approximation, calculated by simulation with $n = 10^7$. For such a large n there is no visible variation of the density over different realizations, or different methods of simulation.) The right panel of Figure 1 repeats the exercise, but this time using the GLAE in (9) rather than the NPKDE.

The estimator (9) exhibits better small-sample properties than the NPKDE. The replications are more tightly clustered around the true distribution both at the center of the distribution and at the tails. (This occurs despite the fact that, by construction, the NPKDE foregoes unbiasedness in order to obtain lower variance.) To quantify the results of Figure 1, we looked at the L_1-norm deviations from the true density ψ. We computed average L_1 deviations over 1,000 replications. For $n = 500$, the ratio of the GLAE L_1 deviation to the NPKDE L_1 deviation was 0.5854. In other words, average L_1 error for the NPKDE was 71% larger than that of the GLAE.

This simulation exercise considered a one-dimensional problem. The strong performance of the GLAE relative to the NPKDE is likely to be significantly greater in higher-dimensional problems, since the rate of converge of the NPKDE falls as the dimension of the state space increases.

6. Technical appendix. This section contains the proofs. In what follows, $\|\cdot\|$ represents either the norm on $L_1(\mu)$ or the norm on $L_2(\mu)$, depending on the context. (The proofs of Lemma 4.1 and Theorem 4.1 are set in $L_1(\mu)$, while that of Theorem 4.2 is set in $L_2(\mu)$.)

Consider the setting of §§3 and 4. We begin with the following lemma:

LEMMA 6.1. *If $\mathcal{L}X = \phi$, then $\mathscr{E}q(\cdot \mid X) = \psi$ in $L_1(\mu)$. If $\mathcal{L}X = \phi$ and the conditions of Theorem 4.2 hold, then $\psi \in L_2(\mu)$ and $\mathscr{E}q(\cdot \mid X) = \psi$ in $L_2(\mu)$.*

PROOF. To begin with the case of $L_1(\mu)$, let $\|\cdot\|$ be the $L_1(\mu)$ norm and observe that, by the definition of q, we have $\|q(\cdot \mid x)\| = 1$ for all $x \in \mathbb{X}$. As a result, $\mathbf{E}\|q(\cdot \mid X)\| = 1 < \infty$, and the Bochner-Pettis expectation $\mathscr{E}q(\cdot \mid X)$ is well defined. To show that $\mathscr{E}q(\cdot \mid X) = \psi$, we must prove in addition that $\mathbf{E}\langle q(\cdot \mid X), h\rangle = \langle \psi, h\rangle$ for all $h \in L_\infty(\mu)$. Fixing $h \in L_\infty(\mu)$, Fubini's theorem and (3) yield

$$\mathbf{E}\langle q(\cdot \mid X), h\rangle = \mathbf{E}\int q(y \mid X)h(y)\mu(dy) = \int \mathbf{E}q(y \mid X)h(y)\mu(dy).$$

By (3) this equals $\int \psi h \, d\mu = \langle \psi, h\rangle$, as was to be shown.

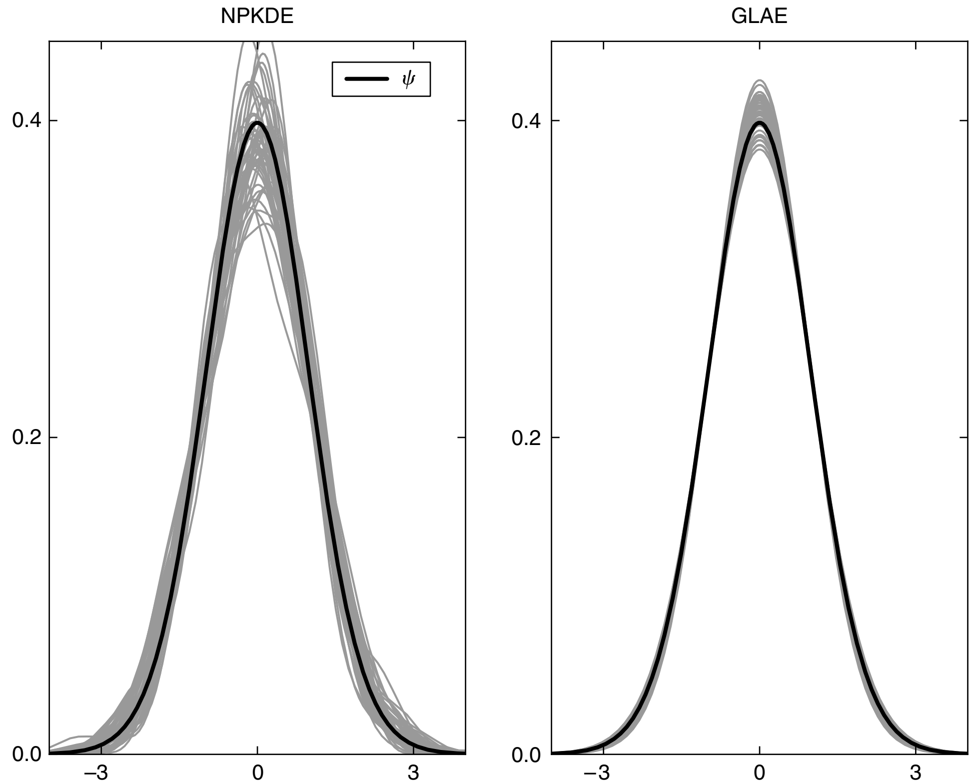

FIGURE 1. Relative performance, $n = 500$.

For the proof of the $L_2(\mu)$ case, let $\|\cdot\|$ be the $L_2(\mu)$ norm. We first verify that $\psi \in L_2(\mu)$. Let m_0, m_1, V and γ be as in the statement of Theorem 4.2. Using the expression for ψ in (3), two applications of Jensen's inequality (first with a convex and then with a concave function), and the bound in Theorem 4.2, we have

$$\begin{aligned}\int \psi(y)^2 \mu(dy) &= \int \left[\int q(y \mid x)\phi(dx)\right]^2 \mu(dy)\\ &\le \int\int q(y\mid x)^2 \phi(dx)\mu(dy)\\ &= \int\int q(y\mid x)^2\mu(dy)\phi(dx)\\ &\le m_0 + m_1 \int V(x)^\gamma \phi(dx) \le m_0 + m_1\left[\int V(x)\phi(dx)\right]^\gamma.\end{aligned}$$

The final expression is finite by (2).

Next we verify that $\mathbf{E}\|q(\cdot \mid X)\| < \infty$, which is necessary to ensure that the expectation $\mathscr{E}q(\cdot \mid X)$ is well defined in $L_2(\mu)$. For this, it suffices that

$$\mathbf{E}\|q(\cdot \mid X)\|^2 = \mathbf{E}\int q(y\mid X)^2\mu(dy) = \int\int q(y\mid x)^2 \mu(dy)\phi(dx) < \infty.$$

Finiteness of this term was established immediately above, and hence $\mathscr{E}q(\cdot \mid X)$ is well defined in $L_2(\mu)$. Thus it remains only to verify that $\mathscr{E}q(\cdot \mid X) = \psi$ in $L_2(\mu)$. The proof of this claim is essentially identical to the L_1 case so we omit it. □

We can now prove Lemma 4.1:

PROOF OF LEMMA 4.1. Assume the conditions of the lemma. Since ϕ is stationary for P and $\mathscr{L}X_0 = \phi$, we have $\mathscr{L}X_t = \phi$ for all $t \ge 0$. From linearity of $\mathscr{E}$ and Lemma 6.1, we conclude that

$$\mathscr{E}\psi_n = \mathscr{E}\left[\frac{1}{n}\sum_{t=1}^n q(\cdot \mid X_t)\right] = \frac{1}{n}\sum_{t=1}^n \mathscr{E}q(\cdot\mid X_t) = \psi.$$

In other words, ψ_n is unbiased, as was to be shown. □

6.1. Proof of Theorem 4.1. In this section we provide the proof of Theorem 4.1, and $\|\cdot\|$ always represents the $L_1(\mu)$ norm. The arguments are standard constructions from laws of large numbers in Banach space. Our first observation is that part (ii) of the theorem (asymptotic unbiasedness) follows from part (i) (strong consistency). Indeed, suppose that $\psi_n \to \psi$ almost surely in $L_1(\mu)$. Using standard properties of the Bochner-Pettis integral, we obtain

$$\|\mathscr{E}\psi_n - \psi\| = \|\mathscr{E}\psi_n - \mathscr{E}\psi\| \le \mathbf{E}\|\psi_n - \psi\|.$$

Since $\|f - g\| \le 2$ for any pair of densities f and g, the right-hand side converges to zero by the dominated convergence theorem.

Let us turn now to the claim that $\psi_n \to \psi$ almost surely in $L_1(\mu)$. As in the statement of the theorem, let P be an ergodic stochastic kernel on $(\mathbb{X}, \mathscr{X})$ with stationary distribution ϕ. Let $(X_t)_{t\ge 0}$ be P-Markov and let X^* be a random variable on $\mathbb{X}$ with $\mathscr{L}X^* = \phi$. Define $T(x) := q(\cdot \mid x) - \psi$, which is a measurable function from $\mathbb{X}$ to $L_1(\mu)$. Note that $\mathscr{E}T(X^*) = 0$ by Lemma 6.1.

We need to show that

$$\lim_{n\to\infty} \|\psi_n - \psi\| = \lim_{n\to\infty} \left\| \frac{1}{n} \sum_{t=1}^{n} T(X_t) \right\| = 0 \quad (\mathbf{P}\text{-almost surely}). \tag{12}$$

Fix $\epsilon > 0$. Since $L_1(\mu)$ is separable, we can choose a partition $\{B_j\}_{j\in\mathbb{N}}$ of $L_1(\mu)$ such that each B_j has diameter less than ϵ. For any $L_1(\mu)$-valued random variable U, we let $L_J U := \sum_{j=1}^{J} b_j \mathbb{1}\{U \in B_j\}$, where, for each j, b_j is a fixed point in B_j. Thus, $L_J U$ is a simple random variable that approximates U. In particular, we have the following result, a proof of which can be found in Bosq [1, pp. 27–28]:

$$\exists J \in \mathbb{N} \text{ with } \mathbf{E}\|T(X^*) - L_J T(X^*)\| < \epsilon. \tag{13}$$

Our first claim is that

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{t=1}^{n} L_J T(X_t) - \mathscr{E} L_J T(X^*) \right\| = 0 \quad (\mathbf{P}\text{-almost surely}). \tag{14}$$

To establish (14), we can use the real ergodic law (1) to obtain

$$\frac{1}{n} \sum_{t=1}^{n} L_J T(X_t) = \sum_{j=1}^{J} b_j \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{T(X_t) \in B_j\} \to \sum_{j=1}^{J} b_j \mathbf{P}\{T(X^*) \in B_j\} = \mathscr{E} L_J T(X^*)$$

almost surely, where the last equality follows immediately from the definition of $\mathscr{E}$. Thus (14) is established.

Returning to (12), we have

$$\left\| \frac{1}{n} \sum_{t=1}^{n} T(X_t) \right\| \le \frac{1}{n} \sum_{t=1}^{n} \|T(X_t) - L_J T(X_t)\| + \left\| \frac{1}{n} \sum_{t=1}^{n} L_J T(X_t) - \mathscr{E} L_J T(X^*) \right\| + \|\mathscr{E} L_J T(X^*)\|.$$

Using real-valued ergodicity again, as well as (14), we get

$$\limsup_{n\to\infty} \left\| \frac{1}{n} \sum_{t=1}^{n} T(X_t) \right\| \le \mathbf{E}\|T(X^*) - L_J T(X^*)\| + \|\mathscr{E} L_J T(X^*)\|.$$

But the fact that $\mathscr{E}T(X^*) = 0$ now gives

$$\|\mathscr{E} L_J T(X^*)\| = \|\mathscr{E}T(X^*) - \mathscr{E} L_J T(X^*)\| \le \mathbf{E}\|T(X^*) - L_J T(X^*)\|.$$

In view of (13) we then have

$$\limsup_{n\to\infty} \left\| \frac{1}{n} \sum_{t=1}^{n} T(X_t) \right\| \le 2\epsilon \quad (\mathbf{P}\text{-almost surely}).$$

Since ϵ is arbitrary, the proof of (12) is now done.

6.2. Proof of Theorem 4.2. In this section we provide the proof of Theorem 4.2, and the assumptions of that theorem will be in force throughout. Moreover, the symbol $\|\cdot\|$ always represents the $L_2(\mu)$ norm. Throughout the proof, for $x \in \mathbb{X}$ we let $T(x)$ be the function $y \mapsto q(y \mid x) - \psi(y)$. In this notation, Theorem 4.2 amounts to the claim that

$$\mathcal{L}\left[n^{-1/2}\sum_{t=1}^{n} T(X_t)\right] \to N(0, C) \quad (n \to \infty), \tag{15}$$

where C is the operator defined in §4.

Our first lemma shows that, given our ergodicity assumptions on P, we can restrict attention to the case where $\mathcal{L}X_1 = \phi$ when proving (15).

Lemma 6.2. *Let $(X_t)_{t\geq 1}$ and $(X'_t)_{t\geq 1}$ be two P-Markov chains, where $\mathcal{L}X_1 = \phi$ and $X'_1 = x \in \mathbb{X}$. For any Borel probability measure ν on $L_2(\mu)$,*

$$\mathcal{L}\left[n^{-1/2}\sum_{t=1}^{n} T(X_t)\right] \to \nu \qquad \textit{implies} \qquad \mathcal{L}\left[n^{-1/2}\sum_{t=1}^{n} T(X'_t)\right] \to \nu.$$

Proof. Given our assumption of geometric ergodicity (and hence ergodicity), it is well known (see, for example, Meyn and Tweedie [17, Theorem 17.1.7], and Lindvall [15, Theorem 21.12]) that one can construct P-Markov processes $(X_t)_{t\geq 1}$ and $(X'_t)_{t\geq 1}$ on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that

$$\tau := \inf\{t \in \mathbb{N}\colon X_t = X'_t\}$$

is finite almost surely, and $X_t = X'_t$ for all $t \geq \tau$. Let $S_n := \sum_{t=1}^{n} T(X_t)$ and $S'_n := \sum_{t=1}^{n} T(X'_t)$, and assume as in the statement of the lemma that $n^{-1/2}S_n \to \nu$. To prove that $n^{-1/2}S'_n \to \nu$ it suffices to show that the (norm) distance between $n^{-1/2}S'_n$ and $n^{-1/2}S_n$ converges to zero in probability (cf., e.g., Dudley [7, Lemma 11.9.4]). Fixing $\epsilon > 0$, we need to show that

$$\mathbf{P}\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \epsilon\} \to 0 \quad (n \to \infty). \tag{16}$$

Clearly

$$\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \epsilon\} \subset \left\{\sum_{t=1}^{n} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\}.$$

Fix $k \in \mathbb{N}$, and partition the last set over $\{\tau \leq k\}$ and $\{\tau > k\}$ to obtain the disjoint sets

$$\left\{\sum_{t=1}^{n} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\} \cap \{\tau \leq k\} \subset \left\{\sum_{t=1}^{k} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\},$$

and

$$\left\{\sum_{t=1}^{n} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\} \cap \{\tau > k\} \subset \{\tau > k\}.$$

Together, these lead to the bound

$$\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \epsilon\} \subset \left\{\sum_{t=1}^{k} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\} \cup \{\tau > k\}$$

$$\therefore \quad \mathbf{P}\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \epsilon\} \leq \mathbf{P}\left\{\sum_{t=1}^{k} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\} + \mathbf{P}\{\tau > k\}.$$

For any fixed k, we have

$$\lim_{n\to\infty} \mathbf{P}\left\{\sum_{t=1}^{k} \|T(X'_t) - T(X_t)\| > n^{1/2}\epsilon\right\} = 0. \tag{17}$$

Hence

$$\limsup_{n\to\infty} \mathbf{P}\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \epsilon\} \leq \mathbf{P}\{\tau > k\}, \quad \forall k \in \mathbb{N}.$$

Since $\mathbf{P}\{\tau < \infty\} = 1$ taking $k \to \infty$ yields (16). □

In view of Lemma 6.2, we can continue the proof of (15) while considering only the case $\mathcal{L}X_1 = \phi$. In this case $(T(X_t))$ is a centered (see Lemma 6.1) strict sense stationary stochastic process in L_2, and we can apply the stationary Hilbert CLT in Merlevède et al. [16, Theorem 4, Corollary 1]. From the latter we obtain the following result: Let $\xi_t := T(X_t)$ for all t. Define the corresponding mixing coefficients by

$$\alpha(t) := \sup |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)|,$$

where the supremum is over all $A \in \sigma(\xi_1)$ and $B \in \sigma(\xi_{t+1})$. In this setting, the convergence in (15) will be valid whenever there exists a constant $\delta > 0$ such that

$$\mathbf{E}\|\xi_t\|^{2+\delta} < \infty \qquad \text{and} \qquad \sum_{t=1}^{\infty} t^{2/\delta}\alpha(t) < \infty. \tag{18}$$

(The definition of the mixing coefficient used here is slightly different from the one used in Merlevède et al. [16, Definition 1]. However, in the Markov case it is well known that the two are equivalent. See, for example, Bradley [2, Section 3].)

We first establish the finite expectation on the left-hand side of (18). Let $\delta := 2(1-\gamma)/\gamma$, where γ is the constant in (10). Let m_0, m_1, and V be the constants and function in the same equation. From the definitions we have

$$\begin{aligned}\|\xi_t\|^{2+\delta} = \|T(X_t)\|^{2+\delta} &= \left\{\int [q(y \mid X_t) - \psi(y)]^2 \mu(dy)\right\}^{(2+\delta)/2} \\ &= \left\{\int [q(y \mid X_t) - \psi(y)]^2 \mu(dy)\right\}^{1/\gamma}.\end{aligned}$$

Applying an elementary inequality followed by the bound in (10), we obtain

$$\begin{aligned}\int [q(y \mid x) - \psi(y)]^2 \mu(dy) &\le 2\int q(y \mid x)^2 \mu(dy) + 2\int \psi(y)^2 \mu(dy) \\ &\le 2m_0 + 2m_1 V(x)^\gamma + 2\int \psi(y)^2 \mu(dy).\end{aligned}$$

(Finiteness of $\int \psi^2\, d\mu$ is guaranteed by Lemma 6.1.) By assumption, $\gamma < 1$. Hence $1/\gamma > 1$, and we can apply the previous bound and Jensen's inequality to obtain

$$\begin{aligned}\|\xi_t\|^{2+\delta} &\le \left\{2m_0 + 2m_1 V(X_t)^\gamma + 2\int \psi(y)^2 \mu(dy)\right\}^{1/\gamma} \\ &= \left\{\frac{6m_0}{3} + \frac{6m_1 V(X_t)^\gamma}{3} + \frac{6\|\psi\|^2}{3}\right\}^{1/\gamma} \\ &\le \frac{1}{3}\{[6m_0]^{1/\gamma} + [6m_1 V(X_t)^\gamma]^{1/\gamma} + [6\|\psi\|^2]^{1/\gamma}\}.\end{aligned}$$

In other words, there exist finite constants c_1 and c_2 such that

$$\|\xi_t\|^{2+\delta} \le c_1 V(X_t) + c_2.$$

Taking expectations and applying the first expression in (2) gives $\mathbf{E}\|\xi_t\|^{2+\delta} < \infty$ as required.

The last step of the proof of Theorem 4.2 is to verify the finiteness of the sum on the right-hand side of (18). An elementary argument shows the following ordering of σ-algebras:

$$\sigma(\xi_j) = \sigma(T(X_j)) \subset \sigma(X_j), \quad \forall\, j.$$

As a result, we have

$$\alpha(t) := \sup_{\substack{A \in \sigma(\xi_1) \\ B \in \sigma(\xi_{t+1})}} |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| \le \sup_{\substack{A \in \sigma(X_1) \\ B \in \sigma(X_{t+1})}} |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)|.$$

The right-hand side gives the strong mixing coefficients for (X_t), which, in the geometrically ergodic case, are known to be $O(\lambda^t)$ for the constant λ in (2). (See, for example, Jones [11, p. 304].) As a consequence, we have $\alpha(t) = O(\lambda^t)$, and hence $\sum_{t=1}^{\infty} t^{2/\delta}\alpha(t)$ will be finite if $\sum_{t=1}^{\infty} t^{2/\delta}\lambda^t$ is finite. Since $\lambda < 1$, this last sum is clearly finite. This completes the proof of Theorem 4.2.

6.3. Properties of the GARCH process. Consider the process $X_{t+1} = \alpha_0 + \beta X_t + \alpha_1 X_t W_t^2$ in (8), where (W_t) is IID and standard normal. Let $\bar{x} := \alpha_0/(1-\beta)$ and $\mathbb{X} := [\bar{x}, \infty)$. Let P be the associated stochastic kernel on $\mathbb{X}$. The following lemma establishes geometric ergodicity of P:

LEMMA 6.3. *If α_0, β, and α_1 are positive constants with $\beta + \alpha_1 < 1$, then P is geometrically ergodic. In particular, there exist constants $\lambda < 1$ and $L < \infty$ such that (2) is satisfied with $V(x) = x$.*

PROOF. We use the conditions of Hairer and Mattingly [9, Theorem 1.2], which imply that the claim in the lemma is true whenever the following two conditions hold:

(i) There exist constants $\kappa < 1$ and $K < \infty$ such that

$$\mathbf{E}(\alpha_0 + \beta x + \alpha_1 x W_t^2) \leq \kappa x + K \quad \text{for all } x \in \mathbb{X}.$$

(ii) For some $R > 2K/(1-\kappa)$ there exist an $\epsilon > 0$ and a probability measure ν such that

$$\inf_{\bar{x} \leq x \leq R} P(x, B) \geq \epsilon \nu(B) \quad \text{for all Borel subsets } B \text{ of } \mathbb{X}.$$

Condition (i) is obvious because $\mathbf{E}(\alpha_0 + \beta x + \alpha_1 x W_t^2) = \alpha_0 + (\beta + \alpha_1)x$ and $\beta + \alpha_1 < 1$ by assumption. Regarding (ii), let R be any constant strictly greater than the maximum of $2K/(1-\kappa)$ and $\bar{x}$. Define $U_t := \beta + \alpha_1 W_t^2$, so that $X_{t+1} = \alpha_0 + X_t U_{t+1}$. Let g be the density of the real random variable U_t. Observe that g is continuous and strictly positive on $[\beta + \alpha_1, \infty)$. Since $\beta + \alpha_1 < 1 < R/\bar{x}$, it follows that

$$\delta := \min\{g(u)\colon \beta + \alpha_1 \leq u \leq R/\bar{x}\} > 0.$$

Since $X_{t+1} = \alpha_0 + X_t U_{t+1}$, the conditional density $p(x, y)$ of X_{t+1} given $X_t = x$ has the form

$$p(x, y) = g\left(\frac{y - \alpha_0}{x}\right)\frac{1}{x}.$$

Let q be the function

$$q(y) = \frac{\delta}{R}\mathbb{1}\{\alpha_0 + (\beta + \alpha_1)R \leq y \leq \alpha_0 + R\} \quad (y \in \mathbb{X}).$$

If we can show that

$$p(x, y) \geq q(y) \quad \text{whenever } \bar{x} \leq x \leq R \text{ and } y \in \mathbb{X}, \tag{19}$$

then condition (ii) is verified. Indeed, since $\delta > 0$ and $(\beta + \alpha_1)R < R$, the function q has a nonzero Lebesgue integral $\epsilon = \int_{\bar{x}}^{\infty} q(y)\, dy$, and hence, applying (19),

$$\inf_{\bar{x} \leq x \leq R} P(x, B) \geq \int_B q(y)\, dy = \epsilon \int_B \frac{q(y)}{\epsilon}\, dy.$$

It remains only to show that (19) is valid. To see that this is so, suppose first that $\bar{x} \leq x \leq R$ and $\alpha_0 + (\beta + \alpha_1)R \leq y \leq \alpha_0 + R$. In this case we have

$$\beta + \alpha_1 = \frac{\alpha_0 + (\beta + \alpha_1)R - \alpha_0}{R} \leq \frac{y - \alpha_0}{x} \leq \frac{\alpha_0 + R - \alpha_0}{\bar{x}} = \frac{R}{\bar{x}},$$

and hence, by the definition of δ,

$$p(x, y) = g\left(\frac{y - \alpha_0}{x}\right)\frac{1}{x} \geq \delta \frac{1}{x} \geq \frac{\delta}{R} \geq q(y).$$

Regarding the other case, suppose now that $\bar{x} \leq x \leq R$ holds but $\alpha_0 + (\beta + \alpha_1)R \leq y \leq \alpha_0 + R$ fails. In this case $q(y) = 0$, so (19) is trivially satisfied. Hence condition (ii) is verified, and the proof of Lemma 6.3 is done. □

References

[1] Bosq D (2000) *Linear Processes in Function Space* (Springer-Verlag, New York).
[2] Bradley RC (2005) Basic properties of strong mixing conditions: A survey and some open questions. *Probab. Surveys* 2:107–144.
[3] Brandt MW, Santa-Clara P (2002) Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *J. Financial Econom.* 63:161–210.
[4] Chen X, White H (1998) Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory* 14:260–284.
[5] Danielsson J (1994) Stochastic volatility in asset prices estimation with simulated maximum likelihood. *J. Econometrics* 64(1–2):375–400.
[6] Devroye L, Lugosi G (2001) *Combinatorial Methods in Density Estimation* (Springer-Verlag, New York).
[7] Dudley RM (2002) Cambridge studies in advanced mathematics No. 74. *Real Analysis and Probability* (Cambridge University Press).
[8] Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* 85:398–409.
[9] Hairer M, Mattingly JC (2011) Yet another look at Harris' ergodic theorem for Markov chains. Dalang RC, Dozzi M, Russo F, eds. *Seminar on Stochastic Analysis, Random Fields, and Applications VI* (Springer Basel) 109–117.
[10] Henderson SG, Glynn PW (2001) Computing densities for Markov chains via simulation. *Math. Oper. Res.* 26:375–400.
[11] Jones GL (2004) On the Markov chain central limit theorem. *Probab. Surveys* 1:299–320.
[12] Kamihigashi T (2007) Stochastic optimal growth with bounded or unbounded utility and with bounded or unbounded shocks. *J. Math. Econom.* 43(3–4):477-500.
[13] Kristensen D (2008) Uniform ergodicity of a class of Markov chains with applications to time series models. Mimeo, Columbia University.
[14] Liebscher E (2005) Toward a unified approach to proving geometric ergodicity and mixing properties of nonlinear autoregressive processes, *J. Time Series Anal.* 26(5):669–689.
[15] Lindvall T (2002) *Lectures on the Coupling Method* (Dover Publications, Mineola, NY).
[16] Merlevède F, Peligrad M, Utev S (1997) Sharp conditions for the CLT of linear processes in a Hilbert space. *J. Theor. Probab.* 10(3): 681–693.
[17] Meyn S, Tweedie RL (2009) *Markov Chains and Stochastic Stability*, 2nd ed. (Cambridge University Press).
[18] Nishimura K, Stachurski J (2005) Stability of stochastic optimal growth models: A new approach. *J. Econom. Theory* 122(1):100–118.
[19] Pedersen AR (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian J. Statist.* 22:55–71.
[20] Stachurski J, Martin V (2008) Computing the distributions of economic models via simulation. *Econometrica* 76(2):443–450.