



# Generalized Look-Ahead Methods for Computing Stationary Densities

R. Anton Braun  
Research Department, Federal Reserve Bank of Atlanta  
[r.anton.braun@atl.frb.org](mailto:r.anton.braun@atl.frb.org)

Huiyu Li  
Graduate School, Department of Economics, Stanford  
University  
[tohuiyu@gmail.com](mailto:tohuiyu@gmail.com)

John Stachurski  
Research School of Economics, Australian National  
University  
[john.stachurski@anu.edu.au](mailto:john.stachurski@anu.edu.au)

ANU Working Papers in Economics and Econometrics  
No. 558

October, 2011

ISBN: 086831 558 3

# Generalized Look-Ahead Methods for Computing Stationary Densities\*

R. Anton Braun

Research Department, Federal Reserve Bank of Atlanta

`r.anton.braun@atl.frb.org`

Huiyu Li

Graduate School, Department of Economics, Stanford University

`tohuiyu@gmail.com`

John Stachurski

Research School of Economics, Australian National University

`john.stachurski@anu.edu.au`

October 11, 2011

## Abstract

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 much 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.

## 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

---

\*The authors gratefully acknowledge helpful comments from Manuel Arellano, Fumio Hayashi, Hidehiko Ichimura, Albert Marcet, Enrique Sentana and Steve Stern.

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 (2001), Brandt and Santa-Clara (2002), Danielsson (1994), or Gelfand and Smith (1990)).

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 (2001), 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 (1990) contains a related method. Brandt and Santa-Clara (1995) and Pedersen (1995) 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. Our setting incorporates the look-ahead estimator and useful extensions. We prove global consistency and a functional central limit theorem under minimal assumptions. From the functional central limit theorem it follows that the  $L_2$  deviation between the true and computed densities is  $O_P(n^{-1/2})$ , independent of the dimension of the state space.

To reiterate the contributions of this paper, we:

1. extend the look-ahead estimator to incorporate a significantly wider range of applications, and
2. provide a more complete asymptotic theory, with a focus on minimal assumptions and global (i.e., norm) deviation between the estimator and the target density.

In the process, we connect the concept of the look-ahead estimator to the more general problem of conditional Monte Carlo density estimation.

One way to phrase the nature of the extension in point (i) is as follows: 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.<sup>1</sup>

## 2 Definitions

Let  $(\mathbb{Y}, \mathcal{Y}, \mu)$  be a  $\sigma$ -finite measure space, and let  $(\Omega, \mathcal{F}, \mathbf{P})$  be a probability space. 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  $g: \mathbb{Y} \rightarrow \mathbb{R}$  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, either  $\mathbb{Y} \subset \mathbb{R}^k$ ,  $\mathcal{Y}$  is the Borel sets, and  $\mu$  is Lebesgue measure, or  $\mathbb{Y}$  is countable,  $\mathcal{Y}$  is the set of all subsets, and  $\mu$  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  $\mathbb{Y}$ .<sup>2</sup> 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,<sup>3</sup> 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

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

---

<sup>1</sup>See sections 4 and 5 for more discussion of this example. Note also that Henderson and Glynn (2001) consider generalizations of the basic look-ahead idea, but on a less systematic level than the treatment here.

<sup>2</sup>As usual, functions equal  $\mu$ -almost everywhere are identified.

<sup>3</sup> $\mathcal{Y}$  is called countably generated if there exists a countable family  $\mathcal{A}$  of subsets of  $\mathbb{Y}$  such that  $\mathcal{A}$  generates  $\mathcal{Y}$ .

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  $L_p(\mu)$ .<sup>4</sup> The expectation  $\mathcal{E}F$  of  $F$  is defined as the unique element of  $L_p(\mu)$  such that

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

where  $\mathbf{E}$  is the usual scalar expectation. If  $\mathbf{E}\|F\|_p$  is finite, then  $\mathcal{E}F$  exists.  $\mathcal{E}F$  is also called the Bochner-Pettis integral of  $F$ .<sup>5</sup> 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  $P: \mathbb{X} \times \mathcal{X} \rightarrow [0, 1]$  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

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

A probability measure  $\phi$  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  $\phi$  is stationary,  $(X_t)_{t \geq 0}$  is  $P$ -Markov and  $\mathcal{L}X_0 = \phi$ , then  $(X_t)_{t \geq 0}$  is itself (strict sense) stationary.

$P$  is called *ergodic* if it has a unique stationary distribution  $\phi$ , and  $P^t(x, \cdot)$  converges to  $\phi$  in total variation norm for every  $x \in \mathbb{X}$ . If this case, for every  $P$ -Markov process  $(X_t)_{t \geq 0}$  and every measurable  $h: \mathbb{X} \rightarrow \mathbb{R}$  with  $\int |h| d\phi < \infty$ , we have

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

$P$  is called *V-uniformly ergodic* if, in addition, there exist a measurable function  $V: \mathbb{X} \mapsto [1, \infty)$  and nonnegative constants  $\lambda < 1$  and  $L < \infty$  satisfying

$$\sup_{|h| \leq V} \left| \int h(x') P^t(x, dx') - \int h(x') \phi(dx') \right| \leq \lambda^t L V(x) \quad \text{for all } x \in \mathbb{X}, t \in \mathbb{N}$$

---

<sup>4</sup>Measurability requires that  $F^{-1}(B) \in \mathcal{F}$  for every Borel subset  $B$  of  $L_p(\mu)$ . By the Pettis measurability theorem, if  $L_p(\mu)$  is separable, then a sufficient condition is Borel measurability of  $\Omega \ni \omega \mapsto \langle F(\omega), h \rangle \in \mathbb{R}$  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.

<sup>5</sup>For more details, see, e.g., Bosq (2000).

(If  $V$  can be chosen identically equal to 1, then the left-hand side becomes the total variation distance between  $P^t(x, dx')$  and  $\phi$ , while the right hand side is independent of  $x$ . This is the uniformly ergodic case.) Under the  $V$ -uniform ergodicity assumption, the central limit theorem can be established for a broad class of functions.  $V$ -uniform ergodicity has been shown to hold in a range of applications in operations research, finance, economics and time series analysis.<sup>6</sup>

### 3 Methodology

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

$$\psi(y) = \int q(y | x) \phi(dx) \quad \text{for all } y \in \mathbb{Y} \quad (2)$$

In order to simulate the target density  $\psi$ , we assume the existence of a decomposition (2) such that

1. The conditional density  $q$  can be evaluated, at least numerically.
2. There exists a stochastic kernel  $P$  on  $(\mathbb{X}, \mathcal{X})$  such that  $\phi$  is the unique stationary distribution of  $P$ .
3. 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  $\psi$  as

$$\psi_n(y) = \frac{1}{n} \sum_{t=1}^n q(y | X_t) \quad \text{where } (X_t)_{t=1}^n \text{ is } P\text{-Markov} \quad (3)$$

Examples are presented below. The simplest case is where direct IID sampling from  $\phi$  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  $\psi_n$  as in (3). The estimator (3) is very natural in this setting because we then have

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

---

<sup>6</sup>See, e.g., Meyn and Tweedie (2009), Kamihigashi (2007), Kristensen (2008) or Nishimura and Stachurski (2005).

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  $\phi$  is infeasible. The Markov chain Monte Carlo solution is to construct a kernel  $P$  such that  $\phi$  is the stationary distribution of  $P$ , and then generate  $P$ -Markov time series. Inserting this series into (3) gives an implementation of the GLAE.

### 3.1 Examples

The GLAE in (3) generalizes the stationary density look-ahead estimator of Henderson and Glynn (2001).<sup>7</sup> 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  $P(x, B) = \int_B q(y|x) \mu(dy)$  for some conditional density  $q: \mathbb{Y} \times \mathbb{Y} \rightarrow \mathbb{R}_+$ . Suppose that a unique stationary distribution exists. In this setting, it is well-known that the stationary distribution can be represented by a density  $\psi$  on  $\mathbb{Y}$ , and, moreover, the density  $\psi$  satisfies

$$\psi(y) = \int q(y|x) \psi(x) \mu(dx) \quad \text{for all } y \in \mathbb{Y} \quad (4)$$

Suppose that  $q$  is tractable but  $\psi$  is not, and one wishes to compute  $\psi$ . Although there are several techniques for doing this, the stationary density look-ahead estimator of Henderson and Glynn (2001) is perhaps the most attractive. The look-ahead estimator of  $\psi$  is defined as  $\psi_n(y) = n^{-1} \sum_{t=1}^n q(y|X_t)$ , where  $(X_t)_{t=1}^n$  is a simulated  $P$ -Markov time series. This is a special case of (3). In particular, comparing (2) and (4), 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 the reduced form model

$$\begin{aligned} k_{t+1} &= h(k_t) z_{t+1} \\ z_{t+1} &= g(z_t) \xi_{t+1} \end{aligned}$$

where  $(\xi_t)$  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

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

---

<sup>7</sup>Properties of the stationary density look-ahead estimator were also investigated by Stachurski and Martin (2008).

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 (2001) 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 (4) 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  $(\xi_t)$ . 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 computation of the stationary density of the capital stock. Letting  $\psi$  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

$$\psi(k') = \int q(k' | k, z) d\phi(k, z) \quad (5)$$

where  $\phi$  is the stationary density of  $P$ , and  $q(k' | k, z)$  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

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

Since (5) is a special case of (2), we can apply the GLAE, simulating  $(k_1, z_1), \dots, (k_n, z_n)$  and then calculating

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

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

$$r_t = \sigma_t W_t \quad \text{where } (W_t) \stackrel{\text{iid}}{\sim} N(0, 1) \quad \text{and} \quad \sigma_{t+1}^2 = \alpha_0 + \beta \sigma_t^2 + \alpha_1 r_t^2 \quad (6)$$

Suppose we wish to compute the stationary density  $\psi$  of the returns process  $(r_t)_{t \geq 0}$ .<sup>8</sup> Let  $\phi$  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 stationary distribution and  $V$ -uniform ergodicity. See section 5 for details.) The look-ahead estimator cannot be applied to this problem, for reasons similar to the macroeconomic

---

<sup>8</sup>Researchers are interested in the stationary density of returns for a variety of reasons, including density forecasting, value at risk, exact likelihood estimation and model assessment.



model discussed above. However, we can use the GLAE as follows: Equation (6) implies that  $r_t = \sqrt{X_t}W_t$ , and hence the conditional density  $q(r|x)$  of  $r_t$  given  $X_t = x$  is centered Gaussian with variance  $x$ . For this  $q$  we have  $\psi(r) = \int q(r|x)\phi(dx)$ , which is a version of (2). The process  $(X_t)_{t \geq 0}$  can be expressed as

$$X_{t+1} = \alpha_0 + \beta X_t + \alpha_1 X_t W_t^2 \quad (7)$$

After simulating a time series  $(X_t)_{t=1}^n$  from this process, the conditional Monte Carlo estimator of  $\psi$  can be formed as

$$\psi_n(r) = \frac{1}{n} \sum_{t=1}^n q(r|X_t) = \frac{1}{n} \sum_{t=1}^n (2\pi X_t)^{-1/2} \exp \left\{ -\frac{r^2}{2X_t} \right\} \quad (8)$$

## 4 Results

In this section we provide a general asymptotic theory of the GLAE. To begin, notice that when  $\psi_n$  was defined in (3), the distribution of  $X_0$  was not specified. When it is possible to draw  $X_0$  from  $\phi$ , we have the following result:

**Lemma 4.1.** *If  $\mathcal{L}X_0 = \phi$ , then  $\psi_n$  is unbiased, in the sense that  $\mathcal{E}\psi_n = \psi$ .<sup>9</sup>*

In many applications, there is no obvious way to sample directly from  $\phi$ , and lemma 4.1 cannot be applied. However, with sufficient ergodicity,  $\psi_n$  is asymptotically unbiased for large  $t$ , and also consistent:

**Theorem 4.1.** *If  $P$  is ergodic, then*

1.  *$\psi_n$  is strongly globally consistent, in the sense that  $\psi_n \rightarrow \psi$  in  $L_1(\mu)$  as  $n \rightarrow \infty$  with probability one; and*
2.  *$\psi_n$  is asymptotically unbiased, in the sense that  $\mathcal{E}\psi_n \rightarrow \psi$  in  $L_1(\mu)$  as  $n \rightarrow \infty$ .*

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.)

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  $\psi_n \rightarrow \psi$  in  $L_1$ ,

---

<sup>9</sup>Here  $\mathcal{E}$  is the Bochner-Pettis expectation, as defined in section 2.

then the maximum deviation in probabilities over all events converges to zero.<sup>10</sup> 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.

For each  $x \in \mathbb{X}$ , let  $T(x)$  represent the function  $y \mapsto q(y|x) - \psi(y)$ , and define the linear operator  $C: L_2(\mu) \rightarrow L_2(\mu)$  by

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

for arbitrary  $h, g \in L_2(\mu)$ , where  $(X_t^*)_{t \geq 0}$  is stationary and  $P$ -Markov.<sup>11</sup> We can now state the following result:

**Theorem 4.2.** *If  $P$  is  $V$ -uniformly ergodic and the second moment condition*

$$\int q(y|x)^2 \mu(dy) \leq V(x) \quad \text{for all } x \in \mathbb{X} \quad (9)$$

*holds, then  $\sqrt{n}(\psi_n - \psi)$  converges in distribution to a centered Gaussian  $G$  on  $L_2(\mu)$  with covariance operator  $C$ .*<sup>12</sup>

One immediate comment on theorem 4.2 is that since  $h \mapsto \|h\|_2$  is continuous on  $L_2(\mu)$ , theorem 4.2 and the continuous mapping theorem imply that  $\|\psi_n - \psi\|_2 = O_P(n^{-1/2})$ . In other words,  $\psi_n$  is globally  $\sqrt{n}$ -consistent for  $\psi$  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  $\mathcal{L}X_0 = \phi$ . In many settings the stationary distribution  $\phi$  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

<sup>10</sup>For further discussion of the advantages of using  $L_1$  norm, see Devroye and Lugosi (2001).

<sup>11</sup>That is,  $(X_t^*)_{t \geq 0}$  is  $P$ -Markov and  $X_0^*$  is drawn from the stationary distribution  $\phi$ . That  $C$  is indeed a well-defined operator from  $L_2(\mu)$  to itself follows from the proof of theorem 4.2.

<sup>12</sup>A centered Gaussian  $G$  has 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  $G_n \rightarrow G_0$  in distribution if  $\mathbf{E}h(G_n) \rightarrow \mathbf{E}h(G)$  for every  $h \in \mathcal{C}$ . For more details and a related central limit theorem, see Chen and White (1998).

theorems typically assume stationarity of the process. For this reason, we provide a direct proof based on extending the asymptotic theory of  $V$ -uniformly ergodic Markov processes to  $L_p$ -valued functions of the process.

As a final remark, note that where IID sampling from  $\phi$  is possible, the conclusions of theorem 4.1 hold without any conditions on  $q$  and  $\phi$ , and theorem 4.2 holds whenever

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

## 5 Discussion

As an example of how the theory applies, consider again the estimator (8) presented in section 3.1. Assume as before that  $\alpha_1 + \beta < 1$ . Using the sufficient conditions of Meyn and Tweedie (2009), the process  $(X_t)_{t \geq 0}$  defined in (7) can be shown to be  $V$ -uniformly ergodic on  $\mathbb{X} := [\alpha_0/(1-\beta), \infty)$  for  $V(x) = x + c$ , where  $c$  is any constant in  $[1, \infty)$ .<sup>13</sup> Regarding the moment condition (9) in theorem 4.2, we have

$$\int q(r|x)^2 dr = \int (2\pi x)^{-1} \exp\left\{-\frac{r^2}{x}\right\} dr = (4\pi x)^{-1/2} \leq \left\{\frac{4\pi\alpha_0}{1-\beta}\right\}^{-1/2}$$

Recall that  $V(x) = x + c$ , where  $c$  can be chosen arbitrarily large. For large enough  $c$ , then, we have  $\int q(r|x)^2 dr \leq c \leq x + c = V(x)$ , and (9) is satisfied. As a result, both theorems 4.1 and 4.2 apply.

Before continuing, let us make a brief comparison of (3) 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  $Y_1, \dots, Y_n$  from  $\psi$ . The NPKDE  $f_n$  is then defined in terms of a kernel (i.e., density)  $K$  on  $\mathbb{Y}$  and a “bandwidth” parameter  $\delta_n$ :

$$f_n(y) := \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{y - Y_i}{\delta_n}\right) \quad (10)$$

The estimate  $f_n$  is known to be consistent, in the sense that  $\mathbb{E}\|f_n - \psi\|_1 \rightarrow 0$  whenever  $\delta_n \rightarrow 0$  and  $n\delta_n^k \rightarrow \infty$  (Devroye and Lugosi, 2001). 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  $\psi$  to be twice differentiable, then, for suitable choice of  $K$ , it can be shown that

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

---

<sup>13</sup>Details are omitted, since the  $V$ -uniform ergodicity result is known. For example, Kristensen (2008) establishes  $V$ -uniform ergodicity of a larger class of GARCH models.

Thus, even with this smoothness assumption on  $\psi$ —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  $\psi_n$ . Moreover, the rate of convergence slows as the dimension  $k$  of  $\mathbb{Y}$  increases.<sup>14</sup>

## 5.1 Simulation Results

Consider the GARCH application in section 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 the  $\psi_n$  implied by theorem 4.2 is illustrated in figure 1. The left panel of the figure contains the true density  $\psi$ , 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).<sup>15</sup> The right panel of figure 1 repeats the exercise, but this time using the GLAE in (8) rather than the NPKDE.

The estimator (8) 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  $\psi$ . We computed average  $L_1$  deviations over 1000 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 41% larger than that of the GLAE.

This simulation exercise considered a one-dimensional problem. The stronger 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.

---

<sup>14</sup>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 (3) 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.

<sup>15</sup>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 realization, or different methods of simulation.

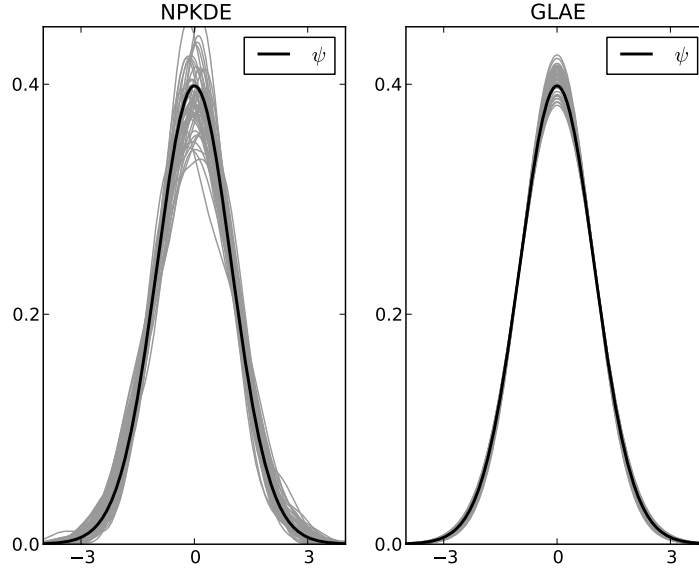


Figure 1: Relative performance,  $n = 500$

## 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  $L_2(\mu)$ .)

Also, to simplify the presentation, we will use the well-known operator notation

$$(\nu P)(B) := \int P(x, B) \nu(dx) \quad \text{and} \quad (Ph)(x) := \int h(y) P(x, dy)$$

Here  $P$  is a stochastic kernel on  $\mathbb{X}$ ,  $\nu$  is a probability measure on  $(\mathbb{X}, \mathcal{X})$  and  $h: \mathbb{X} \rightarrow \mathbb{R}$  is a measurable function such that the integral is defined. We also use  $\nu(h)$  to indicate the integral  $\int h d\nu$ . Given these definitions, recall that (c.f., e.g., Meyn and Tweedie, 2009, chapter 3), for each  $k \in \mathbb{N}$ , we have

$$\mathbb{E}[h(X_{t+k}) \mid \mathcal{F}_t] = P^k h(X_t) \tag{11}$$

where  $P^k$  denote the  $k$ -th iterate of the operator  $P$ , and  $\mathcal{F}_t$  is the  $\sigma$ -algebra generated by  $X_1, \dots, X_t$ .

Consider the setting of section 3. We begin with the following lemma:

**Lemma 6.1.** *If  $\mathcal{L}X = \phi$ , then  $\mathcal{E}q(\cdot | X) = \psi$  in both  $L_1(\mu)$  and  $L_2(\mu)$ .*

*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  $\|q(\cdot | x)\| = 1$  for all  $x \in \mathbb{X}$ . As a result,  $\mathbf{E}\|q(\cdot | X)\| = 1 < \infty$ , and the Bochner-Pettis expectation  $\mathcal{E}q(\cdot | X)$  is well-defined. To show that  $\mathcal{E}q(\cdot | X) = \psi$ , we must prove in addition that  $\mathbf{E}\langle q(\cdot | X), h \rangle = \langle \psi, h \rangle$  for all  $h \in L_\infty(\mu)$ . Fixing  $h \in L_\infty(\mu)$ , Fubini's theorem and (2) yield

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

By (2) this equals  $\int \psi h d\mu = \langle \psi, h \rangle$ , as was to be shown.

For the proof of the  $L_2(\mu)$  case, let  $\|\cdot\|$  be the  $L_2(\mu)$  norm. Since other parts of the proof are almost identical to the  $L_1(\mu)$  case, we verify only that  $\mathbf{E}\|q(\cdot | X)\| < \infty$ , which is necessary to ensure that the expectation  $\mathcal{E}q(\cdot | X)$  is well-defined in  $L_2(\mu)$ . For this, it suffices to show that

$$\mathbf{E}\|q(\cdot | X)\|^2 = \mathbf{E} \int q(y | X)^2 \mu(dy)$$

is finite. In view of (9), this quantity is bounded above by  $\mathbf{E}V(X)$ . Finiteness of  $\mathbf{E}V(X)$  when  $\mathcal{L}X = \phi$  is implicit from the definition of  $V$ -uniform ergodicity.  $\square$

*Proof of lemma 4.1.* Assume the conditions of the lemma. Since  $\phi$  is stationary for  $P$  and  $\mathcal{L}X_0 = \phi$ , we have  $\mathcal{L}X_t = \phi$  for all  $t \geq 0$ . From linearity of  $\mathcal{E}$  and lemma 6.1, we conclude that

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

In other words,  $\psi_n$  is unbiased, as was to be shown.  $\square$

## 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 2 of the theorem (asymptotic unbiasedness) follows from part 1 (strong consistency). Indeed, suppose that  $\psi_n \rightarrow \psi$  almost surely in  $L_1(\mu)$ . Using standard properties of the Bochner-Pettis integral, we obtain

$$\|\mathcal{E}\psi_n - \psi\| = \|\mathcal{E}\psi_n - \mathcal{E}\psi\| \leq \mathbf{E}\|\psi_n - \psi\|$$

Since  $\|f - g\| \leq 2$  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  $\psi_n \rightarrow \psi$  almost surely in  $L_1(\mu)$ . As in the statement of the theorem, let  $P$  be an ergodic stochastic kernel on  $(\mathbb{X}, \mathcal{X})$  with stationary distribution  $\phi$ . Let  $(X_t)_{t \geq 0}$  be  $P$ -Markov and let  $\mathcal{L}X^* = \phi$ . Define  $T(x) := q(\cdot | x) - \psi$ , which is a measurable function from  $\mathbb{X}$  to  $L_1(\mu)$  (see footnote 4). Note that  $\mathcal{E}T(X^*) = 0$  by lemma 6.1.

We need to show that

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

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  $\epsilon$ . 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 (2000, pp. 27–28):

$$\exists J \in \mathbb{N} \text{ with } \mathbf{E}\|T(X^*) - L_J T(X^*)\| < 2\epsilon \quad (13)$$

Our first claim is that

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

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

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

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

Returning to (12), we have

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

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

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{t=1}^n T(X_t) \right\| \leq \mathbf{E}\|T(X^*) - L_J T(X^*)\| + \|\mathcal{E}L_J T(X^*)\|$$

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

$$\|\mathcal{E}L_J T(X^*)\| = \|\mathcal{E}T(X^*) - \mathcal{E}L_J T(X^*)\| \leq \mathbf{E}\|T(X^*) - L_J T(X^*)\|$$

In view of (13) we then have

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{t=1}^n T(X_t) \right\| \leq 4\epsilon \quad (\mathbf{P}\text{-almost surely})$$

Since  $\epsilon$  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  $\|\cdot\|$  always represents the  $L_2(\mu)$  norm. Throughout the proof, for  $x \in \mathbb{X}$  we let  $T_0(x)$  be the function  $y \mapsto q(y|x)$ , and  $T(x)$  be the function  $y \mapsto q(y|x) - \psi(y)$ . In this notation, theorem 4.2 amounts to the claim that

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

where  $C$  is the operator defined in section 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  $X'_1 \equiv x \in \mathbb{X}$ . For any Borel probability measure  $\mu$  on  $L_2(\mu)$ ,*

$$\mathcal{L} \left[ n^{-1/2} \sum_{t=1}^n T(X_t) \right] \rightarrow \mu \quad \text{implies} \quad \mathcal{L} \left[ n^{-1/2} \sum_{t=1}^n T(X'_t) \right] \rightarrow \mu$$

*Proof.* As is well known (see, for example, Roberts and Rosenthal, 2004), one can construct copies of  $(X_t)_{t \geq 1}$  and  $(X'_t)_{t \geq 1}$  on a common probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  such that if  $X_t = X'_t$  for some  $t$ , then  $X_k = X'_k$  for all  $k \geq t$ , and  $\mathbf{P}\{\tau < \infty\} = 1$ , where  $\tau$  is the stopping time (coupling time)

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

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  $n^{-1/2}S_n \rightarrow \mu$ . To prove that  $n^{-1/2}S'_n \rightarrow \mu$  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, 2002, lemma 11.9.4). Fixing  $\epsilon > 0$ , we need to show that

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

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

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

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

For any fixed  $k$ , we have

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

Hence

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

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

In view of lemma 6.2, we can continue the proof of (15) while considering only the case  $\mathcal{L}X_1 = \phi$ . Another result we will find useful to establish (15) is given in the next lemma, and is an easy corollary of Bosq (2000, Theorem 2.3). The proof is omitted.

**Lemma 6.3.** *Let  $(Y_n)_{n \geq 1}$  and  $V$  be  $L_2(\mu)$  valued random variables. The following statements are equivalent:*

1.  $(\mathcal{L}Y_n)_{n \geq 1}$  is tight and  $\mathcal{L}\langle e, Y_n \rangle \rightarrow \mathcal{L}\langle e, V \rangle$  as  $n \rightarrow \infty$  for every  $e \in L_2(\mu)$  with  $\|e\| \leq 1$ ; and
2.  $\mathcal{L}Y_n \rightarrow \mathcal{L}V$  as  $n \rightarrow \infty$ .

Taking this result as given, and taking  $\mathcal{L}X_1 = \phi$ , define  $S_n := T(X_1) + \dots + T(X_n)$ , and  $Y_n := n^{-1/2}S_n$ . The strategy of the proof is to establish (i) of Lemma 6.3 for  $V \sim N(0, C)$ . As a first step, let us consider tightness of the partial sums.

**Lemma 6.4.** *Under the hypotheses of theorem 4.2, the sequence  $(\mathcal{L}Y_n)_{n \geq 1}$  is tight.*

*Proof.* As a first step, observe that  $\mathbf{E}\|T(X_1)\|^2 < \infty$ , because

$$\mathbf{E}\|T(X_1)\|^2 = \mathbf{E}\|q(\cdot | X_1) - \psi\|^2 \leq 2(\mathbf{E}\|q(\cdot | X_1)\|^2 + \|\psi\|^2)$$

and  $\mathbf{E}\|q(\cdot | X_1)\|^2 < \infty$  as shown in lemma 6.1. Since  $\mathbf{E}\|T(X_1)\|^2$  is finite, the covariance operator of  $T(X_1)$  exists, and we denote it by  $D$ . The operator  $D$  admits a decomposition of the form

$$Dh = \sum_{j \geq 1} \lambda_j \langle v_j, h \rangle v_j \quad (h \in L_2(\mu)) \quad (18)$$

where  $(v_j)_{j \geq 1}$  is an orthonormal basis for  $L_2(\mu)$ , and  $(\lambda_j)_{j \geq 1}$  is a sequence satisfying  $\sum_{j \geq 1} \lambda_j = \mathbf{E}\|T(X_1)\|^2 < \infty$  (c.f., e.g., Bosq, 2000, Chapter 1).<sup>16</sup> It follows from (18) that  $\mathbf{E}\langle v_j, T(X_1) \rangle^2 = \lambda_j$  for all  $j$ .

Pick any  $j \in \mathbb{N}$ . Our first claim is that if for some constant  $\gamma$  independent of  $j$  and  $n$  one has  $\mathbf{E}\langle v_j, Y_n \rangle^2 \leq \gamma \lambda_j$ , then  $(Y_n)_{n \geq 1}$  is tight. To see this, fix  $\epsilon > 0$  and consider the real sequence  $(a_N)$  defined by

$$a_N := \sum_{j \geq N} \mathbf{E}\langle v_j, Y_n \rangle^2 \leq \gamma \sum_{j \geq N} \lambda_j$$

Since  $a_N \downarrow 0$ , we can choose an increasing sequence of integers  $(N_k)_{k \geq 1}$  and a positive real sequence  $(b_k)_{k \geq 1}$  with  $b_k \uparrow \infty$  and  $\sum_{k \geq 1} b_k a_{N_k} < \epsilon$ . Let

$$B_k := \left\{ h \in L_2(\mu) : \sum_{j \geq N_k} \langle v_j, h \rangle^2 \leq b_k^{-1} \right\}$$

<sup>16</sup>As usual, an orthonormal basis of  $L_2(\mu)$  is an orthonormal set the span of which is dense in  $L_2(\mu)$ .

and let  $K$  be the compact set  $\cap_{k \geq 1} B_k$ .<sup>17</sup> Using the Chebychev Inequality now gives

$$\mathbf{P}\{Y_n \notin K\} \leq \sum_{k \geq 1} \mathbf{P}\{Y_n \notin B_k\} \leq \sum_{k \geq 1} b_k a_{N_k} < \epsilon, \quad \forall n \geq 1$$

It follows that  $(\mathcal{L}Y_n)_{n \geq 1}$  is tight.

It remains to prove that there exists a constant  $\gamma$  independent of  $j$  and  $n$  with  $\mathbf{E}\langle v_j, Y_n \rangle^2 \leq \gamma \lambda_j$ . So fix  $j$  and  $n$ , and observe that

$$\begin{aligned} \mathbf{E}\langle v_j, Y_n \rangle^2 &= \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n \mathbf{E}\langle v_j, T(X_s) \rangle \langle v_j, T(X_t) \rangle \leq \mathbf{E}\langle v_j, T(X_1) \rangle^2 + 2 \sum_{t=2}^n \mathbf{E}\langle v_j, T(X_1) \rangle \langle v_j, T(X_t) \rangle \\ \therefore \quad \mathbf{E}\langle v_j, Y_n \rangle^2 &\leq \lambda_j + 2 \sum_{t=2}^n \mathbf{E}\langle v_j, T(X_1) \rangle \langle v_j, T(X_t) \rangle \end{aligned} \quad (19)$$

Let  $\kappa$  be defined by  $\kappa(x) = \langle v_j, T(x) \rangle$ . Observe that  $\mathbf{E}\kappa(X_1) = 0$  and

$$|\kappa(x)| = |\langle v_j, T(x) \rangle| \leq \|v_j\| \cdot \|T(x)\| \leq V(x)^{1/2}$$

By Meyn and Tweedie (1993, Theorem 15.2.9), the Markov chain  $(X_t)_{t \geq 1}$  is also  $V^{1/2}$ -uniformly ergodic, and since  $|\kappa| \leq V^{1/2}$  we have

$$|P^n \kappa(x)| = |P^n \kappa(x) - \phi(\kappa)| \leq \alpha^t c V(x)^{1/2} \quad (20)$$

for some constant  $c$  and some  $\alpha < 1$  (Meyn and Tweedie, 1993, Theorem 16.1.2).

Now consider the term

$$\mathbf{E}\langle v_j, T(X_1) \rangle \langle v_j, T(X_t) \rangle = \mathbf{E}\kappa(X_1)\kappa(X_t) = \mathbf{E}\{\kappa(X_1)\mathbf{E}[\kappa(X_t) | \mathcal{F}_1]\}$$

In view of (11) we can write this as

$$\mathbf{E}\langle v_j, T(X_1) \rangle \langle v_j, T(X_t) \rangle = \mathbf{E}[\kappa(X_1)P^{t-1}\kappa(X_1)]$$

and the Cauchy-Schwartz inequality combined with (20) gives

$$\mathbf{E}\langle v_j, T(X_1) \rangle \langle v_j, T(X_t) \rangle \leq \mathbf{E}[\kappa(X_1)^2] \mathbf{E}[(P^{t-1}\kappa(X_1))^2] \leq \mathbf{E}[\kappa(X_1)^2] \mathbf{E}[(\alpha^{t-1}c)^2 V(X_1)]$$

Since  $\mathbf{E}[\kappa(X_1)^2] = \lambda_j$  we have

$$\mathbf{E}\langle v_j, T(X_1) \rangle \langle v_j, T(X_t) \rangle \leq \lambda_j (\alpha^{t-1}c)^2 \phi(V)$$

---

<sup>17</sup>Let  $C$  be any subset of  $L_2(\mu)$ , and let  $(e_k)_{k \geq 1}$  be an orthonormal basis. It is known that  $C$  is compact if  $\sum_{k \geq n} \langle e_k, h \rangle \rightarrow 0$  as  $n \rightarrow \infty$  uniformly over  $h \in C$ .

Returning to (19), then,

$$\begin{aligned}\mathbf{E}\langle v_j, Y_n \rangle^2 &\leq \lambda_j + 2 \sum_{t=2}^n \lambda_j (\alpha^{t-1} c)^2 \phi(V) \\ \therefore \quad \mathbf{E}\langle v_j, Y_n \rangle^2 &\leq \gamma \lambda_j, \quad \gamma := 1 + \frac{2c^2 \phi(V)}{1 - \alpha^2}\end{aligned}$$

As this was the last claim to be proven, the tightness of  $(\mathcal{L}Y_n)_{n \geq 1}$  is established.  $\square$

The next step in the proof of Theorem 4.2 is to verify the second part of (i) in Lemma 6.3.

**Lemma 6.5.** *Let  $(Y_n)_{n \geq 1}$  be as above, and let  $V$  be a Gaussian random variable on  $L_2(\mu)$  with distribution  $N(0, C)$ . Given any  $e \in L_2(\mu)$  with  $\|e\| \leq 1$ , we have  $\mathcal{L}\langle e, Y_n \rangle \rightarrow \mathcal{L}\langle e, V \rangle$  as  $n \rightarrow \infty$ .*

*Proof.* Fix such an  $e$ , let  $\kappa_0(x) := \langle e, T_0(x) \rangle$  and  $\kappa(x) := \langle e, T(x) \rangle$ . We have

$$\langle e, Y_n \rangle = n^{-1/2} \sum_{t=1}^n \langle e, T(X_t) \rangle = n^{-1/2} \sum_{t=1}^n \kappa(X_t)$$

Observe that  $\kappa(x) = \kappa_0(x) - \phi(\kappa_0)$ , and  $\kappa_0(x)^2 \leq \|T_0(x)\|^2 \leq V(x)$  for all  $x \in \mathbb{X}$ . In view of the scalar CLT for  $V$ -uniformly ergodic Markov chains (Meyn and Tweedie, 1993, Theorem 17.0.1) we have

$$\mathcal{L} \left[ n^{-1/2} \sum_{t=1}^n \kappa(X_t) \right] \rightarrow N(0, \sigma^2)$$

where

$$\begin{aligned}\sigma^2 &:= \mathbf{E}\kappa(X_1)^2 + 2 \sum_{t \geq 2} \mathbf{E}\kappa(X_1)\kappa(X_t) \\ &= \mathbf{E}\langle e, T(X_1) \rangle^2 + 2 \sum_{t \geq 2} \mathbf{E}\langle e, T(X_1) \rangle \langle e, T(X_t) \rangle = \langle e, Ce \rangle\end{aligned}$$

In other words,  $\mathcal{L}\langle e, Y_n \rangle \rightarrow \mathcal{L}\langle e, V \rangle$ , where  $V \sim N(0, C)$ .  $\square$

The result in Theorem 4.2 follows from lemmas 6.2, 6.3, 6.4 and 6.5.

## References

- [1] Bosq, Denis (2000): *Linear Processes in Function Space*, Springer-Verlag, New York.
- [2] Brandt, Michel W. and Pedro Santa-Clara (2002): "Simulated Likelihood Estimation Of Diffusions With An Application To Exchange Rate Dynamics In Incomplete Markets," *Journal of Financial Economics*, 63 (2), 161–210.
- [3] Chen, Xiaohong and Halbert White (1998): "Central Limit and Functional Central Limit Theorems for Hilbert-Valued Dependent Heterogeneous Arrays with Applications," *Econometric Theory*, 14, 260–284.
- [4] Danielsson, Jon (1994): "Stochastic Volatility in Asset Prices Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 64 (1-2), 375–400.
- [5] Devroye, Luc and Gabor Lugosi (2001): "Combinatorial Methods in Density Estimation" Springer-Verlag, New York.
- [6] Dudley, Richard M. (2002): *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics No. 74, Cambridge University Press.
- [7] Gelfand, Alan E. and Adrian F.M. Smith (1990): "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association*, Vol. 85, 398-409.
- [8] Henderson, Shane G. and Peter W. Glynn (2001): "Computing densities for Markov chains via simulation," *Mathematics of Operations Research* Vol. 26, 375-400.
- [9] Kamihigashi, Takashi (2007): "Stochastic optimal growth with bounded or unbounded utility and with bounded or unbounded shocks," *Journal of Mathematical Economics*, 43 (3-4), 477-500
- [10] Kristensen, Dennis (2008): "Uniform Ergodicity of a Class of Markov Chains with Applications to Time Series Models," mimeo, Columbia University.
- [11] Meyn, Sean and Richard L. Tweedie (2009): *Markov Chains and Stochastic Stability*, 2nd Edition, Cambridge University Press.
- [12] Nishimura, Kazuo and John Stachurski (2005): "Stability of Stochastic Optimal Growth Models: A New Approach," *Journal of Economic Theory*, 122 (1), 100–118.
- [13] Pedersen, Asger R. (1995): "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55–71.

- [14] Roberts, Gareth O. and Jeffrey S. Rosenthal (2004): "General State Markov Chains MCMC Algorithms," *Probability Surveys*, 1, 20–71.
- [15] Stachurski, John and Vance Martin (2008): "Computing the Distributions of Economic Models via Simulation," *Econometrica*, Vol. 76(2), 443–450.
- [16] Stachurski, John (2009): "A Hilbert Space Central Limit Theorem for Geometrically Ergodic Markov Chains," mimeo, Kyoto University.