CIRJE-F-620

Computing Densities and Expectations in Stochastic Recursive Economies: Generalized Look-Ahead Techniques

Richard Anton Braun
University of Tokyo

Huiyu Li
Graduate School of Economics, University of Tokyo

John Stachurski
Kyoto University

May 2009

Computing Densities and Expectations in Stochastic Recursive Economies: Generalized Look-Ahead Techniques

Richard Anton Braun
[*]University of Tokyo

Huiyu Li
[†]University of Tokyo

John Stachurski
[‡]Kyoto University

May 20, 2009

Abstract

We propose a generalized look-ahead estimator for computing densities and expectations in economic models. We provide conditions under which the estimator converges globally with probability one, and exhibit the asymptotic distribution of the error. Our estimator is more efficient than other Monte Carlo based approaches. Numerical experiments indicate that the estimator can provide large increases in accuracy and speed relative to traditional methods. Particular applications we consider are the stochastic growth model and an income fluctuation problem.

Keywords: Distributions, numerical methods, simulation
JEL Classification Codes: C61, C63

[*]Faculty of Economics, The University of Tokyo email:toni@e.u-tokyo.ac.jp
[†]Faculty of Economics, The University of Tokyo email:tohuiyu@gmail.com
[‡]Institute of Economic Research, Kyoto University email: john@kier.kyoto-u.ac.jp

1 Introduction

This paper proposes a new strategy for simulating marginal distributions, stationary distributions, and expectations of functions of the state. The methods we propose are general and can be applied in any economic model that can be formulated as stochastic dynamic programming problem with Markovian shocks.

Our strategy for computing distributions involves simulating densities. Densities are more useful than cumulative distribution functions or probability measures, partly because they can be used in applications such as maximum likelihood estimation, and partly because computing distributions from densities is straightforward, while computing densities from (empirical) distributions is ill-posed.[1]

We treat the problem of computing a density as an *estimation* problem, and propose a simulation-based estimator using conditional Monte Carlo. We prove that our density estimator, referred to below as the Generalized Look Ahead Estimator (GLAE), converges with probability one to the true invariant density under some weak regularity conditions on the data generating process. We also characterize the asymptotic distribution of the error in density space. Finally, we provide asymptotic approximations to the true density that can be used to bound the approximation errors of any measurable set for a given length of the simulation.

The GLAE has a wide range of applications. It can be applied in situations where shocks have either finite support or continuous support and can be applied both when the dynamics are stationary and non-stationary. We provide two versions of our estimator. The Generalized Time-series Look Ahead Estimator (GTLAE) works well in environments where the stochastic process eventually settles into a stationary distribution. The Generalized Cross-Sectional Look Ahead Estimator (GCLAE) can be used to estimate densities for finite k-period ahead events. This estimator is thus well-suited for use in nonstationary environments.

[1] In essence, this is because the first operation involves integration, while the second requires differentiation.

GLAE has some attractive properties as compared to other approaches for computing densities. GLAE is accurate. We show that GLAE provides estimates of the density that are more accurate than other existing simulation-based approaches using the same data. GLAE is also fast. It produces estimates to a given precision in less time than other approaches.

We consider two substantive economic applications. First, we illustrate the properties of GTLAE and GCLAE to estimate the density of capital in the neoclassical growth model with continuous shocks. GTLAE is shown to be more accurate than a Gaussian kernel estimator, producing L_1 errors that are 5 to 10 percent lower when calculating the stationary density. The improvements are much larger when using the GCLAE to estimate finite k-step ahead densities and range from 68-75 percent.

Our second application is estimating the labor productivity-asset density for an income fluctuations problem. The particular problem we consider is that of a household that faces uninsured shocks to labor-productivity, is able to save at a constant real interest rate but is subject to borrowing constraints as in Aiyagari (1994). Following Aiyagari (1994), we discretize the shocks to labor-productivity using Tauchen's (1986) methodology and solve a discrete dynamic programming problem. We compare GTLAE with a Monte Carlo Estimator (MCE). In this application GTLAE produces L_1 errors that are from 32 - 49 percent lower when shocks to labor productivity are persistent. In the presence of transient shocks to labor productivity the accuracy gains are even larger. MCE requires fewer computations and is thus faster than GTLAE for a given sample size. However, once we control for accuracy we find that GTLAE is faster at computing a solution to a pre-specified level of accuracy than MCE and that the benefits of using GTLAE increase with the size of the state space.

We also compare the speed of GTLAE with two other common approaches for computing the labor productivity-asset density. The first approach uses Gaussian elimination (see e.g. Stachurski, 2009). The second approach iterates on the transition matrix (see e.g. Ljungqvist and Sargent, 2004). In our example GTLAE is as much as 600 times faster than Gaussian elimination and as much as 100 times faster than iterating on the transition matrix.

The benefits of GTLAE are particularly significant in settings with large numbers of state variables. Indeed, many researchers are now using spaces of such high dimension that Gaussian and iterative methods are computationally infeasible. (For discrete state spaces, computational time for both Gaussian elimination and iterative solutions is of order N^3, where N is the size of the state space.) We illustrate how this problem can arise and illustrate how our approach can be used to calculate the population density and to estimate it quickly and accurately.

The current solution to this problem is to use traditional Monte Carlo estimation. However, GTLAE is more efficient than Monte Carlo estimation and provides substantially greater accuracy per unit of CPU time than traditional Monte Carlo methods.

In many applications, the object of interest is an expectation of a function of the state, and not the complete density. A second and distinct contribution of the paper is that we develop look-ahead techniques for computing expectations of functions of the state variable. This is convenient because it can be applied in situations when proper densities do not exist due to stochastic singularities. We show that our estimator has lower asymptotic variance than the usual Monte Carlo estimator. And provide examples that illustrate the magnitude of the efficiency gains in practical applications.

Our research is most closely related to Stachurski and Martin (2008) and Henderson and Glynn (2001). They provide asymptotic results for the look-ahead estimator (LAE). They assume that the conditional distribution of the future state given the current state can be represented by a density. This assumption rules out applying their results to many interesting dynamic programming applications, which have the property that the conditional distribution is degenerate along one or more dimensions.[2] Our methods are tailored to deal directly with this situation. Second this previous research limits attention to the case of Lebesgue measure. We

[2]To illustrate how this problem arises, consider for example the Brock-Mirman (1972) model with serially correlated shocks. In this setting, the equilibrium law of motion is a bivariate process in capital and the productivity shock. This law of motion lacks full stochastic rank because the number of state variables exceeds the number of innovations (i.e., one).

consider a general measure here that allows shocks to have either discrete or continuous support.

Our method for calculating expectations extends similar ideas of conditional Monte Carlo and Rao-Blackwellization developed by McKeague and Wefelmeyer (2000). In particular, we relax two assumptions made in that work (uniform ergodicity and the detailed balance condition) which rarely hold in economic applications.

The remainder of the paper is organized as follows. In Section 2 we propose the GTLAE estimator and derive its asymptotic properties. Then in Section 3 we report results for two applications of this estimator. In Section 4 we propose a look-ahead approach to computing expectations of functions of the state and describe applications of this approach. In Section 5 we propose the GCLAE estimator and describe an application of this estimator. Section 6 contains our concluding remarks.

2 Computing Stationary Densities

Consider the following scenario. Suppose that the laws of motion for an economic system can be formulated as

$$X_{t+1} = F(X_t, \xi_{t+1}), \qquad (\xi_t)_{t\geq 1} \overset{\text{IID}}{\sim} \phi \tag{1}$$

where X_t and ξ_t take values in spaces $\mathbb{X}$ and $\mathbb{Z}$ respectively. Typically, X_t is a vector of endogenous and exogenous state variables, while ξ_t is a vector-valued shock. In the interest of generality, we assume only that $(\mathbb{X}, \mathscr{X})$ is a measurable space where $\mathscr{X}$ is countably generated, and $(\mathbb{Z}, \mathscr{Z})$ is any measurable space.[3]

The process $(\xi_t)_{t\geq 1}$ is defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and assumed to be IID with common distribution ϕ. (Serially correlated shocks are incorporated in this setting by appending them to the state space—see

[3]The function $F\colon \mathbb{X} \times \mathbb{Z} \to \mathbb{X}$ is $(\mathscr{X} \otimes \mathscr{Z}, \mathscr{X})$ measurable. (In section 5, we allow F to depend explicitly on time in order to accommodate nonstationary environments.)

below.) As a result, $(X_t)_{t\geq 0}$ is Markovian, and its dynamics can also be represented in terms of the stochastic kernel

$$P(x, B) = \phi\{z \in \mathbb{Z} : F(x, z) \in B\} = \int \mathbb{1}_B[F(x, z)]\phi(dz) \tag{2}$$

which gives the probability $P(x, B)$ of moving from $x \in \mathbb{X}$ into set $B \in \mathscr{X}$ in one step. Here $\mathbb{1}_B$ is the indicator function of B. The sequence of marginal distributions $(\psi_t)_{t\geq 0}$ for $(X_t)_{t\geq 0}$ is known to satisfy

$$\psi_{t+1}(B) = \int P(x, B)\psi_t(dx) \qquad (B \in \mathscr{X}, t \geq 0) \tag{3}$$

Letting $\mathscr{P}(\mathbb{X})$ denote the set of probability measures on $(\mathbb{X}, \mathscr{X})$, a distribution $\psi^* \in \mathscr{P}(\mathbb{X})$ is called *stationary* for (1) if, for all $B \in \mathscr{X}$, we have

$$\psi^*(B) = \int P(x, B)\psi^*(dx) \tag{4}$$

In view of (3), if X_t has distribution ψ^*, then so does X_{t+k} for all $k \geq 0$. Throughout this section, we assume that ψ^* exists and is unique. In this case, our interest is in computing quantities related to ψ^*.

An ideal setting is where the conditional distribution $P(x, dy)$ can be represented by a density for each x. This is the full stochastic rank case. One can then show that ψ^* can likewise represented by a density, and efficient "look-ahead" methods for computing this density via simulation exist (Henderson and Glynn, 2001, Stachurski and Martin, 2008). In most economic environments, however, serial correlation and other factors mean that realistic models fall outside this ideal setting.

Problem 2.1. Consider an optimization problem of the form

$$\max_{\sigma\in\Sigma} \mathbb{E}\left[\sum_{t\geq 0} \beta^t v(k_t, \sigma(k_t, z_t), z_t)\right]$$

subject to $k_{t+1} = \sigma(k_t, z_t) \in \Gamma(k_t, z_t) \subset \mathbb{R}^m_+$ and $z_{t+1} = h(z_t, \epsilon_{t+1}) \in \mathbb{R}^\ell_+$, where Σ is the set of feasible policy functions[4] and $(\epsilon_t)_{t\geq 1}$ is an IID sequence. Assume the existence of an optimal policy σ. Letting $X_t :=$

[4]That is, σ is measurable and $\sigma(k, z) \in \Gamma(k, z)$ for all state pairs (k, z).

$(k_t, z_t) \in \mathbb{R}_+^{m \times \ell} =: \mathbb{X}$ and $\xi_t := \epsilon_t \in \mathbb{R}^j =: \mathbb{Z}$, the state evolves according to

$$X_{t+1} := \begin{bmatrix} k_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} \sigma(k_t, z_t) \\ h(z_t, \epsilon_{t+1}) \end{bmatrix} =: F(X_t, \xi_{t+1}) \tag{5}$$

which is a special case of (1). This problem nests a large class of individual optimization problems as in, e.g., Guvenen and Smith (2006), firm problems as in, e.g., Khan and Thomas (2008), and recursive competitive equilibria as in Hansen and Prescott (1995). An important feature of the problem is that the transition rule in (5) typically exhibits deficient stochastic rank. The mapping $z \mapsto F(x, z)$ takes values in a measure-zero subset of $\mathbb{X}$.

However, it may be that our interest lies in only a subset of the state variables, or in some other variables depending on the state, such as prices or consumption. A setting which nests these problems is one where we seek to compute the density of a measurable function $Y_t = \tau(X_t)$ of the state. When the dimension of Y_t is not large relative to the dimension of the shock, we show that generalized look-ahead estimators for the density of Y_t can often be constructed, and that these estimators have excellent asymptotic properties.[5]

Let us give a more precise problem statement for this section. Let ψ^* be the unique stationary distribution of P. Let X^* be a random variable on $\mathbb{X}$ with distribution ψ^*, and let $Y^* := \tau(X^*)$, where τ is any measurable function from $(\mathbb{X}, \mathscr{X})$ to σ-finite measure space $(\mathbb{Y}, \mathscr{Y}, \mu)$.[6] We wish to compute the density f^* of Y^* on $\mathbb{Y}$ whenever it exists. In doing so, we cannot sample directly from the unknown distribution ψ^*, but we may generate sequences $(X_t)_{t \geq 0}$ from the model (1).

Here f^* is a density with respect to μ. That is, $\int_B f^* d\mu = \mathbb{P}\{Y^* \in B\}$ for all $B \in \mathscr{Y}$.[7] This formulation nests two common cases. First, it nests

[5] Even when the dimension condition is not satisfied and densities do not exist, we provide efficient look-ahead type estimators for computing probabilities and expectations. See section 4.

[6] In addition to σ-finiteness, we require $\mathscr{Y}$ is countably generated.

[7] A necessary and sufficient condition for existence of such an f^* is that the probability measure $B \mapsto \mathbb{P}\{Y^* \in B\}$ is absolutely continuous with respect to μ.

the standard case where $\mathbb{Y}$ is a subset of $\mathbb{R}^k$, $\mathscr{Y}$ is the Borel sets, and μ is the Lebesgue measure. The other case is when $\mathbb{Y}$ is discrete, $\mathscr{Y}$ is the set of all subsets, and μ is the counting measure. In this case, integration is summation, and densities with respect to μ are probability mass functions.

In order for the density approximation problem treated in this section to be well-defined, we need an assumption which guarantees that f^* exists. Our condition is stated in terms of the conditional probability

$$\begin{aligned}\mathbb{P}\{Y_{t+1} \in B \,|\, X_t = x\} &= \phi\{z \in \mathbb{Z} : \tau[F(x,z)] \in B\} \\ &= \phi\{z \in \mathbb{Z} : F(x,z) \in \tau^{-1}(B)\} = P(x, \tau^{-1}(B))\end{aligned}$$

Assumption 2.1. This conditional distribution can be represented by a density on $\mathbb{Y}$ with respect to the dominating measure μ. In particular, there exists a measurable[8] function $q\colon \mathbb{Y} \times \mathbb{X} \to \mathbb{R}_+$ satisfying

$$\int_B q(y \,|\, x)\mu(dy) = P(x, \tau^{-1}(B)) \qquad (x \in \mathbb{X},\ B \in \mathscr{Y}) \tag{6}$$

Note that in the discrete case, where μ is the counting measure, this assumption is trivially satisfied. Any distribution on $\mathbb{Y}$ can be represented by a discrete density (i.e., probability mass function).[9]

Given assumption 2.1, the density f^* exists. Indeed,

$$f^*(y) := \int q(y \,|\, x)\psi^*(dx) \qquad (y \in \mathbb{Y}) \tag{7}$$

is a density representing Y^* (i.e., $\int_B f^* d\mu = \mathbb{P}\{Y^* \in B\}$ for all $B \in \mathscr{Y}$). To see this, fix $B \in \mathscr{Y}$. By Fubini's theorem, (6) and (4),

$$\begin{aligned}\int_B \int q(y \,|\, x)\psi^*(dx)\mu(dy) &= \int \int_B q(y \,|\, x)\mu(dy)\psi^*(dx) \\ &= \int P(x, \tau^{-1}(B))\psi^*(dx) = \psi^*(\tau^{-1}(B))\end{aligned}$$

But $\psi^*(\tau^{-1}(B)) = \mathbb{P}\{X^* \in \tau^{-1}(B)\} = \mathbb{P}\{Y^* \in B\}$, as was to be shown.

[8] Measurability with respect to the product σ-field $\mathscr{Y} \otimes \mathscr{X}$.

[9] This is because $\mu(B) = 0$ implies that $B = \emptyset$, so all distributions on $(\mathbb{Y}, \mathscr{Y})$ are absolutely continuous with respect to μ.

In order to compute f^* from $(X_t)_{t\geq 0}$, we require that P is *ergodic*. In other words, for any $h\colon \mathbb{X} \to \mathbb{R}$ and any initial condition $X_0 = x_0 \in \mathbb{X}$ we have

$$\int |h(x)|\psi^*(dx) < \infty \text{ implies } \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} h(X_t) = \int h(x)\psi^*(dx) \tag{8}$$

$\mathbb{P}$-almost surely. To compute f^* we propose a generalized time-series look-ahead estimator (GTLAE) defined by

$$f_n^*(y) := \frac{1}{n}\sum_{t=1}^{n} q(y \mid X_t) \qquad (y \in \mathbb{Y}) \tag{9}$$

where $(X_1, \ldots, X_n)$ is a time-series simulated from (1), and the initial condition X_0 is either given or chosen arbitrarily.[10] Algorithm 1 demonstrates how to obtain an observation of $f_n^*(y)$ for fixed $y \in \mathbb{Y}$.

Algorithm 1: Time series look-ahead estimate of $f^*(y)$

$X_0 \leftarrow x$, where $x \in \mathbb{X}$ is arbitrary ;
for *t in 1 to n* **do**
  draw $\xi \sim \phi$;
  $X_t \leftarrow F(X_{t-1}, \xi)$;
end
return $f_n^*(y) := \frac{1}{n}\sum_{i=1}^{n} q(y \mid X_t)$

The pointwise properties of the GTLAE are easily described. In particular, if $y \in \mathbb{Y}$ is fixed, then

$$\lim_{n\to\infty} f_n^*(y) := \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} q(y \mid X_t) = \int q(y \mid x)\psi^*(dx) = f^*(y) \tag{10}$$

$\mathbb{P}$-almost surely as $n \to \infty$, where the first equality is by definition, the second is by (8), and the last is due to (7). In this sense we have consistency

[10] The GTLAE nests the standard look-ahead estimator discussed in Henderson and Glynn (2001) and Stachurski and Martin (2008). In particular, if $(\mathbb{Y}, \mathscr{Y})$ is a subset of $\mathbb{R}^k$ paired with its Borel sets, if μ is Lebesgue measure, and if τ is the identity map (so that $Y_t = X_t$), then $q(y \mid x)$ is just the conditional (Lebesgue measure) density of X_t given X_{t-1}. In this case, (9) reduces to the definition used in Stachurski and Martin (2008).

at each point in the domain $\mathbb{Y}$. However, this result does not generally imply that $f_n^* \to f$ pointwise with probability one, since (10) holds on the complement of a null set dependent on y, and uncountable families of null sets are not generally null. To obtain global convergence results we will switch to a function-space point of view, beginning in section 2.3.

2.1 An Example with Continuously Distributed Shocks

To illustrate the application of GLAE density estimates in a setting with continuous shocks consider the variant of the stochastic growth model of Nelson and Plosser (1982). This variant of the stochastic growth model is convenient because it admits a closed form solution.

Example 2.1. Consider the following special case of problem 2.1,

$$v(k', k, z) = \ln(k' - Ak^\alpha z) \quad \text{and} \quad \Gamma(k, z) = [0, Ak^\alpha z]$$

In this case, optimality implies that the capital stock evolves according to

$$k_{t+1} = A\beta\alpha k_t^\alpha z_{t+1}$$

We suppose further that the exogenous shock process (z_t) is determined by

$$\ln z_{t+1} = \rho \ln z_t + \epsilon_{t+1}, \quad (\epsilon_t)_{t \geq 1} \overset{\text{IID}}{\sim} N(0, \sigma^2)$$

We take (k_t, z_t) as the state, and seek to compute the distribution of $Y_t := \ln(k_t/\bar{k})$ when $\bar{k} := (A\beta\alpha)^{1/(1-\alpha)}$. Here Y_t takes values in $(\mathbb{Y}, \mathscr{Y}, \mu) = (\mathbb{R}, \mathscr{B}, \lambda)$, where $\mathscr{B}$ is the Borel subsets of $\mathbb{R}$ and λ is Lebesgue measure. The function τ is given by $\tau(k, z) = \ln(k/\bar{k})$. Some manipulations show that

$$Y_t = \ln k_t - \ln \bar{k} = \alpha(\ln k_{t-1} - \ln \bar{k}) + \rho \ln z_{t-1} + \epsilon_t$$

It follows that the conditional density of Y_t given $X_{t-1} = (k, z)$ is

$$q(y \mid k, z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y - \alpha(\ln k - \ln \bar{k}) - \rho \ln z)^2}{2\sigma^2} \right\} \tag{11}$$

Hence assumption 2.1 is satisfied. For this model, the GTLAE is formed by simulating a series $(k_t, z_t)_{t=1}^n$ and evaluating $n^{-1} \sum_{t=1}^n q(y \mid k_t, z_t)$.

The fact that the density of Y^* can be calculated analytically makes this example a useful test case for the numerical experiments conducted below.

2.2 An Example with Discrete Shocks

Now suppose that $\mathbb{X}$ is finite, and that we wish to compute the stationary distribution ψ^* (i.e, the distribution of X^*). Since we are interested in the state variable X^* rather than some function of X^*, τ is the identity mapping. In the present setting, $(\mathbb{Y}, \mathscr{Y}, \mu) = (\mathbb{X}, \mathscr{X}, c)$, where $\mathscr{X}$ is all subsets of $\mathbb{X}$, and c is the counting measure. Let p be the transition matrix corresponding to (1), so that $p(x,y) = \phi\{z \in Z : F(x,z) = y\}$ gives the probability of transitioning from $x \in \mathbb{X}$ to $y \in \mathbb{X}$ in one step. The conditional density $q(y|x)$ in assumption 2.1 is just $p(x,y)$, as $p(x,y)$ satisfies (6) for any $B \subset \mathbb{X}$ and any $x \in \mathbb{X}$:

$$\int_B p(x,y)c(dy) = \sum_{y \in B} p(x,y) = P(x,B) = P(x, \tau^{-1}(B))$$

In this finite setting, the distribution ψ^* can be represented as the solution to a system of linear equations,[11] and solved using Gaussian elimination or similar techniques. Such techniques are typically of order $O(n^3)$, where $n = \text{card}(\mathbb{X})$. As a result they are slow or infeasible for large n, and other techniques for generating approximate solutions must be employed.

The standard alternative is to compute empirical frequencies using Monte Carlo. Let $(X_t)_{t \geq 1}$ be a simulated time series generated from the model. Employing (8) with $h(x) = \mathbb{1}\{x = y\}$, we obtain

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{X_t = y\} \to \sum_{x \in \mathbb{X}} \mathbb{1}\{x = y\} \psi^*(x) = \psi^*(y) \qquad (n \to \infty) \tag{12}$$

Evaluating the expression on the left for each $y \in \mathbb{X}$ gives an approximation to ψ^*. However, estimates of low frequency events tend to be unstable. We propose instead to use the GTLAE, which in this case is

$$f_n^*(y) = \frac{1}{n} \sum_{t=1}^{n} q(y \mid X_t) = \frac{1}{n} \sum_{t=1}^{n} p(X_t, y) \qquad (y \in \mathbb{X}) \tag{13}$$

As shown in (10), this estimator is also consistent. Moreover, we will show below in Section 4 that its asymptotic variance is lower than that of the standard Monte Carlo estimator, independent of the model p.

[11] See, e.g., Stachurski (2009, section 4.3.1).

Example 2.2. As a particular illustration, consider the incomplete market economy of Aiyagari (1994). Following Aiyagari (1994), we begin with the consumption smoothing problem of an individual who insures against idiosyncratic earnings risk by saving at a risk free rate of r. In this model both the real interest rate and the wage rate w are constant but total earnings are random due to idiosyncratic variations in labor productivity. Each individual also faces a borrowing constraint that rules out uncollateralized borrowing.

The household's problem can be described by the Bellman equation

$$V(a,z) = \max_{c,a'} \left\{ u(c) + \beta \mathbb{E}[V(a',z') \mid z] \right\}$$

subject to

$$0 \leq a' \leq wz + a(1+r) - c$$

The serially correlated shock to labor productivity evolves according to

$$\ln z' = \rho \ln z + \sigma \sqrt{1-\rho^2}\epsilon, \quad \epsilon \sim N(0,1) \tag{14}$$

One common approach to solving this problem is to discretize the productivity process using Tauchen's method (1986), obtaining a grid $\{z_1, ..., z_M\}$ and an $M \times M$ stochastic matrix R that represents the dynamics (14) on the grid.[12] Assuming that assets also lie on a finite grid $\{a_1, ...a_L\}$, the solution to Bellman's equation is an $L \times M$ matrix of saving policies $a' = g(a,z)$.

Given the solution to the individual's problem, we are interested in making inferences about the stationary labor productivity-asset density. The path of assets and productivity shocks evolves according to $a_{t+1} = g(a_t, z_t)$ with $(z_t)_{t \geq 0}$ generated by R. Taking $X_t := (a_t, z_t)$ as the state variable, the transition probabilities from state $x = (a,z)$ to state $y = (a',z')$ are given by

$$p(x,y) :=: p((a,z),(a',z')) := \mathbb{1}\{g(a,z) = a'\} R(z,z') \tag{15}$$

In view of (13), the GTLAE of the stationary productivity-asset density is

$$f_n^*(a',z') = \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{g(a_t,z_t) = a'\} R(z_t,z') \tag{16}$$

[12] In particular, $\mathbb{P}\{z_{t+1} = z' \mid z_t = z\} = R(z,z')$ for any $z, z' \in \{z_1, \ldots, z_M\}$.

Implementation proceeds by drawing a sequence of n realizations of labor productivity from the matrix R, and using the policy function g to derive n associated values of assets. The resulting sequence $(a_t, z_t)_{t=1}^n$ is then inserted into (16), and this expression is evaluated at all $a' \in \{a_1, \ldots, a_L\}$ and $z' \in \{z_1, \ldots, z_M\}$.

2.3 Asymptotic Properties of the GTLAE

The convergence result in (10) is pointwise, pertaining to fixed $y \in \mathbb{Y}$. Our interest is in the global properties of the estimators, and generalizing from these local results to global results is not straightforward. Instead, we start from a global perspective, within the setting of random functions in Banach space. This technique leads to a relatively complete characterization of the global properties.

To begin, recall that P in (2) is called *V-uniformly ergodic* (V-UE) if it is ϕ-irreducible, aperiodic (see Meyn and Tweedie, 1993, for definitions) and, moreover, there is a function $V\colon \mathbb{X} \mapsto [1, \infty)$ such that the sublevel sets $C_{V,\alpha} := \{x \in \mathbb{X} : V(x) \leq \alpha\}$ are P-small for all real α, as well as nonnegative constants $\lambda < 1$ and $L < \infty$ satisfying

$$\int V(y)P(x, dy) = \int V[F(x, z)]\phi(dz) \leq \lambda V(x) + L, \quad x \in \mathbb{X} \tag{17}$$

Under the V-UE assumption, a unique stationary distribution ψ^* exists, and $\int V d\psi^* < \infty$.[13] Moreover, the ergodicity in (8) holds, as does the central limit theorem for a broad class of functions. The V-UE condition has been shown to hold in a range of economic and econometric applications.[14]

Since $(\mathbb{Y}, \mathscr{Y}, \mu)$ is σ-finite and $\mathscr{Y}$ is countably generated, it follows that the Banach spaces $L_p(\mathbb{Y}, \mathscr{Y}, \mu) =: L_p(\mu)$ of p-integrable real-valued functions

[13] See, in particular, Meyn and Tweedie (1993, lemma 15.2.8 and theorem 16.1.2). A set $C \in \mathscr{X}$ is P-small if there exists a nontrivial measure ν and $k \in \mathbb{N}$ such that $P^k(x, B) \geq \nu(B)$ for all $B \in \mathscr{X}$ and all $x \in C$.

[14] See, for example, Kristensen (2007) or Nishimura and Stachurski (2005).

on $\mathbb{Y}$ are separable for all $p \in [1, \infty)$. As usual, $L_p(\mu)$ is endowed with the norm

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

with $\|g\|_\infty$ being the essential supremum. We use $q \in (1, \infty]$ to denote the conjugate index of p (i.e., $1/p + 1/q = 1$), so that $L_q(\mu)$ is (isometrically isomorphic to) the norm dual of $L_p(\mu)$ for $1 \leq p < \infty$. For such p and q, let

$$\langle g, h \rangle := \int g h d\mu =: \int g(y) h(y) \mu(dy) \qquad (g \in L_p(\mu),\ h \in L_q(\mu))$$

If $p = q = 2$ then this inner product gives rise to the Hilbert space $L_2(\mu)$.

Let U be a random variable taking values in $L_p(\mu)$, and defined on probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The (vector-valued) *expectation* of U is the point $\mathcal{E}U \in L_p(\mu)$ such that

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

It follows from the Reisz representation theorem that if $\mathbb{E}\|U\|_p$ is finite, then $\mathcal{E}U$ exists and is unique. A random variable U is called *centered Gaussian* on $L_p(\mu)$ if, for every $h \in L_q(\mu)$, the real-valued random variable $\langle U, h \rangle$ is centered Gaussian on $\mathbb{R}$.

Let $X^* \sim \psi^*$. Fixing $y \in \mathbb{Y}$, the term $q(y \,|X^*)$ is a random variable in $\mathbb{R}$. The GTLAE estimate $f_n^*(y) := n^{-1} \sum_{t=1}^n q(y \mid X_t)$ is the sample analogue of the expectation $\mathbb{E}q(y \,|X^*)$. Moreover, the definition of f^* in (7) implies that for any fixed $y \in \mathbb{Y}$ we have $\mathbb{E}q(y \,|X^*) = f^*(y)$. This drives the pointwise consistency result in (10).

Repeating this argument in the function space setting yields global consistency results. To begin, observe that $q(\cdot \mid X^*)$ is a random variable in $L_1(\mu)$, and the GTLAE $f_n^* = n^{-1} \sum_{t=1}^n q(\cdot \mid X_t)$ is the sample analogue of the expectation $\mathcal{E}q(\cdot \mid X^*)$. Moreover, it is proved in the appendix that

Lemma 2.1. *If $X^* \sim \psi^*$, then $\mathcal{E}q(\cdot \mid X^*) = f^*$.*

Continuing with this logic leads us to the following results.

Theorem 2.1. *If P is ergodic, then the GTLAE f_n^* is globally consistent for f^*, in the sense that $f_n^* \to f^*$ in $L_1(\mu)$ as $n \to \infty$ $\mathbb{P}$-almost surely. If, in addition, P is V-UE and*

$$\int q(y \mid x)^2 \mu(dy) \leq V(x) \quad \forall\, x \in \mathbb{X}$$

then $\sqrt{n}(f_n^* - f^*)$ converges in distribution to a centered Gaussian on $L_2(\mu)$ with covariance operator C defined by

$$\begin{aligned}\langle g, Ch \rangle = \mathbb{E}\langle g, q(\cdot \mid X_1^*) - f^* \rangle \langle h, q(\cdot \mid X_1^*) - f^* \rangle \\ + \sum_{t \geq 2}^{\infty} \mathbb{E}\langle g, q(\cdot \mid X_1^*) - f^* \rangle \langle h, q(\cdot \mid X_t^*) - f^* \rangle \\ + \sum_{t \geq 2}^{\infty} \mathbb{E}\langle h, q(\cdot \mid X_1^*) - f^* \rangle \langle g, q(\cdot \mid X_t^*) - f^* \rangle\end{aligned}$$

where $(X_t^*)_{t \geq 0}$ is a stationary P-Markov sequence.

From this theorem it can be shown directly that $\|f_n^* - f^*\|_2 = O_P(n^{-1/2})$.

3 Applications of the GTLAE

We now illustrate the numerical properties of GTLAE and GCLAE using the three examples we described in Section 2 These examples illustrate that these estimators not only have attractive statistical properties but are also quite fast and accurate in practical applications.

3.1 Optimal Growth, Stationary Density

Our first computational experiment concerns the benchmark stochastic growth model with a log-linear optimal capital accumulation policy described in example 2.1. For this log-linear model, the stationary distribution for $Y_t := \ln(k_t / \bar{k})$ can be obtained explicitly, and has the form

$$f^* = N(0, v) \quad \text{for} \quad v := \frac{\sigma^2(1 + \rho\alpha)}{(1 - \rho^2)(1 - \alpha^2)(1 - \rho\alpha)} \tag{18}$$

The GTLAE of this density is

$$f_n^*(y) = \frac{1}{n}\sum_{t=1}^{n} q(y \mid k_t, z_t) \qquad (y \in \mathbb{R})$$

where q is given by (11) and (k_t, z_t) is a time series generated by (i.e., simulated from) the model. The performance of the GTLAE will be compared against that of the standard nonparametric kernel density estimator (NPKDE)

$$p_n(y) = \frac{1}{n\delta_n}\sum_{t=1}^{n} K\left(\frac{y - Y_t}{\delta_n}\right) \qquad (y \in \mathbb{R})$$

where $(Y_t)_{t=1}^n$ is simulated from the model, K is a Gaussian kernel (i.e., standard normal density) and the bandwidth δ_n is chosen using Silverman's rule.[15] These are widely used defaults for the NPKDE.

The performance of the two estimators is measured in terms of the L_1 deviation between a particular estimate and the stationary density (18). That is,

$$E(h_n) = \int |(h_n(y) - f^*(y))|\, dy \tag{19}$$

where h_n is either the GTLAE f_n^* or the NPKDE p_n. Note that $E(h_n)$ is a real-valued random variable, and we approximate its expectation by averaging over m independent replications.[16]

[15] That is, $\delta_n = n^{-1/5} 1.06 s_n$, where s_n is the standard deviation of the sample.

[16] A word is in order as to why we use the L_1 distance to assess the fit of density estimators. First, the L_1 measure of distance has the desirable property that it is always finite. (Other measures of distance such as the L_2 norm do not have this property absent additional assumptions.) Second, the L_1 distance between densities f and g equals

$$2 \times \sup_{B \in \mathscr{Y}} \left| \int_B f(x)\mu(dx) - \int_B g(x)\mu(dx) \right|$$

Thus, if the L_1 distance between f and g is less than ϵ, then the deviation in probability assigned to arbitrary B by f and g is less than $\epsilon/2$. This provides a natural quantitative interpretation. Finally, L_2 is insensitive to divergence in the tails. This has led Devroy and Lugosi (2001) to argue that L_2 distances are be avoided when analyzing nonparametric density estimates. Our comparisons of L_1 and L_2 distances led us to the same conclusion.

Table 1: L1 errors, Generalized Time Series Look-Ahead Estimator (GCLAE) and Nonparametric Kernel Density Estimator (NPKDE).

Sample size (n)	1000	1500	2000	2500	3000	3500	4000
NPKDE	0.149	0.127	0.11	0.101	0.094	0.085	0.081
GTLAE	0.141	0.118	0.102	0.092	0.084	0.076	0.073
Ratio (GTLAE/NPKDE)	0.95	0.93	0.92	0.91	0.9	0.9	0.9

Notes
1) Results are averages across 100 replications for each given sample size.

The present set up is almost ideal for the NPKDE p_n, and hence the comparison of the GTLAE f_n^* against p_n provides a rigorous test of the GTLAE's performance. The reasons that p_n should perform strongly are as follows. First, the target density f^* is known to be Gaussian, matching the Gaussian kernel, and Silverman's rule is optimal for a Gaussian target density. Hence the choices for both the kernel and the bandwidth use strong prior information which is not generally available in applications. Second, in contrast to the GTLAE, the rate of convergence for the NPKDE slows as the dimension of $\mathbb{Y}$ increases. Hence our one-dimensional state space favors the NPKDE.

Table 1 shows average L_1 errors for the two estimators for different sample sizes, n. The first row gives values of n. The second row gives an average of $E(p_n)$ over 100 replications for each different n. The third row gives the same for $E(f_n^*)$. The fourth row gives the ratio of the third row to the second. The parameter values in the experiment are $A = 5$, $\alpha = 0.5$, $\beta = 0.9$, $\rho = 0.9$ and $\sigma = 0.1$. The results show that the GTLAE generally outperforms the NPKDE for these values of n, despite the set up being favorable to the NPKDE. The errors for the GTLAE are between 5 and 10 percent lower than those for the NPKDE, with larger sample size (i.e., larger n) increasing the reduction in error achieved by the GTLAE.

It should be noted that for very small values of n, we found that the GTLAE had higher mean errors than the NPKDE. This is due to the high variance of the sequence (z_t), which appears in the GTLAE but not in the

NPKDE.[17] We proved above in Theorem 2.1 that the error for the GTLAE converges to zero at the parametric rate $O_P(n^{-1/2})$, which is faster than the rate for the NPKDE. Hence the GTLAE outperforms the NPKDE eventually (i.e., for large n), even in a applications suited to the NPKDE (such as the one in question).

3.2 Aiyagari Model

As a second application, we use the GTLAE to compute the joint distribution of assets and labor-productivity in the heterogeneous agent economy analyzed by Aiyagari (1994). We parameterize the model using one of the parameterizations reported in Aiyagari (1994). The relative risk-aversion parameter is set to $\mu = 3$, the preference discount rate is $\beta = 0.96$, the real interest rate is $r = 0.0129$ and the wage rate is 1.3712. Labor productivity is assumed to follow the AR(1) process (14). We report results for two different values of the serial correlation parameter for labor productivity: $\rho = 0.3$ and $\rho = 0.9$. The variance of labor productivity is set to 0.4.

The method for implementing the GTLAE was discussed in example 2.2. We compare results over different grid sizes, with M (the size of the grid over productivity shock space) equal to either 10 or 15, and L (the size of the grid over the asset space) ranging between 812 and 5,469.

We first compare the accuracy of the GTLAE estimates of this joint distribution with the Monte Carlo estimator (MCE) described in (12). The MCE tabulates the frequency of each outcome over n Monte Carlo realizations of assets and labor productivity. Results for the two estimators are reported in Table 2. All of the results are based on $n = 10^6$ pseudo realizations of assets and labor productivity. The criteria used to compare the accuracy of the GTLAE and MCE is the L_1 error criterion. Since there is no closed form solution for the exact invariant distribution in this application, we

[17] The sequence has relatively high variance because of the high value of the correlation coefficient ρ we discuss this point in more detail in Section 4.2 below.

Table 2: Accuracy comparison of GTLAE and Monte Carlo estimator of asset-productivity distribution in Aiyagari model with discrete shocks.

Specification		GTLAE		Monte Carlo	
Grid size (assets *x* productivity)	**Persistence**	**L_1 error**	**CPU Time**	**L_1 error**	**CPU Time**
(812 x 10)	rho=0.9	0.026	0.8	0.038	0.5
(1568 x 10)	rho=0.9	0.034	0.8	0.052	0.5
(5058 x 15)	rho=0.9	0.056	1.3	0.109	0.8
(877 x 10)	rho=0.3	0.011	0.8	0.026	0.6
(1695 x 10)	rho=0.3	0.015	0.9	0.037	0.6
(5469 x 15)	rho=0.3	0.026	1.4	0.083	0.9

Notes

1) The sample length is fixed at 1,000,000 for all runs after dropping the first 500 observations.
2) Time is measured in seconds.
3) For the specifications with 10 productivity nodes L1 erros are calculated using the solution from Gaussian elimination as the true density.
4) For the specification with fifteen productivity nodes the true density is computed using a GTLAE simulation of length 3000e+6.
5) The same time series data is used for GTLAE and Monte Carlo estimation.
6) CPU Time and L_1 errors are averages across 100 replications.

use Gaussian elimination methods to compute it numerically.[18] For the two specifications with the largest number of nodes (15 nodes for labor productivity and over 5000 nodes for assets), Gaussian elimination is not feasible due to insufficient core memory. Instead we appeal to our theoretical results, which establish the global convergence of the GTLAE and calculate the true density by conducting a GTLAE simulation of length 3000e+06.

The first thing to notice about Table 2 is that the GTLAE is more accurate than MCE. The GTLAE yields lower L_1 errors than MCE for all choices of the grid size and both settings of the persistence parameter for labor productivity. The accuracy of both estimators declines as we increase the size of the grid. Notice though that the *relative* accuracy of the GTLAE compared to MCE increases with the number of grid points. For the largest

[18]Calculations were performed using an HP workstation with two quad4 Intel processors and 16 gigabytes of RAM. The programs to perform these calculations are written in Matlab and are available from the authors upon request.

grid, the GTLAE is about twice as accurate as the Monte Carlo estimator when the persistence of labor productivity is 0.9 and it is about 3 times as accurate when the persistence is of labor productivity is set to 0.3.

The second noteworthy feature about Table 2 is that the MCE is faster than the GTLAE. This follows from the fact that there are more calculations involved in computing conditional densities than in calculating frequencies. MCE is about 60 percent faster than the GTLAE when labor productivity is persistent. The gap is a bit smaller when the persistence of labor productivity is 0.3.

To further explore the trade off between speed and accuracy we report simulation results in Table 3 that assess the speed of each estimator in achieving a pre-specified L_1 error of 0.025. The results for the GTLAE are reported in column three and results for the Monte Carlo estimator are reported in column four.[19] These results indicate that if we control for accuracy, the GTLAE is faster than the MCE for all choices of the grid size. Moreover, the advantages of the GTLAE as compared to the MCE once again increase with the grid size. For the largest grid size, the GTLAE is three times faster than Monte Carlo when the persistence parameter is 0.9 and eight times faster than Monte Carlo when the persistence parameter is 0.3. For the smallest grid, the speed advantage of the GTLAE ranges from 36 to 300 percent.

Table 3 also provides results for two other common methods used for calculating invariant distributions of Markov chains. The transition matrix iteration method operates on the transition equation (3) from an initial guess of the labor productivity-asset distribution, ψ_0, until the L_1 error is less than 0.025 (see Ljungqvist and Sargent (2004), chapter 2 for more details on this method). Gaussian elimination applies elementary matrix operations to the transition matrix to calculate the invariant distribution directly (see Stachurski (2009), section 4.3.1 for more details on this method). The GTLAE is much faster than either of these two methods. Once gain its speed advantage increases with the size of the grid. The reason for this is

[19] The time for these calculations is not directly comparable with those reported in Table 2 because the programs here we have an additional loop that checks for convergence.

Table 3: Speed Comparison of GTLAE, Monte Carlo, Transition Matrix Iteration, and Gaussian Elimination for labor productivity-asset distribution in Aiyagari model with discrete shocks.

Specification		GTLAE	Monte Carlo	Transition Matrix iteration	Gaussian Elimination
Grid size (assets *x* productivity)	**Persistence**	**CPU Time**	**CPU Time**	**CPU time**	**CPU time**
(812 x 10)	rho=0.9	2.5	3.4	39.1	94.9
(1568 x 10)	rho=0.9	5.3	8.9	149.7	563.5
(5058 x 15)	rho=0.9	41.0	127.0	NA	NA
(877 x 10)	rho=0.3	0.5	1.7	42.4	106.2
(1695 x 10)	rho=0.3	1.1	4.8	150.7	699.3
(5469 x 15)	rho=0.3	7.7	63.3	NA	NA

Notes
1) Time is measured in seconds.
2) GTLAE, Monte Carlo estimation and transition matrix iteration schemes are all stopped when the L_1 error equals 0.025.
3) The same time series data is used for GTLAE and Monte Carlo schemes.
4) For GTLAE and Monte Carlo CPU times are the average across 100 replications and in each replication the first 500 observations are dropped.

that the number of calculations involved in implementing Gaussian elimination and transition matrix iteration grow in a nonlinear way with the size of the grid.

This result illustrates a further important advantage of the GTLAE. Both Gaussian elimination and transition matrix iteration eventually become infeasible due to memory requirements associated with storing and operating on the transition matrix.[20] The GTLAE and MCE avoid this problem by treating the computation of the invariant distribution as an estimation problem.

Taken together, these results show that the GTLAE is also an attractive method for computing stationary densities in settings where the dynamics of the model can be represented as a Markov chain. These benefits are most significant in settings with multiple shocks and/or state variables,

[20]These problems can be ameliorated by using sparse Arnoldi methods but this does not avoid the curse of dimensionality.

or in situations where one needs to compute a density many times such as when estimating structural parameters of the model.

4 Computing Expectations

Let us now return to the general stochastic model with law of motion (1) and stochastic kernel P defined in (2). As above, we assume that P is V-uniformly ergodic for some $V\colon \mathbb{X} \to \mathbb{R}_+$, with unique stationary distribution ψ^*, and consider the distribution of $Y^* = \tau(X^*)$ when $X^* \sim \psi^*$. However, we now drop assumption 2.1, in which case Y^* does not always have a density which represents its distribution.

In this setting, the problem of computing the density of Y^* does not make sense. At the same time, one can still consider the problem of computing $\mathbb{E}Y^*$ when Y^* is real-valued (i.e., $\mathbb{Y} \subset \mathbb{R}$). This nests several important sub-problems. For example, if $\tau(x) = x^k$, then $\mathbb{E}Y^*$ is the k-th moment of the stationary distribution. Such moments are regularly calculated in calibration exercises. Alternatively, if $\tau(x) = \mathbb{1}_B(x)$ for some $B \in \mathscr{Y}$, then $\mathbb{E}Y^* = \mathbb{E}\mathbb{1}_B(X^*) = \psi^*(B)$, which is the probability of event B under the stationary distribution.

4.1 Asymptotic Theory

To simplify notation, we introduce a common shorthand. Letting $h\colon \mathbb{X} \to \mathbb{R}$ be an integrable function, we set

$$Ph(x) := \int h(y)P(x, dy) \qquad (x \in \mathbb{X})$$

Alternatively, Ph can be expressed as

$$Ph(x) = \mathbb{E}[h(X_{t+1}) \mid X_t = x] = \int h[F(x, z)]\phi(dz)$$

Note that for the stationary distribution ψ^* we have

$$h\colon \mathbb{X} \to \mathbb{R} \text{ and } \int |h| d\psi^* < \infty \implies \int Ph d\psi^* = \int h d\psi^* \qquad (20)$$

This can be proved directly. Alternatively, see Stachurski (2009, thm. 9.2.15).

To compute $\mathbb{E}Y^* = \mathbb{E}\tau(X^*) = \int \tau d\psi^*$, let $(X_t)_{t\geq 0}$ be a simulated observation of our process (1) starting from any initial condition $X_0 = x \in \mathbb{X}$, and consider the two estimators

$$E_n\tau := \frac{1}{n}\sum_{t=1}^{n}\tau(X_t)$$

and

$$E_nP\tau := \frac{1}{n}\sum_{t=1}^{n}P\tau(X_t) := \frac{1}{n}\sum_{t=1}^{n}\int \tau[F(X_t,z)]\phi(dz) \tag{21}$$

The first estimator is the standard Monte Carlo estimate, consistency of which follows immediately from (8). The second is a look-ahead estimator. It is also strongly consistent by the same ergodicity property and (20), which gives

$$E_nP\tau := \frac{1}{n}\sum_{t=1}^{n}P\tau(X_t) \to \int P\tau d\psi^* = \int \tau d\psi^* \quad \mathbb{P}\text{-a.s.}$$

Assuming that the integral in (21) can be calculated accurately, $E_nP\tau$ is a better estimator than $E_n\tau$ because its asymptotic variance is lower. An earlier result along these lines was obtained by by McKeague and Wefelmeyer (2000). Their proof used the assumptions of uniform ergodicity and detailed balance, neither of which holds in most economic applications. We now show that an analogous result holds in our more general setting.

Theorem 4.1. *Let P be V-uniformly ergodic, and let $\tau\colon \mathbb{X} \to \mathbb{R}$ be a measurable function satisfying $\tau^2 \leq V$. Then $n^{1/2}(E_n\tau - \psi\tau)$ and $n^{1/2}(E_nP\tau - \psi\tau)$ both converge in law to centered Gaussian distributions on $\mathbb{R}$, with variances v_1 and v_2 respectively. Moreover, $v_2 \leq v_1$ always holds.*

We now turn to some applications of this result.

4.2 Applications

We have shown that GTLAE is weakly more efficient than MCE when calculating expectations. Before describing an economic application first pro-

vide a time-series example that illustrates when the efficiency gains are large and when they are small.

Let us specialize (1) to the real-valued, linear AR(1) process

$$X_{t+1} = F(X_t, \xi_{t+1}) = \mu + \rho X_t + \xi_{t+1} \tag{22}$$

where $(\xi_t)_{t\geq 1}$ is IID with common distribution ϕ, and $\mathbb{E}\xi_1 = 0$. Let $\tau(x) = x$, so that the object of estimation is the unconditional mean. The GTLAE estimates the unconditional mean by

$$\frac{1}{n}\sum_{t=1}^{n}\int \tau[F(X_t, z)]\phi(dz) = \frac{1}{n}\sum_{t=1}^{n}\int (\mu + \rho X_t + z)\phi(dz) = \mu + \rho\frac{1}{n}\sum_{t=1}^{n} X_t$$

while MCE estimates the unconditional mean as $\frac{1}{n}\sum_{t=1}^{n} X_t$. Letting $\mathbb{V}$ denote the variance operator, the variance (across replications) of the GTLAE estimator of the mean is given by

$$\mathbb{V}(\mu + \rho\frac{1}{n}\sum_{t=1}^{n} X_t) = \rho^2\mathbb{V}(\frac{1}{n}\sum_{t=1}^{n} X_t) \tag{23}$$

while the variance of the Monte Carlo estimator of the mean is

$$\mathbb{V}(\frac{1}{n}\sum_{t=1}^{n} X_t) \tag{24}$$

Inspection of these formulas shows that the relative efficiency of GTLAE compared to MCE is largest when persistence is low. As ρ is reduced towards zero, the conditional expectation tends to the unconditional expectation and the efficiency gains of GTLAE are large. Alternatively, as ρ is increased towards one, the standard deviation of the GTLAE estimator also increases and tends towards the MCE estimator of the standard deviation.

Next we compare the relative efficiency GTLAE and MCE in estimating moments in the Aiyagari model with discrete shocks. Table 4 reports the relative efficiency of GTLAE to MCE in estimating the first three moments of consumption and the first three cross-moments of consumption with labor productivity.

Table 4: Efficiency of GTLAE relative to MCE in estimating moments of consumption.

	E(c)	E(c^2)	E(c^3)	E(cz)	E(c^2z^2)	E(c^3z^3)
rho=0.3						
n=1,000	0.92	0.93	0.93	0.60	0.56	0.50
n=1,000,000	0.91	0.92	0.92	0.60	0.56	0.51
rho=0.9						
n=1,000	0.97	0.97	0.97	0.93	0.92	0.90
n=1,000,000	0.97	0.97	0.97	0.93	0.92	0.90

Notes

1 The results are based on 200 replications of simulations of length n. For each replication the first 1000 observations are dropped.

2. Results report the ratio of the GTLAE standard deviation across replications

Results are reported for two simulation lengths $n = \{1e3, 1e6\}$ and two settings of the persistence of labor productivity $\rho = \{0.3, 0.9\}$. Consider first the results for the case where $\rho = 0.3$. The efficiency gain associated with using GTLAE ranges between 7 and 9 percent for the first three moments of individual consumption. The efficiency gains associated with using GTLAE to estimate cross-moments of consumption and labor-productivity are substantially higher. The standard deviation of GTLAE in some cases is half the size of the MCE standard error.

An important factor that limits the overall efficiency benefits of GTLAE is the persistence of consumption. Consumption smoothing considerations imply that consumption is more persistent than labor productivity. Increasing the persistence of labor-productivity also increases the persistence of individual consumption and this reduces the efficiency benefits of GTLAE. In our simulations the first order serial correlation of individual consumption is 0.89 when $\rho = 0.3$ and it rises to 0.97 when $\rho = 0.9$. When $\rho = 0.9$ the maximum improvement in efficiency from using the GTLAE estimator is correspondingly smaller: 3 percent for the own moments of consumption and 10 percent for the cross-moments.

The fact that the dynamic system being considered is not full rank can also affect the relative efficiency of GTLAE when estimating moments. In our example, the one step ahead conditional density of assets, for instance, is

degenerate. In this situation GTLAE assigns all probability mass to the value of assets given by the asset policy function and the calculation of average assets reduces to the Monte Carlo estimator.

As a final application we illustrate that the computation of the invariant distribution in a model with discrete shocks can also be reformulated as a problem of computing an expectation. To see why this is the case note that when we are computing densities with discrete shocks, the value of the density:

$$f(y) = PX_t = y$$

can be expressed as $P\{X_t = y\} = E\mathbb{1}\{X_t = y\}$.[21] To apply Theorem 4.1 fix y and set $\tau(x) = \mathbb{1}\{X_t = y\}$. Then we have

$$E_n\tau = (1/n)\sum_{t=1}^{n} \mathbb{1}\{X_t = y\}$$

which is the discrete Monte Carlo density estimator (13). Observe next that

$$P\tau := \int \tau(z)P(x,dz) = \sum_z \tau(z)p(x,z) = \sum_z \mathbb{1}\{z = y\}\, p(x,z) = p(x,y)$$

so that

$$E_nP\tau = (1/n)\sum_{t=1}^{n} p(X_t, y)$$

which is the discrete GTLAE density estimator (12). Then from Theorem 4.1 it follows that the discrete GTLAE density estimator has a lower asymptotic variance than the discrete Monte Carlo density estimator. Thus, the improved accuracy of the GTLAE density estimator compared to the Monte Carlo density estimator documented Table 2 is an example of a more general result.

5 Nonstationary Environments

So far we have considered results for the stationary setting only. In nonstationary environments the marginal distributions do not converge, and

[21] We are exploiting the discrete structure of the shock process here. In the continuous case with Lebesque measure, the equality is not valid.

one thinks instead of computing the marginal distribution of $Y_t = \tau(X_t)$ at a given point in time T. The model we consider here is the same as (1), but we now permit the transition rule to depend explicitly on t:

$$X_{t+1} = F_t(X_t, \xi_{t+1}), \qquad (\xi_t)_{t \geq 1} \overset{\text{IID}}{\sim} \phi \tag{25}$$

We replace assumption 2.1 with a nonstationary version:

Assumption 5.1. For every $t \geq 0$, there exists a measurable function q_t from $\mathbb{Y} \times \mathbb{X}$ into $\mathbb{R}_+$ satisfying

$$\int_B q_t(y \mid x)\mu(dy) = \phi\{z \in \mathbb{Z} : \tau[F_t(x, z)] \in B\} \tag{26}$$

for all $x \in \mathbb{X}$ and all $B \in \mathscr{Y}$.

Given this assumption, the distributions in the sequence $(Y_t)_{t \geq 1}$ can all be represented by densities with respect to μ. In particular, the function f_t defined by

$$f_t(y) := \int q_{t-1}(y \mid x)\psi_{t-1}(dx) \qquad (y \in \mathbb{Y}) \tag{27}$$

represents Y_t for any given $t \geq 1$. (Here ψ_{t-1} is the distribution of X_{t-1}. The proof is similar to the justification given for (7) in section 2, and hence is omitted.) Our interest is in computing f_T for fixed $T \in \mathbb{N}$. For this purpose we propose a generalized cross-sectional look-ahead estimator (GCLAE), defined by

$$f_T^n(y) := \frac{1}{n} \sum_{i=1}^n q_{T-1}(y \mid X_{T-1}^i) \qquad (y \in \mathbb{Y}) \tag{28}$$

where $(X_{T-1}^1, \ldots, X_{T-1}^n)$ is n independent copies of X_{T-1}. Algorithm 2 indicates how to obtain an observation of $f_T^n(y)$ for fixed $y \in \mathbb{Y}$.

5.1 Asymptotic theory

Regarding the pointwise properties of f_T^n, pick any $y \in \mathbb{Y}$. Observe that $f_T^n(y)$ is the sample mean of n independent copies of the real-valued random variable $q_{T-1}(y \mid X_{T-1})$. Hence $f_T^n(y)$ is unbiased and consistent for

Algorithm 2: Cross-sectional look-ahead estimate of $f_T(y)$

for *i in 1 to n* **do**
  set $X \leftarrow X_0$ (X_0 is determined in the problem domain);
  for *t in 1 to* $T-1$ **do**
    draw $\xi \sim \phi$ and set $X \leftarrow F_{t-1}(X, \xi)$;
  end
  set $X_{T-1}^i \leftarrow X$;
end
return $f_T^n(y) = \frac{1}{n}\sum_{i=1}^n q_{T-1}(y \mid X_{T-1}^i)$

the expectation of this random variable. In view of (27), this expectation is

$$\mathbb{E}q_{T-1}(y \mid X_{T-1}) = \int q_{T-1}(y \mid x)\psi_{T-1}(dx) = f_T(y)$$

Thus, $f_T^n(y)$ is unbiased and consistent for $f_T(y)$. Let us now extend this pointwise result to a global result, analogous to theorem 2.1:

Theorem 5.1. *Fix $T \in \mathbb{N}$. The estimator f_T^n is a globally unbiased estimator of f_T, in the sense that $\mathcal{E}f_T^n$ exists in $L_1(\mu)$ and is equal to f_T for every $n \in \mathbb{N}$. It is also globally consistent, in the sense that $\|f_T^n - f_T\|_1 \to 0$ almost surely as $n \to \infty$. If there exists a ψ_{T-1}-integrable function $V\colon \mathbb{X} \to \mathbb{R}$ such that*

$$\int q(y \mid x)^2 \mu(dy) \leq V(x), \quad \forall\, x \in \mathbb{X}$$

then $\sqrt{n}(f_T^n - f_T)$ converges in distribution to a centered Gaussian on $L_2(\mu)$.

5.2 Application: Optimal Growth

In the previous example the target density f^* was Gaussian, an optimal setting for the standard NPKDE with Gaussian kernel and default choice of bandwidth. In applications where distributions must be calculated numerically, the target density is almost never Gaussian. (If it were then analytical solutions would typically be available.) In the next experiment we compare the generalized look-ahead estimator with the NPKDE in a non-Gaussian environment.

Table 5: Average L1 errors of Generalized Cross-sectional Look-Ahead Estimator (GCLAE) and NPKDE for period T capital density in neoclassical growth model.

Sample size (*n*)	1000	1500	2000	2500	3000	3500	4000
NPKDE	0.11	0.098	0.087	0.08	0.076	0.071	0.068
GCLAE	0.035	0.029	0.025	0.022	0.02	0.019	0.017
Ratio (GCLAE/NPKDE)	0.32	0.29	0.29	0.27	0.27	0.26	0.25

Notes
1) Results are averages across 100 replications for each given sample size

Since the stationary density f^* is always Gaussian, we consider instead the marginal densities f_T at a fixed point in time T, which is non-Gaussian when the initial distribution for y_0 is non-Gaussian. In particular, we assume that the density f_0 of y_0 is a mixture of two normal densities: $f_0 = g/2 + g'/2$, where $g = N(\mu, v)$ and $g' = N(\mu', v')$. Mixing two normal densities gives a non-Gaussian marginal f_T which can still be calculated analytically (and hence compared against the different estimates).

The marginal density f_T can be approximated either by the GCLAE f_T^n in (28) or by a cross-sectional NPKDE of the form

$$p_n(y) = \frac{1}{n\delta_n} \sum_{i=1}^{n} K\left(\frac{y - \hat{k}_T^i}{\delta_n}\right) \qquad (y \in \mathbb{R})$$

where $(k_T^i)_{i=1}^n$ is n independent (simulated) observations of k_T. Here the kernel is again Gaussian and the bandwidth is chosen by Silverman's rule.

To simplify calculations further the correlation parameter ρ is set to zero, so that shocks in the growth model are IID $N(0, \sigma^2)$. The parameters are thus $A = 5$, $\alpha = 0.5$, $\beta = 0.9$, $\rho = 0.0$ and $\sigma = 0.1$. The simple nature of the model and the mixed Gaussian initial conditions are admittedly contrived, but they do help to illustrate the strong performance of the GCLAE relative to the NPKDE in settings that are less optimal for the latter.

Table 5 gives results for the same values of n as in table 1, and all rows have the same interpretations as in that table. The percentage reductions

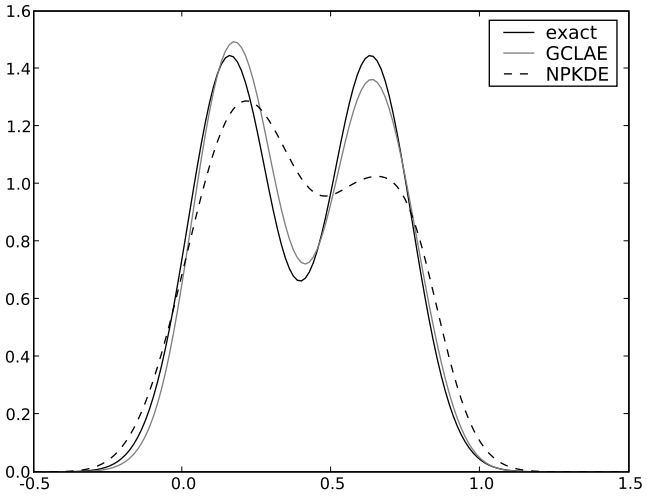

Figure 1: Relative performance, $n = 200$

in error achieved by the GCLAE over the NPKDE are very large. As before, the relative accuracy of the GCLAE increases as n gets larger.

Observations of the deviations between the GCLAE, the NPKDE and the exact marginal f_T are presented in figures 1 and 2. In the first figure $n = 200$, while in the second $n = 2000$. Although the accuracy of the GCLAE relative to the NPKDE is visually more apparent when $n = 200$ (i.e., in figure 1), the percentage reduction in error achieved by the GCLAE over the NPKDE is in fact larger in figure 2 (69.5% versus 53.2%).

To conclude this section we note that the performance of the NPKDE can probably be improved for this particular experiment by a different choice of the bandwidth and kernel. However, it is not immediately obvious how to do this in the present setting, and even less so in applications where the target density is unknown. In contrast, the GCLAE *automatically* incorporates all available information about the model into the estimator.

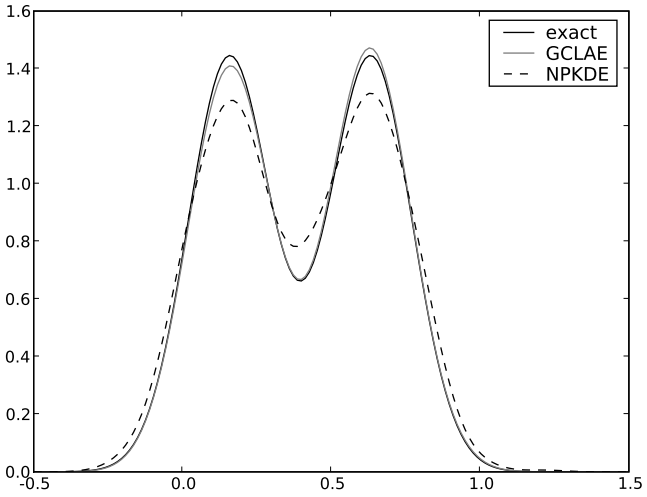

Figure 2: Relative performance, $n = 2000$

6 Conclusion

In this paper we have proposed a generalized look-ahead approach for estimating densities and computing expectations of functions of the state in economic models. Our methods are applicable to a broad class of economic problems that can be expressed as a dynamic programming problem with Markovian shocks. We have provided asymptotic results for our estimators. Our estimators exploit the parametric assumptions that are typically imposed on the shock processes in economic applications and are thus more efficient than standard Monte Carlo estimators of densities and expectations.

We have also shown that GLAE can offer significant computational speed advantages in practical applications. GLAE is a particularly attractive option for estimating densities and expectations in models with multiple endogenous state variables and/or shocks or in situations where computations have to be repeated many times as when e.g. estimating parameters.

7 Technical Appendix

Proof of lemma 2.1. That $q(\cdot \mid X^*)$ is an element of $L_1(\mu)$ for each realization of X^*. That it is also a random variable (i.e., a measurable map) follows easily from separability of $L_1(\mu)$.[22] To show that $\mathcal{E}q(\cdot \mid X^*) = f^*$, we must prove that $\mathbb{E}\langle q(\cdot \mid X^*), h\rangle = \langle f^*, h\rangle$ for all $h \in L_\infty(\mu)$. This follows from Fubini's theorem, because

$$\begin{aligned}\mathbb{E}\langle q(\cdot \mid X^*), h\rangle &= \int\int q(y \mid x)h(y)\mu(dy)\psi^*(dx) \\ &= \int h(y)\int q(y \mid x)\psi^*(dx)\mu(dy) = \int h(y)f^*(y)\mu(dy) = \langle h, f^*\rangle\end{aligned}$$

□

Proof of theorem 2.1. Regarding consistency, let P be ergodic, let $(X_t)_{t\geq 0}$ be P-Markov and let $X^* \sim \psi^*$. Define $Q(x) := q(\cdot \mid x) - f^*$, which is a measurable function from $\mathbb{X}$ to $L_1(\mu)$. Note that $\mathcal{E}Q(X^*) = 0$ by lemma 2.1. We need to show that

$$\lim_{n\to\infty} \|f_n^* - f^*\| = \lim_{n\to\infty} \left\| \frac{1}{n}\sum_{t=1}^{n} Q(X_t) \right\| = 0 \qquad (\mathbb{P}\text{-almost surely}) \tag{29}$$

Our proof is an extention of that for the IID Banach space LLN, as given in Bosq (2000, thm. 2.4). To begin, fix $\epsilon > 0$ and choose a partition $\{B_j\}$ of $L_1(\mu)$ such that each B_j has diameter less than ϵ. For any $L_1(\mu)$-valued random variable Y, we let $L_J Y := \sum_{j=1}^{J} b_j \mathbb{1}\{Y \in B_j\}$. We use the following fact, a proof of which can be found in Bosq (2000, pp. 27-28):

$$\exists\, J \in \mathbb{N} \text{ with } \mathbb{E}\|Q(X^*) - L_J Q(X^*)\| < 2\epsilon \tag{30}$$

Our first claim is that

$$\lim_{n\to\infty} \left\| \frac{1}{n}\sum_{t=1}^{n} L_J Q(X_t) - \mathcal{E}L_J Q(X^*) \right\| = 0 \qquad (\mathbb{P}\text{-almost surely}) \tag{31}$$

[22] See, in particular, Bosq (2000, lemma 1.2).

To establish (31), we can use the real ergodic law (8) to obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} L_J Q(X_t) &= \sum_{j=1}^{J} b_j \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{Q(X_t) \in B_j\} \\ &\to \sum_{j=1}^{J} b_j \mathbb{P}\{Q(X^*) \in B_j\} = \mathcal{E} L_J Q(X^*)\end{aligned}$$

almost surely, where the last equality follows immediately from the definition of $\mathcal{E}$. Thus (31) is established.

Returning to (29), we have

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

Using real-valued ergodicity again, as well as (31), we get

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

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

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

In view of (30) we then have

$$\limsup_{n\to\infty} \left\| \frac{1}{n}\sum_{t=1}^{n} Q(X_t) \right\| \leq 4\epsilon \qquad (\mathbb{P}\text{-almost surely})$$

Since ϵ is arbitrary, the proof of (29) is now done.

Regarding the asymptotic normality result, this follows immediately from the Hilbert space CLT of Stachurski (2009, theorem 3.1), where $x \mapsto q(\cdot \mid x)$ corresponds to T_0 in that theorem. The only point that needs checking vis-a-vis that CLT is that $\mathcal{E} q(\cdot \mid X^*) = f^*$, and this has already been verified in lemma 2.1. □

Proof of theorem 4.1. We make use of the following fact: By the Markov property, for our state process $(X_t)_{t\geq 0}$ and integrable $h\colon \mathbb{X} \to \mathbb{R}$ we have

$$\mathbb{E}_t h(X_{t+k}) = P^k h(X_t)$$

where $\mathbb{E}_t$ is expectation conditional on $\sigma(X_0, \ldots, X_t)$, and P^k is the k-th iterate of the mapping $h \mapsto Ph$. For convenience, let $\bar{h} := h - \int h d\psi^*$ for any $h\colon \mathbb{X} \to \mathbb{R}$. By the Markov chain CLT for V-uniformly ergodic kernels (Meyn and Tweedie, 1993, p 411), if $g\colon \mathbb{X} \to \mathbb{R}$ with $g^2 \leq V$, then

$$n^{-1/2} \sum_{t=1}^{n} \bar{g}(X_t) \to N(0, v)$$

in distribution, where

$$v =: \mathbb{E}\bar{g}^2(X_1^*) + 2 \sum_{t\geq 2} \mathbb{E}\bar{g}(X_1^*)\bar{g}(X_t^*) \tag{32}$$

Here $(X_t^*)_{t\geq 0}$ is a stationary version of the process (1). Regarding the standard Monte Carlo estimator $E_n\tau$, it now follows that

$$n^{1/2}(E_n\tau - \int \tau d\psi^*) = n^{-1/2} \sum_{t=1}^{n} \bar{\tau}(X_t) \to N(0, v_1)$$

where v_1 is given by

$$v_1 =: \mathbb{E}\bar{\tau}^2(X_1^*) + 2 \sum_{t\geq 2} \mathbb{E}\bar{\tau}(X_1^*)\bar{\tau}(X_t^*) \tag{33}$$

Regarding the look-ahead estimator, we claim that $n^{1/2}(E_nP\tau - \int \tau d\psi^*) \to N(0, v_2)$, where

$$v_2 =: \mathbb{E}\bar{P\tau}^2(X_1^*) + 2 \sum_{t\geq 2} \mathbb{E}\bar{P\tau}(X_1^*)\bar{P\tau}(X_t^*) \tag{34}$$

To show this, observe that

$$n^{1/2}(E_nP\tau - \int \tau d\psi^*) = n^{1/2}(E_nP\tau - \int P\tau d\psi^*) = n^{-1/2} \sum_{t=1}^{n} \bar{P\tau}(X_t)$$

where the second equality follows from (20).

Now note that P is also $\hat{V}$-UE, where $\hat{V} := \lambda V + L$. To see this, observe that the drift inequality (17) holds with $\hat{V}$ in place of V, because P is V-UE, and hence

$$P\hat{V} = \lambda PV + L \leq \lambda(\lambda V + L) + L = \lambda \hat{V} + L$$

Second, the sublevel sets $\{\hat{V} \leq \alpha\}$ are P-small, because the sublevel sets V are P-small, and

$$\{\hat{V} \leq \alpha\} = \{\lambda V + L \leq \alpha\} = \{V \leq (\alpha - L)/\lambda\}$$

Thus the CLT claim for $n^{1/2}(E_n P\tau - \int \tau d\psi^*)$ will hold if we can show that $(P\tau)^2 \leq \hat{V}$. Using Jensen's inequality, $\tau^2 \leq V$ and the drift condition (17),

$$(P\tau)^2 \leq P\tau^2 \leq PV \leq \lambda V + L =: \hat{V}$$

It remains only to show that $v_2 \leq v_1$. To see this, consider first the term $\mathbb{E}\bar{P\tau}^2(X_1^*)$. Writing $\psi^*\tau$ for $\int \tau d\psi^*$ and using Jensen's inequality for conditional expectations, we obtain

$$\begin{aligned}
\mathbb{E}(P\tau(X_1^*) - \psi^*\tau)^2 &= \mathbb{E}(\mathbb{E}_1(\tau(X_2^*) - \psi^*\tau))^2 \\
&\leq \mathbb{E}\mathbb{E}_1(\tau(X_2^*) - \psi^*\tau)^2 \\
&= \mathbb{E}(\tau(X_2^*) - \psi^*\tau)^2 = \mathbb{E}(\tau(X_1^*) - \psi^*\tau)^2
\end{aligned}$$

where the last step is by stationarity. In other words, $\mathbb{E}\bar{P\tau}^2(X_1^*) \leq \mathbb{E}\bar{\tau}^2(X_1^*)$. To complete the proof that $v_2 \leq v_1$, then, it is sufficient to show that the autocovariance terms in v_1 and v_2 are equal. That is, for any $t \geq 2$,

$$\mathbb{E}(P\tau(X_1^*) - \psi^*\tau)(P\tau(X_t^*) - \psi^*\tau) = \mathbb{E}(\tau(X_1^*) - \psi^*\tau)(\tau(X_t^*) - \psi^*\tau)$$

To see this, observe that

$$\begin{aligned}
\mathbb{E}(P\tau(X_1^*) - \psi^*\tau)(P\tau(X_t^*) - \psi^*\tau) &= \mathbb{E}(\mathbb{E}_1\tau(X_2^*) - \psi^*\tau)(\mathbb{E}_t\tau(X_{t+1}^*) - \psi^*\tau) \\
&= \mathbb{E}\mathbb{E}_1\mathbb{E}_t(\tau(X_2^*) - \psi^*\tau)(\tau(X_{t+1}^*) - \psi^*\tau) \\
&= \mathbb{E}(\tau(X_2^*) - \psi^*\tau)(\tau(X_{t+1}^*) - \psi^*\tau) \\
&= \mathbb{E}(\tau(X_1^*) - \psi^*\tau)(\tau(X_t^*) - \psi^*\tau)
\end{aligned}$$

The proof is done. □

Proof of Theorem 5.1. Unbiasedness of the estimator f_T^n is equivalent to the claim that $\mathcal{E} q_{T-1}(\cdot \mid X_{T-1}) = f_T$. The proof is almost identical to that for f^* given in the proof of lemma 2.1, and hence is omitted. Regarding consistency, the Banach-space law of large numbers (cf., e.g., Bosq, 2000, Theorem 2.4) implies that if $(U_i)_{i \geq 1}$ is an IID sequence in $L_1(\mu)$ with expectation $\mathcal{E}U$, then

$$\left\| \frac{1}{n} \sum_{i=1}^{n} U_i - \mathcal{E}U \right\|_1 \to 0 \;\; \mathbb{P}\text{-a.s. as } n \to \infty$$

Letting $U_i = q_{T-1}(\cdot \mid X_{T-1}^i)$, where $(X_{T-1}^i)_{i \geq 1} \overset{\text{IID}}{\sim} \psi_{T-1}$, this becomes

$$\| f_T^n - f_T \|_1 \to 0 \;\; \mathbb{P}\text{-a.s. as } n \to \infty$$

which is consistency result that we seek.

Finally, consider the issue of asymptotic normality. From Bosq (2000, theorem 2.7), we can deduce that if $(U_i)_{i \geq 1}$ is an IID sequence in $L_2(\mu)$ such that $\mathbb{E}\|U\|_2^2$ is finite, then

$$n^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} U_i - \mathcal{E}U \right) \qquad (n \geq 1)$$

converges in distribution to a centered Gaussian on $L_2(\mu)$. Once again we take $U_i = q_{T-1}(\cdot \mid X_{T-1}^i)$, where $(X_{T-1}^i)_{i \geq 1} \overset{\text{IID}}{\sim} \psi_{T-1}$. From the condition in Theorem 5.1 we have

$$\int q_{T-1}(y \mid x)^2 \mu(dy) \leq V(x) < \infty \qquad (x \in \mathbb{X})$$

As a consequence, U takes values in $L_2(\mu)$. Moreover, $\mathbb{E}\|U\|_2^2 < \infty$, because

$$\mathbb{E} \int q_{T-1}(y \mid X_{T-1})^2 \mu(dy) \leq \mathbb{E} V(X) = \int V d\psi_{T-1} < \infty$$

Using $\mathbb{E}\|U\|_2 \leq \mathbb{E}\|U\|_2^2 + 1 < \infty$, one can show that $\mathcal{E}U$ is well-defined in $L_2(\mu)$, and $\mathcal{E}U = f_T$. It now follows from the definitions that

$$n^{1/2}(f_T^n - f_T) = n^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} U_i - \mathcal{E}U \right)$$

converges in distribution to a centered Gaussian on $L_2(\mu)$. □

References

[1] Aiyagari, S Rao, 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, Vol. 109(3), pages 659-84.

[2] Bosq, Denis, 2000. *Linear Processes in Function Space,* Springer.

[3] Brock, W. A. and L. Mirman, 1972, "Optimal Economic Growth and Uncertainty: The Discounted Case," *Journal of Economic Theory*, Vol. 4, pages 479-513.

[4] Devroye Luc and Gabor Lugosi, 2001 "Combinatorial Methods in Density Estimation" Springer-Verlag, New York.

[5] Hansen, Gary, D. and Edward C. Prescott, 1995, "Recursive Methods for Computing Equilibria in Business Cycle Models," Chapter 2 in T. F. Cooley, ed., *Frontiers of Business Cycle Research*, Princeton University Press.

[6] Henderson, G. Shane and Peter W. Glynn, 2001, "Computing densities for Markov chains via simulation," *Mathematics of Operations Research* Vol. 26, pages 375-400.

[7] Khan, Aubik and Julia M. Thomas, 2008, "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics." *Econometrica*, Vol. 76(2), pages 395-436.

[8] Kristensen, Dennis, 2007. "Geometric Ergodicity of a Class of Markov Chains with Applications to Time Series Models," mimeo, University of Wisconsin.

[9] Ljungqvist, Lars and Thomas J. Sargent, 2004. *Recursive Macroeconomic Theory*, MIT Press, second edition.

[10] McKeague, Ian W. and Wolfgang Wefelmeyer, 2000, "Markov chain Monte Carlo and Rao-Blackwellization," *Journal of Statistical Planning and Inference,* 85, 171-182.

[11] Meyn, Sean P. and Richard L.Tweedie, 1993, *Markov Chains and Stochastic Stability*, Springer-Verlag: London.

[12] Nelson Charles and Charles Plosser, 1982, "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, vol. 10, pp. 139-162.

[13] Nishimura, Kazuo and John Stachurski, 2005, "Stability of Stochastic Optimal Growth Models: A New Approach," *Journal of Economic Theory,* 122 (1), pp. 100–118.

[14] Saad, Yousef, (1992), *Numerical Methods for Large Eigenvalue Problems*, Halsted Press: New York, NY.

[15] Smith, Anthony M. and Fatih Guvenen 2006 "What do Labor and Consumption Data Jointly Tell About Labor Income Risk?" Unpublished manuscript.

[16] Stachurski, John, 2009, *Economic Dynamics: Theory and Computation*, MIT Press.

[17] Stachurski, John, 2008, "Continuous State Dynamic Programming via Nonexpansive Approximation." *Computational Economics* Vol. 31(2), pages 141-160.

[18] Stachurski, John and Vance Martin, 2008, "Computing the Distributions of Economic Models via Simulation," *Econometrica*, Vol. 76(2), pages 443-450.

[19] Stachurski, John, 2009, "A Hilbert Space Central Limit Theorem for Geometrically Ergodic Markov Chains," mimeo, Kyoto University.

[20] Tauchen, George, 1986, "Finite State Markov Chain Approximations to Univariate and Vector Autoregressions." *Economic Letters*, Vol. 20, pages 507-532.