Burnside, C. (1993) *Consistency of a method of moments estimator based on numerical solutions to asset pricing models.* Econometric Theory, 9 (4). pp. 602-632. ISSN 0266-4666

http://eprints.gla.ac.uk/59421/

Deposited on: 25 January 2012

Econometric Theory, 9, 1993, 602–632. Printed in the United States of America.

CONSISTENCY OF A METHOD OF MOMENTS ESTIMATOR BASED ON NUMERICAL SOLUTIONS TO ASSET PRICING MODELS

CRAIG BURNSIDE
University of Pittsburgh and Queen's University

This paper considers the properties of estimators based on numerical solutions to a class of economic models. In particular, the numerical methods discussed are those applied in the solution of linear integral equations, specifically Fredholm equations of the second kind. These integral equations arise out of economic models in which endogenous variables appear linearly in the Euler equations, but for which easily characterized solutions do not exist. Tauchen and Hussey [24] have proposed the use of these methods in the solution of the consumption-based asset pricing model. In this paper, these methods are used to construct method of moments estimators where the population moments implied by a model are approximated by the population moments of numerical solutions. These estimators are shown to be consistent if the accuracy of the approximation is increased with the sample size. This result depends on the solution method having the property that the moments of the approximate solutions converge uniformly in the model parameters to the moments of the true solutions.

1. INTRODUCTION

In recent macroeconomic research, considerable attention has been given to the numerical solution of economic models that do not yield closed-form solutions for the decision rules of agents or for other endogenous stochastic processes. Examples in the asset pricing literature include Gagnon and Taylor [9], Labadie [15], and Tauchen and Hussey [24]. The properties of several methods used to solve a one-sector growth model are discussed by Taylor and Uhlig [25]. Quite general methods for solving dynamic stochastic economic models have been suggested by Judd [13] and Marcet [18]. These methods are frequently used in modern empirical macroeconomic research. If estimates of the parameters of a model depend in any way on the solution method used

I wish to thank Mark Watson, Ian Domowitz, Marty Eichenbaum, Bo Honoré, Sonal Desai, and Domenico Fanizza for many helpful discussions and comments. I also wish to thank three anonymous referees who provided helpful comments on previous drafts. All errors are my own. My research was supported in part by the Social Sciences and Humanities Research Council of Canada.

to solve a particular model, then clearly the convergence properties of the solution method will be important in determining the reliability of those estimates.

Determining the properties of an econometric estimator that depends on a numerical solution method requires extensions to the usual theoretical results pertaining to these methods. Much of this theory deals only with the convergence of solution functions in compact state spaces and, more importantly, is often limited to pointwise convergence results. For example, Marshall [19] establishes pointwise convergence in compact state spaces for approximate solutions obtained using the method proposed by Marcet [18]. Judd [13] discusses the literature on the pointwise convergence of Galerkin methods of which his method is an example. An exception to this is Tauchen and Hussey [24], who establish uniform convergence in a compact state space for their method. For econometric application of this method, two important extensions are necessary. First, uniform convergence must be established in the parameter space since this is the usual route to establishing the consistency of an estimator. Second, to handle the distributional assumptions commonly used in empirical work, theory must be provided for cases where the state space is unbounded.

This paper discusses a Generalized Method of Moments (GMM) estimator, Hansen [11], which is based on Tauchen and Hussey's [24] numerical solution technique. This solution technique is applicable in models where the state variables are exogenous, or relate trivially to exogenous variables, and in which the unknown solution functions appear linearly in the Euler equations. Their method has been applied in other fields in the solution of Fredholm integral equations of the second kind. It is one of a large number of methods which solve models by discretizing the state space when it is continuous.

The method of moments estimator proposed here does not exploit the Euler equations of a given model. Rather, it minimizes a quadratic form in the difference between a vector of population moments implied by the model and a corresponding vector of sample moments. This being the case, the properties of the estimator will depend on the method used to approximate the population moments. This kind of estimator was used by Hodrick, Kocherlakota, and Lucas [12] and relates closely to other work by Kocherlakota [14]. These estimators can be viewed as substitutes for Euler equation based estimators. For example, they enable the estimation of models that have unobservable variables in the Euler equations. Even when Euler equation based estimation is used, other moments are often used at a diagnostic stage to statistically evaluate a model. For example, see Christiano and Eichenbaum [7], Cecchetti, Lam, and Mark [6], and Burnside, Eichenbaum, and Rebelo [5]. Many of the issues surrounding approximation which are discussed in this paper are relevant for both estimation and diagnostic testing.

In this paper, it is assumed that the solution method allows for analytic approximations to the population moments used in constructing the estimator. Therefore, the discussion here is distinct from a discussion of simulation estimators that compute Monte-Carlo approximations to the population moments. Such estimators have been proposed by Duffie and Singleton [8] and Lee and Ingram [16], among others. Simulation methods are particularly useful in cases where the solution functions are known, but the moments of these functions are difficult, or impossible, to calculate analytically. Rather than allow for simulation error, this paper focuses on the quite distinct issue of analytic approximation error.

To illustrate this class of models, an example that fits the description above is the consumption-based Intertemporal Capital Asset Pricing Model (ICAPM) of Lucas [17]. Consider the case where there is an economic agent with constant relative risk aversion preferences who solves the following optimization problem:

$$\max_{\{c_t\}_{t=0}^{\infty}\{s_{i,t}\}_{t=1}^{\infty} i=1,\dots,M} \quad E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}$$

$$\text{s.t.} \quad c_t + \sum_{i=1}^{M} p_{i,t} s_{i,t+1} = \sum_{i=1}^{M} (p_{i,t} + d_{i,t}) s_{i,t} \qquad t = 0,1,\dots$$

$s_{1,0}, s_{2,0}, \dots, s_{M,0}$ given,

where c_t is consumption at time t, $p_{i,t}$ is the price of one share of the ith asset at time t relative to the price of the consumption good, $s_{i,t}$ is the number of shares of the ith asset held by the agent at time t, $d_{i,t}$ is the dividend payable on each share of the ith asset at time t, β is the discount factor, and γ is the constant rate of relative risk aversion.

In this model, the dividend processes are often taken to be exogenous. In its simplest form, there is only one asset held in net supply, and the agent's entire income is obtained from the dividend yielded by that asset. As a result, consumption and dividends will be equivalent. In exchange economies with more than one asset in net supply, it is still possible to determine consumption as a function of purely exogenous processes. This makes it possible to specify an exogenous joint process governing consumption and the dividends from some strict subset of the assets as discussed in Tauchen [23].

The Euler equations for this model are familiar. The price-dividend ratio for the ith asset, $v_{i,t} = p_{i,t}/d_{i,t}$, is given by

$$v_{i,t} = \beta E_t[\exp(-\gamma \lambda_{t+1} + \zeta_{i,t+1})(1 + v_{i,t+1})] \qquad i = 1,\dots,M, \tag{1}$$

where $\lambda_{t+1} = \log(c_{t+1}/c_t)$, and $\zeta_{i,t+1} = \log(d_{i,t+1}/d_{i,t})$.

Equation (1) is typical of the kind of equation to which the method presented here can be applied. However, some more complicated models can be solved. Throughout the body of the paper, the extent to which different assumptions might limit the method's application will be discussed. Assuming that the joint distribution of consumption growth and dividend growth has an unbounded continuous domain, the solution method used here discretizes the state space only in order to approximate the integral implicit in (1) by a sum. This is one sense in which this paper is distinct from Tauchen and Hussey [24]. They discretize the state space not only to approximate the integral, but also in order to construct a new conditional distribution function over the discrete state space. Tauchen and Hussey [24] develop theory for compact univariate state spaces, although their theory extends straightforwardly to multivariate state spaces. The theory developed here allows explicitly for state spaces of any finite dimension, and unbounded state spaces with strict distributional assumptions.[1]

In Section 2, the method of moments estimator is defined and is shown to be consistent under a set of assumptions about the data-generating process and the quality of the approximation. This serves to motivate the rest of the paper which verifies that these assumptions can be maintained when a particular numerical method is used to solve the model. In Section 3, the numerical method to be applied to the model is introduced, and its properties are derived. In Section 4, an example is used to illustrate the relative importance and restrictiveness of the assumptions made, and to illustrate the solution method itself. The example used is a simple version of the example given in this introduction, where there is a single risky asset.

2. THE METHOD OF MOMENTS ESTIMATOR

The estimator discussed in this section is generic in the sense that it is not particular to the example presented later. The assumptions that will be placed on the stochastic processes of the model are typical of the assumptions used in nonlinear econometrics, and are more general than those used later in the section on the solution method. Other assumptions made in this section are satisfied for the solution method discussed in Section 3, but could be satisfied by other solution methods.

To begin, assumptions are placed on the forcing processes of the model, which will ensure almost sure convergence of the sample means of continuous functions of these processes.

Assumption 1. $\{Z_t\}_{t=-\infty}^{+\infty}$ is a sequence of vector-valued random variables defined on the complete probability space $(\Omega, \mathcal{A}, P)$ that is strong mixing of size $-r/(r-2)$ for some $r > 2$. ■

Assumption 2. Define random vectors $R_t = R_t(Z_t^n) \equiv R_t(\omega)$, where $R_t(\cdot)$ is a Borel measurable function of $Z_t^n = (Z_{t-n}, \ldots, Z_{t+n})$ with range in $\mathbf{R}^l$. The function $R_t(\cdot)$ depends on finitely many lags and leads of Z. ■

Assumption 3. Define a vector-valued function $g(R_t): \mathbf{R}^l \to \mathbf{R}^k$ which is continuous in R_t, and assume that there exists a sequence of random variables $\{d_t\}$ with $|g(R_t)| \leq d_t$, and $\|d_t\|_r \leq \Delta < \infty$, $t = 0,1,2,\ldots$.

PROPOSITION 1

$$\lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T}[g(R_t)-Eg(R_t)]\right| = 0 \qquad \text{almost surely.}$$

Proof. From Assumptions 1-3, it follows that $g(R_t)$ is near epoch dependent of size $-\frac{1}{2}$. The result follows immediately from Gallant's [10, p. 515] uniform strong law of large numbers. ■

Assumption 4. Assuming that $\{Z_t\}$ is strictly stationary, and allowing the random vector $g(R_t)$ to depend on the parameter vector θ, define a function $\mu(\theta) = Eg(R_t,\theta)$, where $\mu:\Theta\to\mathbf{R}^k$. ■

Let the true parameter vector be θ_0. In order to estimate θ, a method of moments estimator minimizing the distance between the sample mean of $g(R_t,\theta_0)$ and $\mu(\theta)$ could be used. However, suppose that the function $\mu(\theta)$ is unknown but can be approximated numerically by the function $\mu_T(\theta)$, where $\mu_T:\Theta\to\mathbf{R}^k$. The approximating function is assumed to change with the sample size, T. An example of this would be increasing the number of grid points in a discrete state space method as the sample size increased.

An estimator for the parameter vector θ is

$$\hat\theta_T = \arg\min_{\theta\in\Theta} Q_T(\theta), \qquad Q_T(\theta) = m_T(\theta)'D_T m_T(\theta),$$

where

$$m_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g(R_t,\theta_0) - \mu_T(\theta),$$

and the $k\times k$ matrix D_T is positive definite, and possibly stochastic. The number of moment conditions k is assumed to be no less than the dimension of θ.

In order to obtain consistency of the estimator $\hat\theta_T$, it is necessary that the approximate moment function, μ_T, converges uniformly to the true moment function, μ.

Assumption 5

$$\lim_{T\to\infty}\sup_{\Theta}|\mu_T(\theta)-\mu(\theta)| = 0. \qquad ■$$

For a discrete state space method to satisfy Assumption 5, it would be necessary, although not sufficient, for the approximate solutions to converge uniformly in the parameter space, as well as in the state space. The remaining sections in this paper are concerned with verifying Assumption 5 for the solution method described in the introduction.

Notice that

$$\begin{aligned} m_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} g(R_t,\theta_0) - \mu_T(\theta) \\ &= \frac{1}{T}\sum_{t=1}^{T} [g(R_t,\theta_0) - \mu(\theta)] + [\mu(\theta) - \mu_T(\theta)] \\ &= \frac{1}{T}\sum_{t=1}^{T} [g(R_t,\theta_0) - \mu(\theta_0)] + [\mu(\theta_0) - \mu(\theta)] + [\mu(\theta) - \mu_T(\theta)]. \end{aligned}$$

Define $m_0(\theta) = \mu(\theta_0) - \mu(\theta)$. This gives

$$m_T(\theta) - m_0(\theta) = \frac{1}{T}\sum_{t=1}^{T} [g(R_t,\theta_0) - \mu(\theta_0)] + [\mu(\theta) - \mu_T(\theta)].$$

The triangle inequality implies that

$$|m_T(\theta) - m_0(\theta)| \le \left| \frac{1}{T}\sum_{t=1}^{T} [g(R_t,\theta_0) - \mu(\theta_0)] \right| + |\mu(\theta) - \mu_T(\theta)|.$$

But from Proposition 1 and Assumption 5, this establishes the following:

PROPOSITION 2

$$\lim_{T\to\infty} \sup_{\Theta} |m_T(\theta) - m_0(\theta)| = 0 \qquad \text{almost surely.} \qquad \blacksquare$$

Assumption 6. The sequence of weighting matrices D_T, converges almost surely to a constant positive definite matrix D_0.

Assumption 7. θ_0 is the unique zero of the function $m_0(\theta)$. ■

Under Assumptions 6 and 7, the following proposition is true.

PROPOSITION 3 (CONSISTENCY). *The sequence of functions $Q_T(\theta)$ converges almost surely uniformly in Θ as $T \to \infty$ to the nonstochastic function $Q(\theta) = m_0(\theta)'D_0 m_0(\theta)$. $Q(\theta)$ attains a unique global minimum at θ_0. Furthermore, $\hat{\theta}_T$ converges almost surely to θ_0 as $T \to \infty$.* ■

Proposition 3 is established by Amemiya's [1] Theorem 4.1.1, which does not require that $\hat{\theta}_T$ be the unique minimizer of $Q_T(\theta)$, as long as it is chosen in such a way that it is a measurable function of the data.

Asymptotic normality presents greater difficulties, since more needs to be known about the approximate moment functions than that they converge uniformly. To be specific, in a discrete state space method, the major issue is at what rate the number of grid points must be expanded, relative to the sample size, in order to obtain the distributional result. To illustrate the problem, consider the asymptotic distribution of $m_T(\theta_0)$.

$$m_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} g(R_t,\theta_0) - \mu_T(\theta_0)$$
$$= \frac{1}{T}\sum_{t=1}^{T} g(R_t,\theta_0) - \mu(\theta_0) + \mu(\theta_0) - \mu_T(\theta_0).$$

Therefore, multiplying everywhere by $\sqrt{T}$,

$$\sqrt{T}m_T(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} [g(R_t,\theta_0) - \mu(\theta_0)] + \sqrt{T}[\mu(\theta_0) - \mu_T(\theta_0)].$$

Since the first term on the right will be asymptotically normal under regularity conditions, the asymptotic normality of $\sqrt{T}m_T(\theta_0)$ depends upon whether the remaining piece can be driven to zero in probability. Of course, this requires that the error of approximation $\mu_T(\theta_0) - \mu(\theta_0)$ be of order less than or equal to $o_p(T^{-1/2})$. Tauchen and Hussey [24] provide some evidence on rates of convergence of solution functions using their method. While these results suggest that rather rapid convergence is obtained, their evidence does not apply directly to moments of the solution functions.

A second result that would be needed to establish asymptotic normality would be that the function $\partial\mu_T(\theta)/\partial\theta$ converges to some function $d(\theta)$ as $T \to \infty$. This function need not be the same as $\partial\mu(\theta)/\partial\theta$. Let $d_0 = d(\theta_0)$. Under the regularity conditions set forth by Hansen [11], the asymptotic distribution of $\hat{\theta}_T$ would then be

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N[0, (d_0'D_0d_0)^{-1}d_0'D_0SD_0d_0(d_0'D_0d_0)^{-1}],$$

where S is the asymptotic variance covariance matrix of $\sqrt{T}m_T(\theta_0)$. Since the second assumption is not verified in this paper for the solution method discussed in the next section, asymptotic normality is not pursued in greater detail. If a problem is sufficiently smooth, it is reasonable to believe that asymptotic normality could be established.

3. THE NUMERICAL SOLUTION METHOD

In this section, the numerical method to be applied to equations such as (1) is presented. It is almost identical to the discrete state space method proposed by Tauchen and Hussey [24] for problems where the Euler equations can be represented as Fredholm integral equations of the second kind. Differences between this method and Tauchen and Hussey's will be mentioned as they arise. Although the equations in this section will not be written entirely generally, broadly speaking, the methods presented are applicable to any problem in which the solution functions appear linearly in the Euler equations, and in which all relevant functions are continuous.

Solution methods of this variety require a full description of the laws of motion of the forcing processes. In this section, these processes will be assumed to have multivariate normal distributions, as this is the most likely assumption to be used in practice. The assumption of normality is more limiting, but is interesting since it illustrates the special difficulties which would arise under any distributional assumption involving an unbounded state space.

Assumption 8. Let $\{\xi_t\}_{t=-\infty}^{+\infty}$, $\xi_t \in \mathbf{R}^S$, be a sequence of random vectors governed by a stationary probability distribution such that the density of ξ_t conditional on all past information can be written $\tilde{f}(\xi_t \mid \xi_{t-1};\theta_1)$, where $\theta_1 \in \Theta_1 \subset \mathbf{R}^q$, Θ_1 compact.[2] It is assumed that stationarity is maintained for all $\theta_1 \in \Theta_1$. The true parameter value is $\theta_{10} \in \operatorname{int} \Theta_1$. ■

Assumption 9 (Normality). Let

$$\tilde{f}(\xi_t \mid \xi_{t-1};\theta_1) = (2\pi)^{-S/2}|\Sigma_\epsilon|^{-1/2} \times \exp(-\tfrac{1}{2}[(I-\Pi L)\xi_t - \mu]'\Sigma_\epsilon^{-1}[(I-\Pi L)\xi_t - \mu]),$$

where μ is a vector of constants, and Σ_ϵ is a $S \times S$ positive definite symmetric matrix. Thus, θ_1 represents the elements of Π, μ, and Σ_ϵ arranged in a single vector. Clearly, Assumption 8 requires that the roots of $|I - \Pi Z| = 0$ lie outside the unit circle for all $\theta_1 \in \Theta_1$, and that Σ_ϵ is positive definite for all $\theta_1 \in \Theta_1$. ■

The vector-valued process ξ can be thought of as the exogenous process, or some linear combination of exogenous processes, in the economic model. In the endowment economy described in the introduction, these were the consumption and dividend processes. Normality is an important assumption for two reasons. First, it is a natural assumption to make when a Gauss-Hermite rule will be used in numerically solving equation (1). Second, it facilitates change of variables transformations which allow any parameterization of the density to be obtained from a standardized form of the density. While other distributional assumptions could be made, the results presented here might be sensitive to those assumptions.

The stationary joint distribution of ξ_t will also be important later in the paper. If the unconditional mean and covariance matrix of ξ_t are denoted μ_ξ and Σ_ξ, respectively, then ξ_t has a multivariate normal stationary distribution, which is written as

$$\tilde{f}_s(\xi_t) = (2\pi)^{-S/2}|\Sigma_\xi|^{-1/2}\exp[-\tfrac{1}{2}(\xi_t - \mu_\xi)'\Sigma_\xi^{-1}(\xi_t - \mu_\xi)].$$

Suppose, in general, that the economic model in question yields the following Euler equation:

$$\tilde{v}_t = E_t K(\xi_{t+1}, \xi_t; \theta_2)(1 + \tilde{v}_{t+1}), \tag{2}$$

where $\theta_2 \in \Theta_2 \subset \mathbf{R}^r$, Θ_2 compact, is some unknown parameter vector for the economic model, and the function $K: \mathbf{R}^{2S} \times \Theta_2 \to \mathbf{R}$, is continuous in all arguments. Both Lucas's [17] model and Mehra and Prescott's [20] model can be written in the same form as (2). The number of equations is unimportant, and the term $1 + v$ could be generalized to any function that is linear in the unknown function $\tilde{v}$. In what follows, the main concern is to show under what assumptions approximate solutions to (2) converge when this equation is solved numerically.

Let the combined parameter space be denoted $\Theta = \Theta_1 \times \Theta_2$, and let θ denote the stacked vector $(\theta_1' \theta_2')'$.

Assuming that there exists a function $\tilde{v}: \mathbf{R}^S \times \Theta \to \mathbf{R}$ which is a solution to (2), it can be written as

$$\tilde{v}(\xi_t;\theta) - \int_{\mathbf{R}^S} K(\xi_{t+1},\xi_t;\theta_2)\tilde{v}(\xi_{t+1};\theta)\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)\, d\xi_{t+1}$$

$$= \int_{\mathbf{R}^S} K(\xi_{t+1},\xi_t;\theta_2)\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)\, d\xi_{t+1}. \tag{3}$$

What follows is a description of a solution method, which can be used in solving equation (3). The properties of the approximate solutions yielded by this method will be examined.

Equation (3) can be expressed as

$$\tilde{v}(\xi_t;\theta) - \int_{\mathbf{R}^S} K(\xi_{t+1},\xi_t;\theta_2)\frac{\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)}{\tilde{f}_0(\xi_{t+1};\theta_1)}\, \tilde{v}(\xi_{t+1};\theta)\tilde{f}_0(\xi_{t+1};\theta_1)\, d\xi_{t+1}$$

$$= \int_{\mathbf{R}^S} K(\xi_{t+1},\xi_t;\theta_2)\frac{\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)}{\tilde{f}_0(\xi_{t+1};\theta_1)}\, \tilde{f}_0(\xi_{t+1};\theta_1)\, d\xi_{t+1}, \tag{4}$$

where $\tilde{f}_0(\xi_{t+1},\theta_1) = \tilde{f}(\xi_{t+1} \mid \mu_\xi;\theta_1)$. $\tilde{f}_0$ is introduced because the solution method requires a weighting function that is dependent only on ξ_{t+1}. Tauchen and Hussey [24] show that $\tilde{f}_0$ is a good choice because of its shape relative to both the conditional distribution and the unconditional distribution of ξ_t.

Letting $\tilde{\psi}(\xi_{t+1},\xi_t;\theta) = K(\xi_{t+1},\xi_t;\theta_2)\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)/\tilde{f}_0(\xi_{t+1};\theta_1)$, equation (4) can be rewritten as

$$\tilde{v}(\xi_t;\theta) - \int_{\mathbf{R}^S} \tilde{\psi}(\xi_{t+1},\xi_t;\theta)\tilde{v}(\xi_{t+1};\theta)\tilde{f}_0(\xi_{t+1};\theta_1)\, d\xi_{t+1}$$

$$= \int_{\mathbf{R}^S} \tilde{\psi}(\xi_{t+1},\xi_t;\theta)\tilde{f}_0(\xi_{t+1};\theta_1)\, d\xi_{t+1}. \tag{5}$$

The method that will be used to solve equation (5) is Gaussian numerical quadrature which involves approximating an integral by a finite sum, for example,

$$\int_a^b K(y)f(y)\, dy \approx \sum_{i=1}^{n} K(y_{i,n})\omega_{i,n},$$

where the weights, $\omega_{i,n}$, and the abscissas, $y_{i,n}$, depend directly on the shape of the weighting function f. Thorough discussions of the properties of Gaussian quadrature rules are to be found in Szegö [22] and Tauchen and Hussey [24]. Since the density of the forcing process ξ_t is Gaussian, it is appropriate to use a Gauss-Hermite quadrature rule. Rather than repeat many of the properties of Gaussian quadrature, reference will be made, when necessary, to Szegö [22]. A slight modification to Gaussian quadrature is required for a proof which follows later.

In order to prepare equation (5) for Gaussian quadrature, it is useful to rewrite it in terms of random vectors with standard multivariate normal distributions. Define the change of variables transformation $\nu: \mathbf{R}^S \times \Theta_1 \to \mathbf{R}^S$, with $\nu(\xi_t, \theta_1) = C'^{-1}(\xi_t - \mu_\xi)$, where C is the Cholesky decomposition of Σ_ϵ. Therefore, $\nu(\xi_t, \theta_1)$ is continuous in both ξ_t and θ_1.

Then equation (5) can be rewritten:

$$v(x;\theta) - \int_{\mathbf{R}^S} \psi(y,x;\theta) v(y;\theta) f_0(y)\, dy = \int_{\mathbf{R}^S} \psi(y,x;\theta) f_0(y)\, dy, \tag{6}$$

where $y = \nu(\xi_{t+1}, \theta_1)$, $x = \nu(\xi_t, \theta_1)$, $v(x;\theta) = \tilde{v}[\nu^{-1}(x;\theta_1);\theta]$, and $\psi(y,x;\theta) = \tilde{\psi}[\nu^{-1}(y,\theta_1), \nu^{-1}(x,\theta_1);\theta]$. The function $\psi(y,x;\theta)$ is continuous in its arguments under the assumptions that were made about the functions K, $\tilde{f}$, and ν. The function $f_0(y) = (2\pi)^{-S/2} \exp(-\frac{1}{2}y'y)$ is the multivariate normal distribution function with mean zero and variance-covariance matrix equal to the identity matrix of dimension S. Another function that will be encountered later is the stationary joint distribution of ξ_t multiplied by the Jacobian of the inverse of the transformation ν. This is written as

$$\begin{aligned} f_s(x;\theta_1) &= \tilde{f}_s[\nu^{-1}(x,\theta_1);\theta_1]\,|C'| \\ &= (2\pi)^{-S/2} |\Sigma_x|^{-1/2} \exp(-\tfrac{1}{2}x'\Sigma_x^{-1}x), \end{aligned}$$

where $\Sigma_x = C^{-1\prime}\Sigma_\xi C^{-1}$.

The particular form of $f_0(y)$ is very convenient. As Tauchen and Hussey [24] state, if the multivariate density function $f_0(y)$ can be written as the product of univariate density functions, the theory of numerical quadrature with respect to f_0 is considerably simpler. This is because the theory for product rules borrows directly from the theory for univariate quadrature rules, since multivariate Gaussian integration rules can be constructed as the product of univariate Gaussian rules. Here, the function $f_0(y)$ can be written $f_0(y) = \prod_{i=1}^S f_0^i(y_i) = \prod_{i=1}^S (2\pi)^{-1/2} \exp(-\frac{1}{2}y_i^2)$, where y_i is the ith element of the vector y. Note that there is no assumption as to the independence of the underlying economic variables. The model has been transformed so that the equations are written in terms of normalized random variables that have no economic interpretation.

The functions $f_0^i(y_i)$ satisfy the basic assumptions necessary for the construction of a Gauss-Hermite integration rule (see Szegö [22]). These assumptions are satisfied since, for all i, $\int_{-\infty}^{+\infty} f_0^i(y_i)\,dy_i = \int_{-\infty}^{+\infty}(2\pi)^{-1/2}\exp(-\frac{1}{2}y_i^2)\,dy_i = 1 > 0$ and $\int_{-\infty}^{+\infty} y_i^n f_0^i(y_i)\,dy_i = \int_{-\infty}^{+\infty}(2\pi)^{-1/2} y_i^n \exp(-\frac{1}{2}y_i^2)\,dy_i$ exists for all non-negative integers n.

To approximate the integrals in equation (6), a Gauss-Hermite rule is used which, for each of the state variables, $i = 1, \ldots, S$, divides the real line into N_i adjacent intervals $(z_{i,1}, z_{i,2}), [z_{i,j}, z_{i,j+1})$, $j = 2, \ldots, N_i$, based on the weighting function f_0^i. The abscissas for this integration rule, $y_{i,j}$, must then lie within the subintervals in such a way that

$$-\infty = z_{i,1} < y_{i,1} < z_{i,2} < y_{i,2} < \cdots < z_{i,N_i} < y_{i,N_i} < z_{i,N_i+1} = \infty, \qquad i = 1, \ldots, S.$$

A property of Gauss-Hermite quadrature rules is that $|z_{i,j} - z_{i,j+1}|$ converges to zero for $j = 2, \ldots, N_i - 1$ as N_i goes to infinity. Furthermore, $\lim_{N_i \to \infty} z_{i,1} = -\infty$ and $\lim_{N_i \to \infty} z_{i,N_i+1} = \infty$ (see Szegö [22]).

In a proof to follow later in this paper, a slight modification to this rule becomes necessary. Let M be some large positive constant. Then, if either $x_{i,j} > M$ or $z_{i,j} < M < z_{i,j+1} < \infty$, the abscissa at which the function is evaluated is not taken to be $y_{i,j}$, rather it is taken to be $z_{i,j}$. Similarly, if either $z_{i,j+1} < -M$ or $-\infty < z_{i,j} < -M < z_{i,j+1}$, then the abscissa at which the function is evaluated is not taken to be $y_{i,j}$, rather it is taken to be $z_{i,j+1}$. The necessity of this modification will be discussed later.

Let

$$\Xi^{i,j} = \begin{cases} (z_{i,j}, z_{i,j+1}) & j = 1, \\ [z_{i,j}, z_{i,j+1}) & j = 2, \ldots, N_i. \end{cases}$$

In what follows step functions will be defined over the state space using the subintervals above. Since this would involve subscripts running over the S elements of y, an invertible mapping, which reduces the number of subscripts, is useful. It provides a more compact way of enumerating the different regions over which these step functions will be defined. Define $J: \mathbf{N}^S \to \mathbf{N}$ such that

$$J(i_1, i_2, \ldots, i_S) = i_1 + (i_2 - 1)N_1 + (i_3 - 1)N_2N_1 + \cdots + (i_S - 1)N_{S-1}N_{S-2}\cdots N_1$$

and define $N = N_1N_2\cdots N_S$ and $\bar{N} = (N_1, N_2, \ldots, N_S)$. Since each i_r runs from 1 to N_r, $r = 1, \ldots, S$, the subscript j will run from 1 to N. The following definitions are also useful.

$$y_j = (y_{1,i_1}, y_{2,i_2}, \ldots, y_{S,i_S})' \qquad \text{if} \quad j = J(i_1, i_2, \ldots, i_S)$$

$$x_j = (y_{1,i_1}, y_{2,i_2}, \ldots, y_{S,i_S})' \qquad \text{if} \quad j = J(i_1, i_2, \ldots, i_S)$$

$$\Xi_j = \Xi^{1,i_1} \times \Xi^{2,i_2} \times \cdots \times \Xi^{S,i_S} \qquad \text{if} \quad j = J(i_1, i_2, \ldots, i_S)$$

$$\mathbf{1}_j(y) = \begin{cases} 1 & \text{if } y \in \Xi_j \\ 0 & \text{otherwise.} \end{cases}$$

With these definitions it is possible to define a step function to approximate ψ:

$$\psi_{\bar{N}}(y,x;\theta) = \sum_{h=1}^{N} \sum_{j=1}^{N} \psi(y_j, x_h;\theta)\mathbf{1}_h(x)\mathbf{1}_j(y).$$

Now, related to equation (6), which is defined in terms of ψ, there is an integral equation corresponding to the approximation $\psi_{\bar{N}}$:

$$v_{\bar{N}}(x;\theta) - \int_{\mathbf{R}^S} \psi_{\bar{N}}(y,x;\theta) v_{\bar{N}}(y;\theta) f_0(y)\, dy = \int_{\mathbf{R}^S} \psi_{\bar{N}}(y,x;\theta) f_0(y)\, dy. \tag{7}$$

That a solution to (7) exists will be established later. For the moment, suppose that the solution is of the following form:

$$v_{\bar{N}}(x;\theta) = \sum_{h=1}^{N} a_h(\theta)\mathbf{1}_h(x). \tag{8}$$

Substitute (8) into (7), yielding

$$\sum_{h=1}^{N} a_h(\theta)\mathbf{1}_h(x) - \int_{\mathbf{R}^S} \sum_{h=1}^{N} \sum_{j=1}^{N} \psi(y_j, x_h;\theta) a_j(\theta)\mathbf{1}_h(x)\mathbf{1}_j(y) f_0(y)\, dy$$

$$= \int_{\mathbf{R}^S} \sum_{h=1}^{N} \sum_{j=1}^{N} \psi(y_j, x_h;\theta)\mathbf{1}_h(x)\mathbf{1}_j(y) f_0(y)\, dy.$$

Since all the functions except f_0 are step functions, the integral can be brought inside the summation signs to give

$$\sum_{h=1}^{N} a_h(\theta)\mathbf{1}_h(x) - \sum_{h=1}^{N} \sum_{j=1}^{N} \psi(y_j, x_h;\theta) a_j(\theta)\mathbf{1}_h(x)\omega_j$$

$$= \sum_{h=1}^{N} , \sum_{j=1}^{N} \psi(y_j, x_h;\theta)\mathbf{1}_h(x)\omega_j, \tag{9}$$

where $\omega_j = \int_{\Xi_j} f_0(y)\, dy$. Since (9) must hold for all x, the first summation sign can be eliminated.

$$a_h(\theta) - \sum_{j=1}^{N} \psi(y_j, x_h; \theta) a_j(\theta) \omega_j = \sum_{j=1}^{N} \psi(y_j, x_h; \theta) \omega_j.$$

Thus, the approximate integral equation reduces to a set of N linear equations in the N unknowns $a_h(\theta)$, $h = 1, \ldots, N$, that is, the solution is given by solving

$$(I - \Upsilon) a = \Upsilon 1,$$

where $\Upsilon_{hj} = \psi(y_j, x_h; \theta)\omega_j$, a is a vector containing the coefficients a_h, and 1 is a vector of ones. Later, in Proposition 7, a sufficient condition for the existence and uniqueness of the solution to (7) is provided. Proposition 7 also establishes that the solution is of the form (8) with

$$a = (I - \Upsilon)^{-1} \Upsilon 1.$$

Tauchen and Hussey [24] solve a slightly different set of equations because they construct a probability transition matrix for the discrete state space. Since the concern here is with finding numerical approximations to functions, there is no need to construct a discrete state space, which might be convenient for Monte-Carlo simulation. To discuss the convergence of solutions of (7) to the solutions of (6) requires the application of some of the theory of linear operators on spaces of functions.

Define a linear operator Ψ_θ as follows:

$$\Psi_\theta v = \int_{\mathbf{R}^S} \psi(y, x; \theta) v(y; \theta) f_0(y)\, dy.$$

Similarly, define an operator $\Psi_{\bar{N},\theta}$ with

$$\Psi_{\bar{N},\theta} v = \int_{\mathbf{R}^S} \psi_{\bar{N}}(y, x; \theta) v(y; \theta) f_0(y)\, dy.$$

In operator notation, equation (6) can be written as

$$(\mathrm{I} - \Psi_\theta) v = \Psi_\theta 1, \tag{10}$$

where 1 is the constant function equal to 1 for all $(x, y; \theta)$ and I is the identity linear operator which maps any function to itself. Similarly, equation (7) can be written as

$$(\mathrm{I} - \Psi_{\bar{N},\theta}) v_{\bar{N}} = \Psi_{\bar{N},\theta} 1. \tag{11}$$

Most of the remainder of this section is concerned with determining under what assumptions solutions to (10) and (11) exist and whether the solution to (11) can be made to converge to the solution to (10) by increasing N_1 through N_S.

Define two function spaces

$$L_1^2 = \left\{ g_1(x,y) \,\middle|\, \int_{\mathbf{R}^S} \int_{\mathbf{R}^S} |g_1(x,y)|^2 f_0(y) f_0(x)\, dy\, dx < \infty \right\}$$

and

$$L_2^2 = \left\{ g_2(x) \,\middle|\, \int_{\mathbf{R}^S} |g_2(x)|^2 f_0(x)\, dx < \infty \right\}.$$

These spaces are normed linear vector spaces, with the norms

$$\|g_1\|_1 = \left\{ \int_{\mathbf{R}^S} \int_{\mathbf{R}^S} |g_1(x,y)|^2 f_0(y) f_0(x)\, dy\, dx \right\}^{1/2}$$

and

$$\|g_2\|_2 = \left\{ \int_{\mathbf{R}^S} |g_2(x)|^2 f_0(x)\, dx \right\}^{1/2}. \qquad \blacksquare$$

Assumption 10. For each $\theta \in \Theta$, $\psi(y,x;\theta) \in L_1^2$.

For fixed θ, define the set of functions $\mathcal{V}_\theta = \{ v(x;\theta) \mid v \in L_2^2 \}$.

PROPOSITION 4. *$\mathcal{V}_\theta$ is a normed linear vector space, and under Assumption* 10, *the operator $\Psi_\theta : \mathcal{V}_\theta \to \mathcal{V}_\theta$.*

Proof. The first statement follows easily by the definition of a linear space. The inner product of two functions v_1 and v_2 in $\mathcal{V}_\theta$ can be defined to be

$$(v_1, v_2) = \int_{\mathbf{R}^S} v_1(x;\theta) v_2(x,\theta) f_0(x)\, dx.$$

Similarly, the norm of any function $v \in \mathcal{V}_\theta$ can be defined as

$$\|v\|_{\mathcal{V}_\theta} = (v,v)^{1/2} = \|v\|_2.$$

To show the second part, consider the function

$$h(x;\theta) = \Psi_\theta v(y,\theta) = \int_{\mathbf{R}^S} \psi(y,x;\theta) v(y;\theta) f_0(y)\, dy.$$

Then, using the Schwarz inequality in the second step (see Wouk [26]),

$$\begin{aligned} h(x;\theta)^2 &= \left\{ \int_{\mathbf{R}^S} \psi(y,x;\theta) v(y;\theta) f_0(y)\, dy \right\}^2 \\ &\le \left\{ \int_{\mathbf{R}^S} \psi(y,x;\theta)^2 f_0(y)\, dy \right\} \left\{ \int_{\mathbf{R}^S} v(y;\theta)^2 f_0(y)\, dy \right\}, \qquad \forall x. \end{aligned}$$

Now, all that remains to be proven is that $h(x;\theta)$ is square integrable with respect to f_0.

$$\int_{\mathbf{R}^S} h(x;\theta)^2 f_0(x)\,dx \le \left\{\int_{\mathbf{R}^S}\int_{\mathbf{R}^S} \psi(y,x;\theta)^2 f_0(y) f_0(x)\,dy\,dx\right\}$$
$$\times\left\{\int_{\mathbf{R}^S} v(y;\theta)^2 f_0(y)\,dy\right\}$$
$$\le \|\psi(y,x;\theta)\|_1^2\,\|v(y;\theta)\|_2^2 < \infty.$$

Therefore, $h(x;\theta) \in \mathcal{V}_\theta$. ∎

The norm of the operator Ψ_θ is defined as

$$\|\Psi_\theta\| = \sup_{\{v\,|\,\|v\|_{\mathcal{V}}=1\}} \|\Psi v\|_{\mathcal{V}_\theta}.$$

It follows from the proof of Proposition 4 that

$$\|\Psi_\theta\| = \sup_{\{v\,|\,\|v\|_{\mathcal{V}}=1\}}\left\{\int_{\mathbf{R}^S}\left|\int_{\mathbf{R}^S}\psi(y,x;\theta)v(y;\theta)f_0(y)\,dy\right|^2 f_0(x)\,dx\right\}^{1/2}$$
$$\le \sup_{\{v\,|\,\|v\|_{\mathcal{V}}=1\}}\|\psi(y,x;\theta)\|_1\|v(y;\theta)\|_2$$
$$= \|\psi(y,x,\theta)\|_1.$$

This implies that the operator Ψ_θ is a bounded operator, in that its norm is finite.

Assumption 11. Assume that there exists some square integrable function $h(y,x)$ such that $h(y,x) > \psi(y,x;\theta)$, $\forall\theta \in \Theta$.

Under Assumption 11, Ψ_θ is uniformly bounded in θ. That is to say, $\|\Psi_\theta\| \le m < \infty$, for all θ.

In order to ensure the existence of a solution to (10), the following assumption is useful.

Assumption 12. Assume that

$$\int_{\mathbf{R}^S}\int_{\mathbf{R}^S}\psi(y,x;\theta)^2 f_0(y)f_0(x)\,dy\,dx \le (1-\delta)^2,$$

for all $\theta \in \Theta$, for some $0 < \delta < 1$. This implies that $\sup_{\theta\in\Theta}\|\Psi_\theta\| \le |1-\delta|$. ∎

Assumptions 10–12 implicitly put restrictions on the parameter vector θ. These restrictions are illustrated in Section 4, for the asset pricing example.

PROPOSITION 5 [EXISTENCE AND UNIQUENESS OF A SOLUTION TO (10)]. *From Assumption* 12, *the inverse of the operator* $(\mathrm{I} - \Psi_\theta)$, *denoted* $(\mathrm{I} - \psi_\theta)^{-1}$, *exists and, for each* θ, *there is a unique solution to equation* (10) *which satisfies* $v = (\mathrm{I} - \Psi_\theta)^{-1}\Psi_\theta 1$.

Proof. By the Geometric Series Theorem for bounded linear operators on Banach spaces, (Atkinson [2], Baker [3]),

$$(\mathrm{I} - \Psi_\theta)^{-1} = \sum_{i=0}^{\infty} \Psi_\theta^i.$$

Therefore, a solution to (10) may be written as

$$v = \left(\sum_{i=0}^{\infty} \Psi_\theta^i\right)\Psi_\theta 1 + \Psi_\theta 1 + \Psi_\theta^2 1 + \Psi_\theta^3 1 + \cdots.$$

To see that this solution is unique, consider another solution v'. Then $v - v' = \Psi_\theta(v - v')$. This implies that $\| v - v' \| \leq \| \Psi_\theta \| \, \| v - v' \|$. But since $\| \Psi_\theta \| < 1$, this implies a contradiction. Thus, the solution must be unique. ∎

In order to prove the existence and uniqueness of $v_{\bar{N}}$ for a sufficiently fine approximation, it is convenient to use another change of variables transformation. Define the invertible function $\Lambda(x): \mathbf{R}^S \to (0,1)^S$ such that if $\hat{x} = \Lambda(x)$, then $\hat{x}_i = e^{x_i}/(1 + e^{x_i})$. Let $\hat{x} = \Lambda(x)$ and $\hat{y} = \Lambda(y)$. Let

$$\hat{J}(\hat{x},\hat{y}) = \begin{cases} \prod_{i=1}^{S} [\hat{x}_i(1 - \hat{x}_i)\hat{y}_i(1 - \hat{y}_i)]^{-1} & \\ & \text{if } \; 0 < \hat{x}_i < 1,\ 0 < \hat{y}_i < 1, \qquad \text{for all } i \\ 0 & \text{otherwise.} \end{cases}$$

Also define the set $\mathbf{U} = [0,1]$. Then

$$\int_{\mathbf{R}^S}\int_{\mathbf{R}^S} \psi(y,x;\theta)^2 f_0(y) f_0(x)\,dy\,dx$$
$$= \int_{\mathbf{U}^S}\int_{\mathbf{U}^S} \hat{\psi}(\hat{y},\hat{x};\theta)^2 \hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y})\,d\hat{y}\,d\hat{x},$$

and

$$\int_{\mathbf{R}^S}\int_{\mathbf{R}^S} \psi_{\bar{N}}(y,x;\theta)^2 f_0(y) f_0(x)\,dy\,dx$$
$$= \int_{\mathbf{U}^S}\int_{\mathbf{U}^S} \hat{\psi}_{\bar{N}}(\hat{y},\hat{x};\theta)^2 \hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y})\,d\hat{y}\,d\hat{x},$$

where

$$\hat{\psi}(\hat{y},\hat{x};\theta) = \begin{cases} \psi(\Lambda^{-1}(\hat{y}),\Lambda^{-1}(\hat{x});\theta) & \\ \qquad \text{if} \quad 0 < \hat{y}_i < 1 \quad \text{and} \quad 0 < \hat{x}_i < 1, & \text{for all} \quad i, \\ 0 \qquad \text{otherwise,} & \end{cases}$$

$$\hat{\psi}_{\bar{N}}(\hat{y},\hat{x};\theta) = \begin{cases} \psi_{\bar{N}}(\Lambda^{-1}(\hat{y}),\Lambda^{-1}(\hat{x});\theta) & \\ \qquad \text{if} \quad 0 < \hat{y}_i < 1 \quad \text{and} \quad 0 < \hat{x}_i < 1, & \text{for all} \quad i, \\ 0 \qquad \text{otherwise,} & \end{cases}$$

$$\hat{f}_0(\hat{y}) = \begin{cases} f_0(\Lambda^{-1}(\hat{y})) \quad \text{if} \quad 0 < \hat{y}_i < 1, & \text{for all} \quad i, \\ 0 \qquad\qquad\quad \text{otherwise.} & \end{cases}$$

It is clear from the definition of Λ that it preserves continuity (wherever it exists) in the interior of $\mathbf{U}^S$. However, to obtain continuity of $\hat{\psi}^2\hat{f}_0\hat{f}_0\hat{J}$ at the endpoints, it is sufficient to make the following assumption.

Assumption 13. Assume that

$$\lim_{\omega\to\infty} h(aw,bw)^2 f_0(aw) f_0(bw)\hat{J}[\Lambda(bw),\Lambda(aw)] = 0, \qquad \forall a,b$$

where h is the function referred to in Assumption 11, and a and b are directional vectors. ∎

PROPOSITION 6. *Let* $\hat{N} = \min\{N_1,N_2,\ldots,N_S\}$. *Then* $\sup_{\theta\in\Theta}\|\Psi_{\bar{N},\theta}\| \leq 1 - d$, *for some* $0 < d < 1$ *and* $\hat{N}$ *sufficiently large.*

Proof

$$\|\Psi_{\bar{N},\theta}\| \leq \left\{\int_{\mathbf{R}^S}\int_{\mathbf{R}^S} \psi_{\bar{N}}(y,x;\theta)^2 f_0(y) f_0(x)\,dy\,dx\right\}^{1/2}$$
$$= \left\{\int_{\mathbf{U}^S}\int_{\mathbf{U}^S} \hat{\psi}_{\bar{N}}(\hat{y},\hat{x};\theta)^2 \hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y})\,d\hat{y}\,d\hat{x}\right\}^{1/2}.$$

The transformation Λ is such that the fact that quadrature intervals on the real line narrow as $\hat{N}\to\infty$ is preserved when these intervals are mapped to the unit interval.[3] Recall the modified Gaussian quadrature rule. It follows from the properties of Gaussian quadrature rules that for $\hat{N}$ sufficiently large, there is at least one quadrature interval in each dimension completely to the right of M and one completely to the left of $-M$. On the unit cube this translates into at least one quadrature interval in each dimension completely to the right of $\Lambda(M)$ and one completely to the left of $\Lambda(-M)$. Furthermore, for any $\epsilon > 0$, there exist points $\hat{x}_{i1} < \Lambda(-M)$, $\hat{x}_{i2} > \Lambda(M)$, $\hat{y}_{i1} < \Lambda(-M)$, and $\hat{y}_{i2} > \Lambda(M)$ for $i = 1, \ldots, S$ such that

$$\hat{\psi}(\hat{y},\hat{x};\theta)^2\hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y}) < \epsilon/2$$

if $(\hat{x},\hat{y}) \notin \mathbf{X} \times \mathbf{Y}$, where

$$\mathbf{X} = [\hat{x}_{11},\hat{x}_{12}] \times [\hat{x}_{21},\hat{x}_{22}] \times \cdots \times [\hat{x}_{S1},\hat{x}_{S2}]$$

and

$$\mathbf{Y} = [\hat{y}_{11},\hat{y}_{12}] \times [\hat{y}_{21},\hat{y}_{22}] \times \cdots \times [\hat{y}_{S1},\hat{y}_{S2}].$$

This last statement follows from the continuity of all the functions in the expression and from Assumption 13. To simplify notation, let $\mathbf{Z} = \mathbf{X} \times \mathbf{Y}$ and let $\bar{\mathbf{Z}}$ denote the complement of $\mathbf{Z}$ in $\mathbf{U}^S \times \mathbf{U}^S$. Then

$$\begin{aligned}\int_{\mathbf{U}^S}\int_{\mathbf{U}^S} \hat{\psi}_{\hat{N}}^2 \hat{f}_0 \hat{f}_0 \hat{J}\, d\hat{y}\, d\hat{x} &= \iint_{\mathbf{Z}} \hat{\psi}_{\hat{N}}^2 \hat{f}_0 \hat{f}_0 \hat{J}\, d\hat{y}\, d\hat{x} + \iint_{\bar{\mathbf{Z}}} \hat{\psi}_{\hat{N}}^2 \hat{f}_0 \hat{f}_0 \hat{J}\, d\hat{y}\, d\hat{x} \\ &\le \iint_{\mathbf{Z}} \hat{\psi}_{\hat{N}}^2 \hat{f}_0 \hat{f}_0 \hat{J}\, d\hat{y}\, d\hat{x} + \frac{\epsilon}{2}.\end{aligned}$$

The second step above is possible since $\hat{\psi}_{\hat{N}}$ is equal to $\hat{\psi}$ at selected points. At those selected points, it follows from above that $\hat{\psi}_{\hat{N}}^2\hat{f}_0\hat{f}_0\hat{J} = \hat{\psi}^2\hat{f}_0\hat{f}_0\hat{J} < \epsilon/2$. The function $\hat{f}_0\hat{f}_0\hat{J}$ is monotonically decreasing toward 0 or 1 and the quadrature rule has been modified beyond $\pm M$ so that the abscissas are the inner endpoints of the quadrature intervals. Therefore, $\hat{\psi}_{\hat{N}}^2\hat{f}_0\hat{f}_0\hat{J} < \hat{\psi}^2\hat{f}_0\hat{f}_0\hat{J} < \epsilon/2$ over the entire set $\bar{\mathbf{Z}}$. Notice also that this is true for all θ.

Since $\mathbf{Z}$ is compact, $\hat{\psi}^2$ is uniformly continuous over that set. Therefore, since $\hat{\psi}_{\hat{N}}$ is a step function and the quadrature intervals narrow, there is some $\hat{N}^*$ such that $|\hat{\psi}^2 - \hat{\psi}_{\hat{N}}^2| < \epsilon/2$ for all $\hat{N} > \hat{N}^*$, on $\mathbf{Z}$. $\hat{N}^*$ can be chosen independently of θ since $\hat{\psi}^2$ is continuous in θ, and Θ is a compact set. Thus,

$$\begin{aligned}\int_{\mathbf{U}^S}\int_{\mathbf{U}^S} \hat{\psi}_{\hat{N}}^2 \hat{f}_0 \hat{f}_0 \hat{J}\, d\hat{y}\, d\hat{x} &\le \iint_{\mathbf{Z}} \hat{\psi}^2 \hat{f}_0 \hat{f}_0 \hat{J}\, d\hat{y}\, d\hat{x} + \epsilon \\ &\le (1-\delta)^2 + \epsilon\end{aligned}$$

for $\hat{N} > \hat{N}^*$. This implies that $\|\Psi_{\hat{N},\theta}\|$ can be made uniformly less than 1, simply by making ϵ small and choosing $\hat{N}$ large enough. ■

This theorem implies that for sufficiently large $\hat{N}$, the operator $\Psi_{\hat{N},\theta}$ maps from $\mathcal{V}_\theta$ to $\mathcal{V}_\theta$ and is uniformly bounded. Similarly, the operator $(\Psi_\theta - \Psi_{\hat{N},\theta})$, with

$$(\Psi_\theta - \Psi_{\hat{N},\theta})v = \int_{\mathbf{R}^S} [\psi(y,x;\theta) - \psi_{\hat{N}}(y,x;\theta)]\, v(y;\theta) f_0(y)\, dy,$$

maps from $\mathcal{V}_\theta$ to $\mathcal{V}_\theta$ and inherits the property of uniform boundedness from Ψ_θ and $\Psi_{\bar{N},\theta}$.

PROPOSITION 7 [EXISTENCE AND UNIQUENESS OF A SOLUTION TO (11)]. *From Assumption* 13 *and Proposition* 6, *for sufficiently large* $\hat{N}$, *the inverse of the operator* $(\mathrm{I} - \Psi_{\bar{N},\theta})$, *denoted* $(\mathrm{I} - \Psi_{\bar{N},\theta})^{-1}$, *exists and, for each* θ, *there is a unique solution to equation* (11) *which satisfies* $v_{\bar{N}} = (\mathrm{I} - \Psi_{\bar{N},\theta})^{-1}\Psi_{\bar{N},\theta}1$. *Furthermore, the solution is a step function as in equation* (8).

Proof. The method of proof for the first part is identical to that for the proof of Proposition 5. The geometric series theorem implies that

$$v_{\bar{N}} = \left(\sum_{i=0}^{\infty} \Psi^i_{\bar{N},\theta}\right)\Psi_{\bar{N},\theta}1.$$

Notice that

$$\Psi_{\bar{N},\theta}1 = \sum_{h=1}^{N}\sum_{j=1}^{N} \psi(y_j,x_j;\theta)\mathbf{1}_h(x)\omega_j,$$

which is a step function in x. Successive applications of the operator $\Psi_{\bar{N},\theta}$ lead to more step functions. Therefore, $v_{\bar{N}}$ is a step function. ∎

Given that solutions to the equations (10) and (11) exist, it is naturally of interest to know whether solutions to the approximate integral equation (11) converge to the solution to equation (10). This will center on whether the function ψ can be approximated arbitrarily well in norm by the step function $\psi_{\bar{N}}$. That is to say, it will rely on demonstrating that

$$\int_{\mathbf{R}^S}\int_{\mathbf{R}^S} [\psi(y,x;\theta) - \psi_{\bar{N}}(y,x;\theta)]^2 f_0(y)f_0(x)\,dy\,dx$$

converges to zero.

PROPOSITION 8. *The function* $\psi_{\bar{N}}$ *converges in norm to the function* ψ *as* $\hat{N}\to\infty$. *That is, for any* $\epsilon > 0$, $\exists\bar{N}^*$, *such that* $\forall\hat{N} > \hat{N}^*$, $\|\psi - \psi_{\bar{N}}\| < \epsilon$. *In addition, this convergence is uniform in* θ. *That is to say,* $\exists\hat{N}^*$, *such that* $\forall\hat{N} > \hat{N}^*$, $\sup_{\theta\in\Theta}\|\psi - \psi_{\bar{N}}\| < \epsilon$.

Proof. As above, this problem is mapped to the unit $2S$-dimensional cube. Doing this as before, the integral

$$\int_{\mathbf{R}^S}\int_{\mathbf{R}^S} (\psi(y,x;\theta) - \psi_{\bar{N}}(y,x;\theta))^2 f_0(y)f_0(x)\,dy\,dx$$

can be written as

$$\int_{\mathbf{U}^S}\int_{\mathbf{U}^S} (\hat{\psi}(\hat{y},\hat{x};\theta) - \hat{\psi}_{\hat{N}}(\hat{y},\hat{x};\theta))^2 \hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y})\, d\hat{y}\, d\hat{x}.$$

As in the proof to Proposition 6, for any $\epsilon > 0$, a compact region $\mathbf{Z}$ can be defined such that

$$\hat{\psi}(\hat{y},\hat{x};\theta)^2 \hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y}) < \epsilon/3$$

and

$$\hat{\psi}_{\hat{N}}(\hat{y},\hat{x};\theta)^2 \hat{f}_0(\hat{y})\hat{f}_0(\hat{x})\hat{J}(\hat{x},\hat{y}) < \epsilon/3$$

for $(\hat{x},\hat{y}) \notin \mathbf{Z}$, for all $\theta \in \Theta$, and for $\hat{N}$ sufficiently large. It follows that for $\hat{N}$ sufficiently large, and for all θ

$$\begin{aligned}
\int_{\mathbf{U}^S}\int_{\mathbf{U}^S} (\hat{\psi} - \hat{\psi}_{\hat{N}})^2 \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} &= \iint_{\mathbf{Z}} (\hat{\psi} - \hat{\psi}_{\hat{N}})^2 \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} \\
&\quad + \iint_{\bar{\mathbf{Z}}} (\hat{\psi} - \hat{\psi}_{\hat{N}})^2 \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} \\
&< \iint_{\mathbf{Z}} (\hat{\psi} - \hat{\psi}_{\hat{N}})^2 \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} \\
&\quad + \iint_{\bar{\mathbf{Z}}} (\hat{\psi}^2 + \hat{\psi}_{\hat{N}}^2) \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} \\
&< \int\int_{\mathbf{Z}} (\hat{\psi} - \hat{\psi}_{\hat{N}})^2 \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} + \frac{2\epsilon}{3}.
\end{aligned}$$

Now, since $\mathbf{Z} \times \Theta$ is a compact set, $\hat{\psi}$ is uniformly continuous over that set. Therefore, there is some $\hat{N}^*$ such that $|\hat{\psi} - \hat{\psi}_{\hat{N}}| < \sqrt{\epsilon/3}$ for all $\hat{N} > \hat{N}^*$, on $\mathbf{Z} \times \Theta$. This $\hat{N}^*$ can be chosen independently of θ since θ lies in a compact set. Thus,

$$\int_{\mathbf{U}^S}\int_{\mathbf{U}^S} (\hat{\psi} - \hat{\psi}_{\hat{N}})^2 \hat{f}_0\hat{f}_0\hat{J}\, d\hat{x}\, d\hat{y} < \epsilon, \qquad \text{for} \quad \hat{N} > \hat{N}^*$$

for all θ. ■

PROPOSITION 9 (CONVERGENCE). *The approximate solution function $v_{\hat{N}}(x;\theta)$ converges in norm to the true solution function $v(x;\theta)$, uniformly in θ.*

Proof. Define $\varepsilon_{\bar{N}}(x;\theta) = v(x;\theta) - v_{\bar{N}}(x;\theta)$. It follows from (10) and (11) that $\varepsilon_{\bar{N}}$ satisfies the following integral equation:

$$(\mathrm{I} - \Psi_\theta)\varepsilon_{\bar{N}} = (\Psi_\theta - \Psi_{\bar{N},\theta})(1 + v_{\bar{N}})$$

$$\text{or} \quad \varepsilon_{\bar{N}} = (\mathrm{I} - \Psi_\theta)^{-1}(\Psi_\theta - \Psi_{\bar{N},\theta})(1 + v_{\bar{N}})$$

By the Geometric Series Theorem, $\|(\mathrm{I} - \Psi_\theta)^{-1}\| \le 1/(1 - \|\Psi_\theta\|)$. Similarly, $\|(\mathrm{I} - \Psi_{\bar{N},\theta})^{-1}\| \le 1/(1 - \|\Psi_{\bar{N},\theta}\|)$. Thus,

$$\begin{aligned}
\|\varepsilon_{\bar{N}}\| &\le \|(\mathrm{I} - \Psi_\theta)^{-1}\| \, \|(\Psi_\theta - \Psi_{\bar{N},\theta})\| [1 + \|(\mathrm{I} - \Psi_{\bar{N},\theta})^{-1}\| \, \|\Psi_{\bar{N},\theta}\|] \\
&\le \|(\Psi_\theta - \Psi_{\bar{N},\theta})\| \left[\frac{1}{1 - \|\Psi_\theta\|}\right]\left[1 + \frac{1}{1 - \|\Psi_{\bar{N},\theta}\|}\|\Psi_{\bar{N},\theta}\|\right] \\
&= \|(\Psi_\theta - \Psi_{\bar{N},\theta})\| \left[\frac{1}{1 - \|\Psi_\theta\|}\right]\left[\frac{1}{1 - \|\Psi_{\bar{N},\theta}\|}\right].
\end{aligned}$$

Choose some $\epsilon > 0$. It has been assumed above that $\sup_{\theta\in\Theta}\|\Psi_\theta\| \le 1 - \delta$. Furthermore, it has been established that for sufficiently large $\hat{N}$, $\sup_{\theta\in\Theta}\|\Psi_{\bar{N},\theta}\| \le 1 - d$. Proposition 8 establishes that for sufficiently large $\hat{N}$, $\sup_{\theta\in\Theta}\|\Psi_\theta - \Psi_{\bar{N},\theta}\| < \epsilon\delta d$. Therefore, there exists $\hat{N}$ sufficiently large, such that

$$\begin{aligned}
\sup_{\theta\in\Theta}\|\varepsilon_{\bar{N}}\| &\le \sup_{\theta\in\Theta}\|(\Psi_\theta - \Psi_{\bar{N},\theta})\| \sup_{\theta\in\Theta}\left(\frac{1}{1 - \|\Psi_\theta\|}\right) \sup_{\theta\in\Theta}\left(\frac{1}{1 - \|\Psi_{\bar{N},\theta}\|}\right) \\
&< \epsilon.
\end{aligned}$$

Uniformity of convergence is guaranteed by the fact that there are uniform bounds in θ for each of the norms on the right-hand side of the above inequality. ∎

For the purposes of estimation, the ultimate goal is to be able to calculate the moments of variables of interest. Therefore, to illustrate the possibilities, the first and second moments of the solution function $\tilde{v}$ will be discussed. In order to take advantage of the theory presented in Section 2, the moments of the approximate solutions, which will be obtained analytically, must converge uniformly. For example, the population mean of $\tilde{v}$ is given by

$$Ev = \int_{\mathbf{R}^S} \tilde{v}(\xi_t;\theta)\tilde{f}_s(\xi_t;\theta_1)\, d\xi_t,$$

where $\tilde{f}_s(\xi_t;\theta_1)$ is the stationary distribution of ξ_t, defined above. The result of the above integration, Ev, is a function of the parameter vector θ and will be called a moment function.

It is convenient to write the expectation in terms of the transformed variables. Some algebra shows that Ev can be expressed as

$$Ev = \int_{\mathbf{R}^S} v(x;\theta)f_s(x;\theta_1)\, dx,$$

where $f_s(x;\theta_1)$ is defined above. To approximate this expectation, the expectation of the approximating step-function is used.

$$Ev_{\bar{N}} = \int_{\mathbf{R}^S} v_{\bar{N}}(x;\theta) f_s(x;\theta_1)\, dx.$$

The difference between these two functions of θ will be denoted

$$\begin{aligned}
\Delta(\theta) &= \int_{\mathbf{R}^S} [v(x;\theta) - v_{\bar{N}}(x;\theta)]\,(2\pi)^{-S/2}|\Sigma_x|^{-1/2}\exp\left(-\frac{1}{2}\,x'\Sigma_x^{-1}x\right) dx \\
&= \int_{\mathbf{R}^S} [v(x;\theta) - v_{\bar{N}}(x;\theta)]|\Sigma_x|^{-1/2}\exp\left[-\frac{1}{2}\,x'(\Sigma_x^{-1} - I)x\right](2\pi)^{-S/2} \\
&\quad \times \exp\left(-\frac{1}{2}\,x'x\right) dx \\
&\le \int_{\mathbf{R}^S} |v(x;\theta) - v_{\bar{N}}(x;\theta)|\ |\Sigma_x|^{-1/2}\exp\left[-\frac{1}{2}\,x'(\Sigma_x^{-1} - I)x\right](2\pi)^{-S/2} \\
&\quad \times \exp\left(-\frac{1}{2}\,x'x\right) dx.
\end{aligned}$$

Notice that this is the inner product of two positive functions in the space $\mathcal{V}_\theta$. Therefore, by the Schwarz inequality it follows that

$$\begin{aligned}
\Delta(\theta) &\le \left\{\int_{\mathbf{R}^S} [v(x;\theta) - v_{\bar{N}}(x;\theta)]^2(2\pi)^{-S/2}\exp\left(-\frac{1}{2}\,x'x\right) dx\right\}^{1/2} \\
&\quad \times \left\{\int_{\mathbf{R}^S} (2\pi)^{-S/2}|\Sigma_x|^{-1}\exp\left[-\frac{1}{2}\,x'(2\Sigma_x^{-1} - I)x\right] dx\right\}^{1/2} \\
&= \|v(x;\theta) - v_{\bar{N}}(x;\theta)\|\ |\Sigma_x|^{-1/2}|2\Sigma_x^{-1} - I|^{-1/4}.
\end{aligned}$$

Clearly the difference between the two moment functions converges uniformly to zero given the previous results. Notice that $2\Sigma_x^{-1} - I$ must be positive definite for the result to go through. This translates into restrictions on the set θ_1 which will be illustrated in the next section. Proving that the second moments of the solution function converge is similar, although it requires that from the beginning the theory be developed for functions lying in L^4.

This concludes the section describing the numerical solution method and its properties. In the next section, an example is presented to illustrate under what circumstances the assumptions made in this section are reasonable.

4. THE ASSET PRICING EXAMPLE

The example presented here is the asset pricing model used in the introduction. Assume that there exists one source of income to the consumer, a stochastic endowment of a single perishable consumption good. In equilibrium, consumption will equal this stochastic endowment. The asset to be priced is the right to the endowment stream in perpetuity. The forcing process for the example is equilibrium consumption which is equal to the value of the exogenous stochastic endowment. By assumption, the first difference of the logarithm of consumption at time t is ξ_t.[4]

To satisfy Assumptions 8 and 9, let ξ_t have a univariate AR(1) representation. That is, let

$$\xi_t = \mu + \rho\xi_{t-1} + \epsilon_t,$$

with ϵ_t distributed normally with mean zero and

$$E\epsilon_t\epsilon_{t-s} = \begin{cases} \sigma^2 & \text{if} \quad s = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the distribution of ξ_t conditional on ξ_{t-1} can be written

$$\tilde{f}(\xi_t \mid \xi_{t-1};\theta) = (2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{1}{2\sigma^2}(\xi_t - \mu - \rho\xi_{t-1})^2\right].$$

As in the introduction, the price-dividend ratio of the asset, $\tilde{v}_t$, must satisfy

$$\tilde{v}_t = \beta E_t \exp[(1-\gamma)\xi_{t+1}](1 + \tilde{v}_{t+1}). \tag{12}$$

Therefore, the vector of model parameters $\theta_2 = (\beta \quad \gamma)'$, while the vector of parameters $\theta_1 = (\mu \quad \rho \quad \sigma)'$.

Equation (12) can be rewritten as an integral equation

$$\tilde{v}(\xi_t;\theta) - \int_{\mathbf{R}} \beta\exp[(1-\gamma)\xi_{t+1}]\tilde{v}(\xi_{t+1};\theta)\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)\,d\xi_{t+1}$$
$$= \int_{\mathbf{R}} \beta\exp[(1-\gamma)\xi_{t+1}]\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)\,d\xi_{t+1}. \tag{13}$$

In the example, $E\xi_t = \mu_\xi = \mu/(1-\rho)$, so that

$$\tilde{f}_0(\xi_{t+1};\theta_1) = (2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{1}{2\sigma^2}(\xi_{t+1} - \mu_\xi)^2\right].$$

Thus, in the example

$$\begin{aligned}\tilde{\psi}(\xi_{t+1},\xi_t;\theta) &= \beta\exp[(1-\gamma)\xi_{t+1}]\tilde{f}(\xi_{t+1} \mid \xi_t;\theta_1)/\tilde{f}_0(\xi_{t+1};\theta_1) \\ &= \beta\exp\left[(1-\gamma)\xi_{t+1} - \frac{1}{2\sigma^2}(\xi_{t+1} - \mu - \rho\xi_t)^2\right. \\ &\qquad \left. + \frac{1}{2\sigma^2}(\xi_{t+1} - \mu_\xi)^2\right].\end{aligned}$$

Applying the change of variables transformation ν to the example, the function $\tilde{\psi}$ can be written in terms of the transformed variables, $y = (\xi_{t+1} - \mu_\xi)/\sigma$ and $x = (\xi_t - \mu_\xi)/\sigma$.

$$\psi(y,x;\theta) = \beta \exp[(1-\gamma)(\mu_\xi + \sigma y) - \tfrac{1}{2}(\rho^2 x^2 - 2\rho x y)],$$

and the normalized weighting function is

$$f_0(y) = (2\pi)^{-1/2} \exp(-\tfrac{1}{2} y^2).$$

With this notation in hand, it is possible to rewrite the integral equation (13) in terms of the normalized variables.

$$\begin{aligned} v(x;\theta) &- \int_{\mathbf{R}} \beta \exp\left[(1-\gamma)(\mu_\xi + \sigma y) - \tfrac{1}{2}(\rho^2 x^2 - 2\rho x y) \right] \\ &\times v(y;\theta)(2\pi)^{-1/2} \exp\left(-\frac{1}{2} y^2 \right) dy \\ &= \int_{\mathbf{R}} \beta \exp\left[(1-\gamma)(\mu_\xi + \sigma y) - \frac{1}{2}(\rho^2 x^2 - 2\rho x y) \right] \\ &\times (2\pi)^{-1/2} \exp\left(-\frac{1}{2} y^2 \right) dy. \end{aligned} \tag{14}$$

Approximate solutions are obtained by setting $v_N(x,\theta) = \sum_{h=1}^{N} a_h(\theta) \mathbf{1}_h(x)$, where

$$\begin{aligned} a_h(\theta) &- \sum_{j=1}^{N} \beta \exp\left[(1-\gamma)(\mu_\xi + \sigma y_j) - \frac{1}{2}(\rho^2 x_h^2 - 2\rho x_h y_j) \right] \\ &\times a_j(\theta) [\Phi(z_{j+1}) - \Phi(z_j)] \\ &= \sum_{j=1}^{N} \beta \exp\left[(1-\gamma)(\mu_\xi + \sigma y_j) - \frac{1}{2}(\rho^2 x_h^2 - 2\rho x_h y_j) \right] [\Phi(z_{j+1}) - \Phi(z_j)]. \end{aligned} \tag{15}$$

where the locations of the points x_h, y_j, z_j are determined using the modified Gauss-Hermite quadrature rule described above. The function Φ is the cumulative distribution function of the standard normal.

To write equation (14) in operator notation, define the linear operator

$$\begin{aligned} \Psi_\theta v = \int_{\mathbf{R}} \beta \exp&\left[(1-\gamma)(\mu_\xi + \sigma y) - \frac{1}{2}(\rho^2 x^2 - 2\rho x y) \right] \\ &\times v(y;\theta)(2\pi)^{-1/2} \exp\left(-\frac{1}{2} y^2 \right) dy. \end{aligned}$$

Then, just as above, equation (14) can be written

$$v - \Psi_\theta v = \Psi_\theta 1. \tag{16}$$

It is not possible to calculate the norm of the operator Ψ_θ explicitly. The norm of Ψ_θ is

$$\|\Psi_\theta\| = \sup_{\{v \mid \|v\|_v = 1\}} \left\{ \int_{\mathbf{R}} \left| \int_{\mathbf{R}} \beta \exp\left[(1-\gamma)(\mu_\xi + \sigma y) - \frac{1}{2}(\rho^2 x^2 - 2\rho x y) \right] \right. \right.$$
$$\left. \left. \times\, v(y;\theta)(2\pi)^{-1/2} \exp\left(-\frac{1}{2} y^2\right) dy \right|^2 (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right) dx \right\}^{1/2}$$

The difficulty in obtaining the norm analytically is that it is impossible to take the required sup over all functions of norm 1. This is unfortunate since it would be useful to know under what circumstances there exist square integrable solutions to equation (16). When $\|\Psi_\theta\| < 1$, the operator $(\mathrm{I} - \Psi_\theta)$ is invertible and there exists a unique solution to (16), for each θ. There is an interesting exercise, which shows the role of the norm of Ψ_θ in determining the existence of solutions for the price-dividend ratios. Imagine that the parameter θ_1 were known. Then, it is clear from the definition of the norm given above that it would be a function of β and γ alone. What is more, in this case, it is possible to obtain lower and upper bounds for the norm. A lower bound can be obtained by substituting the constant function 1, which has norm 1, for the function v in the formula for $\|\Psi_\theta\|$. This lower bound is analytic.

$$\|\Psi_\theta\|_{lb} = \left\{ \int_{\mathbf{R}} \left| \int_{\mathbf{R}} \beta \exp\left[(1-\gamma)(\mu_\xi + \sigma y) - \frac{1}{2}(\rho^2 x^2 - 2\rho x y) \right] \right. \right.$$
$$\left. \left. \times\, (2\pi)^{-1/2} \exp\left(-\frac{1}{2} y^2\right) dy \right|^2 (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right) dx \right\}^{1/2}$$
$$= \beta \exp[(1-\gamma)\mu_\xi + \tfrac{1}{2}(1-\gamma)^2\sigma^2(1+2\rho^2)].$$

The upper bound to the norm is obtainable from the Schwarz inequality, which implies that the norm of Ψ_θ is less than

$$\|\Psi_\theta\|_{ub} = \left\{ \int_{\mathbf{R}} \int_{\mathbf{R}} \beta^2 \exp[2(1-\gamma)(\mu_\xi + \sigma y) - \rho^2 x^2 + 2\rho x y)] \right.$$
$$\left. \times\, (2\pi)^{-1/2} \exp\left(-\frac{1}{2} y^2\right) (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right) dy\, dx \right\}^{1/2}$$
$$= \frac{\beta \exp[(1-\gamma)\mu_\xi + (1-\gamma)^2\sigma^2(1+2\rho^2)/(1-2\rho^2)]}{\sqrt{1-2\rho^2}}$$

For this upper bound to exist, and to guarantee that Ψ_θ is invertible and that any solution to (16) is square integrable, requires that $|\rho| < 1/\sqrt{2}$. This follows from the fact that the marginal rate of substitution was multiplied by the ratio of $\tilde{f}/\tilde{f}_0$ when creating the function $\tilde{\psi}$.

Of course, the parameters θ_1 are unknown. However, they can be estimated consistently if time-series on consumption growth, or equivalently dividend growth, are available. To this end, the annual data used by Mehra and Prescott [20], for real U.S. consumption growth between 1889 and 1979, are used here to estimate θ_1. The estimated parameter values are $\mu = .02$, $\rho = -.14$, and $\sigma = .035$. Now, taking these parameter values as the elements of the true θ_1, a graph of the upper and lower bounds for the norm as functions of β and γ can be drawn. Such a graph is presented in Figure 1. Note that the curve labeled "upper bound" defines the locus of points at which

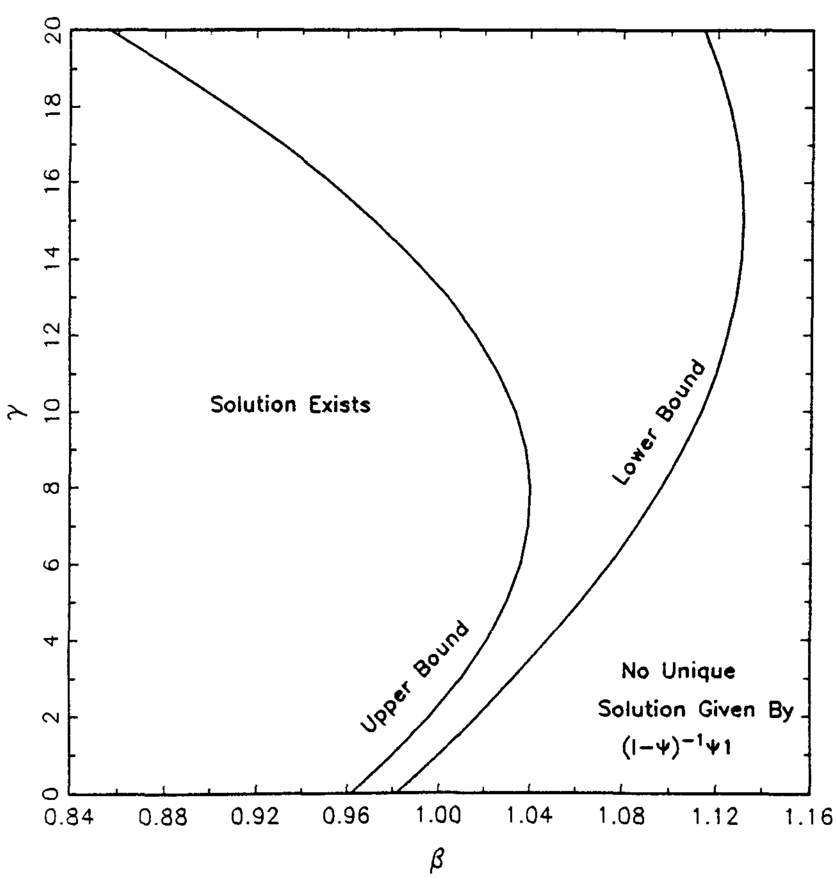

FIGURE 1. The norm of Ψ_θ.

$\| \Psi_\theta \|_{ub} = 1$. Points to the left of that curve represent pairs of β and γ for which a unique square-integrable solution is guaranteed to exist. On the other hand, the curve labeled "lower bound" defines the locus of points at which $\| \Psi_\theta \|_{lb} = 1$. Thus, points to the right of that curve represent pairs of β and γ for which a solution is guaranteed not to exist. Points between the curves lie in a region of uncertainty. Clearly, the choice of the function used to construct the lower bound to the norm is arbitrary, so that other lower bounds could be calculated. The function 1 was chosen purely for simplicity here.

From above it is clear that Assumption 10, which states that $\| \Psi_\theta \|$ is finite for each θ, is satisfied in the example if $|\rho| < 1/\sqrt{2}$. In order to satisfy Assumption 12, it must be less than 1. Experimenting with the parameter vector θ_1 used to construct Figure 1 shows that the admissibility of certain (β, γ) pairs depends on θ_1. Therefore, the parameter space Θ would probably have to be redefined as some compact subset of $\Theta_1 \times \Theta_2$.

In order to satisfy Assumption 11, a function that is square-integrable and is a uniform upper bound for ψ must be obtained. It is clear that for certain specifications of Θ, a function of the form $h(y,x) = a \exp(b|y| + c|xy|)$ will act as a uniform bound for ψ, and will be square integrable under appropriate conditions on a, b, and c. For example, suppose that $\beta \in [\underline{\beta}, \bar{\beta}]$, $\gamma \in [1, \bar{\gamma}]$, $\mu \in [0, \bar{\mu}]$, $\rho \in [\underline{\rho}, \bar{\rho}]$, and $\sigma \in [\underline{\sigma}, \bar{\sigma}]$ with $\bar{\rho} > 0$. Then, $\psi(y,x;\theta) \leq h(y,x) = \bar{\beta} \exp(|1 - \bar{\gamma}| \bar{\sigma} |y| + \bar{\rho} |xy|)$. The function h is square-integrable if $\bar{\rho} < 1/\sqrt{2}$. The conditions of Assumption 13 are satisfied for this choice of h as long as $\bar{\rho} < \frac{1}{2}$. The function $h(y,x)^2 f_0(y) f_0(x) \hat{J}[\Lambda(x), \Lambda(y)]$ is graphed in Figure 2 to illustrate that it converges to zero along any direction in R^2. The parameters used are $\bar{\beta} = 1.2$, $\bar{\gamma} = 20$, $\bar{\rho} = 0.25$, and $\bar{\sigma} = 0.1$. Therefore, it follows that for certain choices of Θ, the approximate solution functions converge in norm to the true solution functions in the example, and the convergence is uniform in the parameters.

Unfortunately, beyond the solution functions, which give the asset prices as functions of the forcing variables, there are other variables of interest in this model. In particular, the properties of the return on the risky asset are often studied. The current theory requires convergence in L^2 which is difficult to establish for arbitrary functions of v. This is generally not a problem when the state space is compact and all functions are continuous as in Burnside [4].

To illustrate estimation, consider the population mean of the price-dividend ratio of the risky asset given by

$$Ev = \int_{\mathbf{R}} \tilde{v}(\xi_t;\theta) \tilde{f}_s(\xi_t;\theta_1)\, d\xi_t,$$

where the price-dividend ratio has been written as a function of the original, untransformed, forcing variables, and $\tilde{f}_s(\xi_t;\theta_1)$ is the stationary distribution of ξ_t:

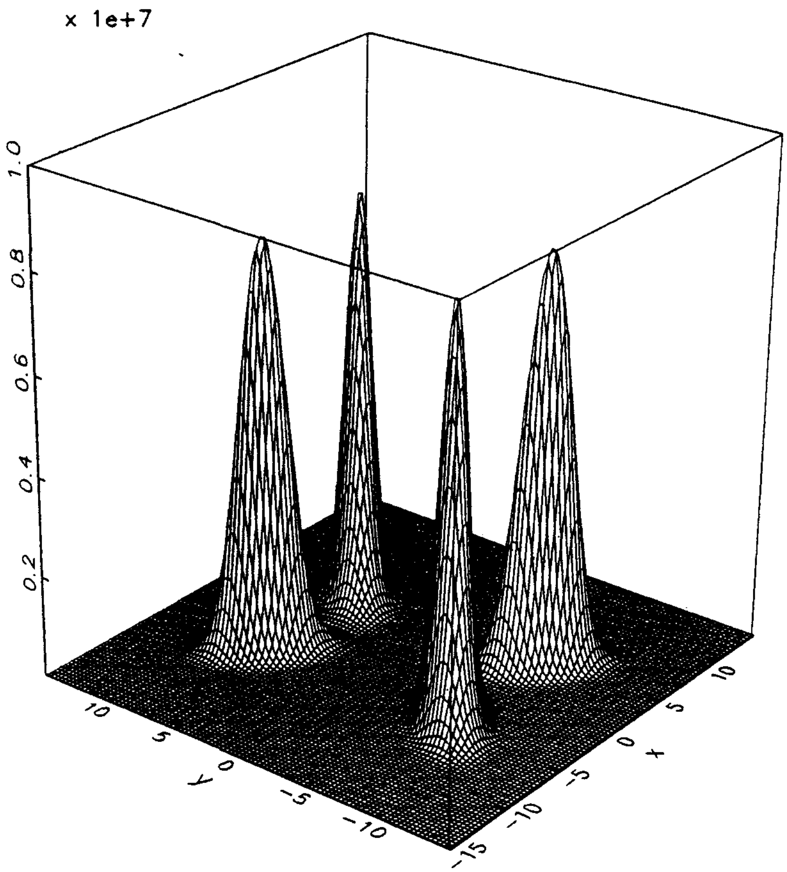

FIGURE 2. The function $h(y,x)^2 f_0(y) f_0(x) \hat{J}[\Lambda(x),\Lambda(y)]$.

$$\tilde{f}_s(\xi_t;\theta_1) = (2\pi)^{-1/2}\sqrt{1-\rho^2}\sigma^{-1}\exp\left[-\frac{(1-\rho^2)}{2\sigma^2}(\xi_t-\mu_\xi)^2\right].$$

The theory in Section 3 implies that Ev can be uniformly approximated by the mean of the approximate solution,

$$Ev_N = \int_{\mathbf{R}} v_N(x;\theta)(2\pi)^{-1/2}\sqrt{1-\rho^2}\exp\left[-\frac{(1-\rho^2)}{2}\right] x^2\,dx.$$

In the context of this section, $\Sigma_x = (1-\rho^2)^{-1}$. Therefore, the condition on Σ_x that was required in the previous section translates into requiring that $1-2\rho^2 > 0$, or $|\rho| < 1/\sqrt{2}$.

In order to have a concrete illustration of the estimator, suppose an exactly identified GMM estimator were used, which exploited the following moment conditions

$$
\begin{aligned}
E(\xi_t - \mu - \rho\xi_{t-1}) &= 0 \\
E[(\xi_t - \mu - \rho\xi_{t-1})\xi_{t-1}] &= 0 \\
E[(\xi_t - \mu - \rho\xi_{t-1})^2 - \sigma^2] &= 0 \\
E[v_t - Ev_N(\beta,\gamma,\mu,\rho,\sigma)] &= 0 \\
E[q_t - Eq(\beta,\gamma,\mu,\rho,\sigma)] &= 0,
\end{aligned}
$$

where v_t are data on price-dividend ratios, and q_t are data on the prices of riskless securities. In the asset pricing model, the price of a riskless asset that pays one unit of consumption one period in the future is given by

$$
\begin{aligned}
q_t &= \beta E_t \exp(-\gamma\xi_{t+1}) \\
&= \beta \exp[-\gamma(\mu + \rho\xi_t) + \tfrac{1}{2}\gamma^2\sigma^2]
\end{aligned}
$$

Therefore,

$$
Eq = \beta \exp\left[-\frac{\gamma\mu}{1-\rho} + \frac{1}{2}\gamma^2 \frac{\sigma^2}{1-\rho^2}\right].
$$

The only moment being approximated numerically is the mean of the price-dividend ratio. Parameter estimates obtained using the Mehra and Prescott [20] data set are presented in Table 1. The sample is annual data from 1889 to 1979. The risky asset is a single share of the S&P 500 index. The riskless asset price is identified with the inverse of the real return on relatively risk-

TABLE 1. Parameter estimates for the asset pricing example

	$\xi_t = \Delta \ln(C_t)$		$\xi_t = \Delta \ln(D_t)$	
Parameter	Estimate	Std. Error	Estimate	Std. Error
μ	0.0204	0.00496	0.0105	0.0130
ρ	−0.139	0.133	0.165	0.112
σ	0.0348	0.00339	0.120	0.0156
β	1.096	0.0530	0.951	0.0406
γ	11.45	3.890	2.618	3.156

These estimates were computed using an exactly identified GMM estimator, where the asymptotic variance covariance matrix of $\sqrt{T}m_T(\theta_0)$ was estimated using the method of Newey and West [21] using 5 lags of the GMM errors. The numerical solutions for the price-dividend ratios were computed using 4-point quadrature rules.

less securities. Two different sets of estimates are presented, where ξ_t is assumed alternately to be consumption growth or dividend growth for the S&P 500. The standard errors presume that these estimators are asymptotically normal. These estimates confirm the feasibility of the estimator. They complement the estimates obtained by Kocherlakota [14], who shows that an estimator that matches the means of the rates of return on these two assets leads to estimates of β greater than 1, and economically large values of γ. Similar parameter estimates are obtained here, where ξ_t is identified as consumption growth. However, the parameter estimates change dramatically when ξ_t is identified as dividend growth.

5. CONCLUSIONS

This paper provides theoretical backing for method of moments estimators which are constructed using a particular solution algorithm for a particular class of models. However, it is also clear that the consistency proof above applies to any method of moments estimator based on a numerical method which yields uniformly accurate approximations to the population moments implied by an economic model. This is subject to the law of motion governing the variables in the model satisfying Assumptions 1–4.

The advantages of the estimator proposed here are that, at least in the asset pricing example, the solution and moment functions can be economically computed, and in fact, are differentiable in the parameters. This may alleviate the burden of computing time involved in obtaining the moments through simulation, and in performing searches for the optimal parameters.

Although this approach is attractive in that it allows model solving and estimation to be unified, it does have its drawbacks. The estimator inherits its properties from the properties of the solution method which in turn may depend on the model itself. Thus, if a different solution method is required in solving the model, its convergence properties must be derived. All that is required for consistency is the uniform convergence of approximate moment functions.

NOTES

1. An earlier version of this paper which discusses compact state spaces is available from the author.

2. In an earlier version of this paper, ξ_t was assumed to be an AR(p) process. In order to reduce notation, p is assumed to be 1 here.

3. The derivative of the transformation $\Lambda(x)$ is bounded between 0 and $\frac{1}{4}$. As a result, each quadrature interval is narrower when mapped to $\mathbf{U}$ from $\mathbf{R}$. Since the widest quadrature interval in $\mathbf{R}^S$ goes to zero as $\hat{N} \to \infty$, this guarantees that the widest quadrature interval in $\mathbf{U}^S$ goes to zero.

4. In an earlier version of this paper, a bivariate forcing process was considered at the cost of additional notation.

REFERENCES

1. Amemiya, T. *Advanced Econometrics.* Cambridge, Mass.: Harvard University Press, 1985.
2. Atkinson, K.E. *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind.* Philadelphia: S.I.A.M., 1976.
3. Baker, C.T.H. *The Numerical Treatment of Integral Equations.* Oxford: Clarendon Press, 1977.
4. Burnside, C. *Solving, Estimating and Testing Nonlinear Asset Pricing Models: A Unified Approach.* Ph.D. dissertation, Northwestern University, 1991.
5. Burnside, C., M. Eichenbaum & S. Rebelo. Labor hoarding and the business cycle. (forthcoming) *Journal of Political Economy* (1993).
6. Cecchetti, S., P. Lam, & N. Mark. The equity premium and the risk free rate: matching the moments. National Bureau of Economic Research, Working Paper No. 3752 (1991).
7. Christiano, L. & M. Eichenbaum. Current real business cycle models and aggregate labor markets. *American Economic Review* 82 (1992): 430–450.
8. Duffie, D. & K.J. Singleton. Simulated moments estimation of markov models of asset prices. Manuscript (1989).
9. Gagnon, J.E. & J.B. Taylor. Solving and estimating equilibrium models with the extended path method. Manuscript (1986).
10. Gallant, A.R. *Nonlinear Statistical Models.* New York: Wiley, 1987.
11. Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* 50 (1982): 1029–1054.
12. Hodrick, R., N. Kocherlakota, & D. Lucas. The variability of velocity in cash-in advance models. *Journal of Political Economy* 99 (1991): 358–384.
13. Judd, K.L. Minimum weighted residual methods for solving aggregate growth models. Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, Discussion Paper 49 (1991).
14. Kocherlakota, N. On tests of representative consumer asset pricing models. *Journal of Monetary Economics* 26 (1990): 285–304.
15. Labadie, P. Solving nonlinear rational expectations models for estimation: an application to an asset pricing model with currency. Manuscript (1986).
16. Lee, B.-S. & B.F. Ingram. Simulation estimation of time-series models. *Journal of Econometrics* 47 (1991): 197–205.
17. Lucas, R.E. Asset prices in an exchange economy. *Econometrica* 46 (1978): 1429–1445.
18. Marcet, A. Solving nonlinear stochastic models by parameterizing expectations. Manuscript (1989).
19. Marshall, D.A. Solving nonlinear models by parameterizing expectations: convergence results. Northwestern University, Department of Finance, Working Paper No. 73 (1990).
20. Mehra, R. & E.C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (1985): 145–161.
21. Newey, W.K. & K.D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (1987): 703–708.
22. Szegö, G. *Orthogonal Polynomials.* Providence, R.I.: American Mathematical Society, 1975.
23. Tauchen, G. Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics* 4 (1986): 397–416.
24. Tauchen, G. & R. Hussey. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica* 59 (1991): 371–396.
25. Taylor, J.B. & H. Uhlig. Solving nonlinear stochastic growth models: A comparison of alternative solution methods. *Journal of Business and Economic Statistics* 8 (1990): 1–17.
26. Wouk, A. *A Course of Applied Functional Analysis.* New York: Wiley, 1979.