Economics Working Paper 76

Solving Nonlinear Rational Expectations Models by Parameterized Expectations: Convergence to Stationary Solutions*

Albert Marcet†
and
David A. Marshall‡

March 1994

Keywords: Numerical Algorithm, Monte–Carlo Integration, Rational Expectations, Stochastic Difference Equations.
Journal of Economic Literature classification: E17.

---

* We thank Larry Christiano, Wouter den Haan, Joan Ketterer, Tom Sargent, Ken Singleton, and three anonymous referees for helpful suggestions. The first part of this paper was presented at Stanford University in January 1989; thanks are due to its attendants, including Ken Judd, Rody Manuelli, John Bryant and Richard Rogerson. Marcet's research was partially supported by the NSF and the DGICYT (Spanish Ministry of Education). All errors are our own.

† Universitat Pompeu Fabra.
‡ Northwestern University.

Abstract

This paper develops the Parameterized Expectations Approach (PEA) for solving nonlinear dynamic stochastic models with rational expectations. The method can be applied to a variety of models, including models with strong nonlinearities, sub-optimal equilibria, and many continuous state variables. In this approach, the conditional expectations in the equilibrium conditions are approximated by flexible functional forms of finite elements. The approach is highly efficient computationally because it incorporates endogenous oversampling and Monte–Carlo integration, and it does not impose a discrete grid on the state variables or the stochastic shocks. We prove that PEA can approximate the correct solution with arbitrary accuracy on the ergodic set by increasing the size of the Monte–Carlo simulations and the dimensionality of the approximating family of functions.

1 Introduction

During the last decade, the use of dynamic stochastic models has extended to all fields of economics. These models are difficult to analyze because they typically have no analytic solution. However, recent increases in the power of computer hardware now allow these models to be studied by numerical simulation techniques. Simulations can be used to study the model from both a theoretical and empirical perspective. Theoretically, one can see if the model reproduces some stylized facts and how it responds to a change in the environment or in policy. Empirically, one can perform calibration exercises or do more formal testing using estimation by simulation or maximum likelihood procedures.

The numerical algorithms most widely used for solving dynamic models with rational expectations in economics are value-function iteration and linear-quadratic approximation. They are based on dynamic programming, so they are not well-suited to models in which the equilibrium does not correspond to the solution of a planner's problem. Furthermore, value-function iteration is affected by the 'curse of dimensionality', so is impractical for models with several continuous endogenous state variables. Linear-quadratic approximation can handle large models, but does not provide an arbitrarily good approximation in nonlinear models; this is especially a problem, for example, when the model has inequality constraints.

This paper introduces the parameterized expectations approach (PEA) for calculating numerical solutions to stochastic nonlinear models with rational expectations. In this approach, the equilibrium conditions of the model are written as a system of stochastic difference equations, and the conditional expectations in these equations are parameterized with flexible functional forms of finite dimensions, such as polynomials or splines. Simulations are then generated using these functions in place of the conditional expectations in the equilibrium conditions. A PEA solution corresponds to a parameterization of the conditional expectations that is consistent with the series it generates.

In this paper, we formally characterize the PEA, and we derive a set of conditions under which arbitrarily good accuracy can be obtained as the solution is refined (for example, as the degree of the polynomial goes to infinity); it contains some of our earlier work in Marcet [1988], Marshall [1988] and Marcet and Marshall [1992].

The PEA algorithm has been successful in delivering solutions to a number of large and complicated models. Applications in the literature include monetary asset pricing models, such as Marshall [1988, 1992], and den Haan [1990, 1991]; models of exchange rates, such as Bansal [1990], Bansal, Gallant, Hussey and Tauchen [1994] and McCurdy and Ricketts [1992]; asset pricing models with heterogeneous agents, such as Marcet and Singleton [1989], Ketterer and Marcet [1989], and den Haan [1993]; models of tax policy, such as Rojas [1991], and Otker [1992]; and nonstandard stochastic growth models, such as Marcet and Marimon's [1992] model with participation constraints and Christiano and Fisher's [1994] model with investment constraints. A related method in which the laws of motion of endogenous variables are parameterized has been used to solve asset-pricing models in papers by Heaton [1993] and Bekaert [1993]. This non-exhaustive list of PEA applications includes models with many continuous state variables, multiple heterogeneous agents, strong non-linearities, inequality constraints, incentive constraints, participation constraints, non-stationarities, discrete choice spaces, and suboptimal equilibria. In short, the PEA has been road-tested: its practical applicability to a wide range of economic models has been demonstrated.[1]

We show that the PEA delivers an arbitrarily close approximation to the true equilibrium as the solution is refined. The set of assumptions we impose is very general: essentially, our proof applies to most models with continuous laws of motion (including non-differentiable cases). To our knowledge, no such proof is yet available for other methods that compute approximate equilibria by solving systems of Euler equations. Our solution procedure is easily adapted to cases where one has to solve for the transition path to the stationary distribution starting from an arbitrary initial condition, but in the current paper we concentrate our discussion on solving for the stationary and ergodic distribution.

In section 2, we formally introduce the parameterized expectations approach and discuss how it can be applied to three well-known examples. In section 3, we relate PEA to two strands of economic literature. We first discuss PEA's relationship to models of least-squares learning; we then compare

[1]This is to be contrasted with proposals to use numerical methods taken from other sciences, without demonstrating their applicability to the problems of concern to economists. For example, Judd (1993) lists a number of numerical techniques used in engineering and physics that have not been tried in full scale economic applications.

our method to other solution algorithms used in economic applications. In section 4, we present the convergence result. Section 5 provides a brief discussion of certain practical considerations in applying the method. Section 6 contains some conclusions and suggestions for extensions.

2 The Parameterized Expectations Approach (PEA)

2.1 A General Framework

We assume that the economy is described by a vector of variables $z_t \in Z \subset R^n$, and a vector of exogenously given shocks $u_t \in U \subset R^s$. The equilibrium process $\{z_t, u_t\}$ is known to satisfy a stochastic difference equation

$$g(E_t[\phi(z_{t+1})], z_t, z_{t-1}, u_t) = 0 \tag{1}$$

for all t, where $g : R^m \times R^n \times R^n \times R^s \to R^q$ and $\phi : R^n \to R^m$. Once the parameters of the model have been fixed, g and ϕ are functions known to the economist. The vector z_t includes all endogenous variables, as well as those exogenous variables that appear inside the expectation. The process u_t is assumed to be Markov of order one. As usual, E_t is the conditional expectation given information up to period t.

The system (1) will typically include Euler equations, resource constraints, laws of motion of exogenous processes, market clearing conditions, incentive constraints, and so forth. Inequality constraints of the form

$$h(E_t[\phi(z_{t+1})], z_t, z_{t-1}, u_t) \geq 0$$

are incorporated into (1) by including equations

$$(h(E_t[\phi(z_{t+1})], z_t, z_{t-1}, u_t))^- = 0$$

as part of the system g, where $(x)^-$ is the function 'negative part of x'. These inequality constraints may include Kuhn-Tucker conditions, second order optimality conditions, or participation constraints in incentive problems.

Often, the sufficient conditions for an equilibrium include, in addition to equation (1), a set of transversality conditions. For most models, a process

$\{z_t\}$ that satisfies (1) also satisfies the transversality conditions if and only if the process is stationary. Stationarity can be imposed as an additional side condition of the model.

We assume that the past information that is relevant for predicting $\phi(z_{t+1})$ can be summarized in a finite-dimensional vector of state variables $x_t \in X \subset R^l$ satisfying

$$x_t = f(z_{t-1}, u_t), \tag{2}$$

where f is a known function.[2] This implies

$$E_t[\phi(z_{t+1})] = E[\phi(z_{t+1})|x_t]. \tag{3}$$

Furthermore, we assume that the model is recursive, so that the conditional expectation is given by a time-invariant function $\mathcal{E}$ such that

$$\mathcal{E}(x_t) \equiv E[\phi(z_{t+1})|x_t]. \tag{4}$$

By definition, the function $\mathcal{E}$ satisfies

$$\mathcal{E} = \arg \min_{\{h: R^l \to R^m\}} E \parallel \phi(z_{t+1}) - h(x_t) \parallel^2 \quad \text{for all } t, \tag{5}$$

since the conditional expectation is the best predictor in the mean square sense.

Throughout the paper we refer to a *solution* as a stochastic process $\{z_t, u_t\}$ satisfying equation (1), (2), and (3), given the exogenous Markov process for $\{u_t\}$. Since we will are dealing with systems g that are invertible in their second argument z_t, finding a solution is equivalent to finding a function $\mathcal{E}$ such that, if $\{z_t, u_t\}$ satisfies

$$g(\mathcal{E}(x_t), z_t, z_{t-1}, u_t) = 0, \tag{6}$$

then $\mathcal{E}$ satisfies (4) for all t. Alternatively, finding a solution is equivalent to finding a law of motion H such that the z_t generated by

$$z_t = H(z_{t-1}, u_t) \tag{7}$$

[2] We are assuming here that the researcher knows how to select a set of sufficient state variables x_t. Sometimes it is difficult to determine the sufficient state vector x_t beforehand; this is often the case, for example, in models with private information. In principle, it is possible to apply PEA and find a sufficient set of state variables *numerically*; we would then have to incorporate into the algorithm a search over functions of past variables that summarize past information. We do not pursue this avenue any further in this paper.

satisfies (1), (2) and (3).

We now show how three well-known models fit into the above framework. These examples are chosen for their simplicity; applications of PEA to more complicated models can be found in the papers cited in the Introduction.

- **Example 1.1 (Lucas [1978] asset pricing model)**

  A representative consumer chooses a stochastic consumption process $\{c_t\}$ of a perishable good in order to maximize $E_0 \sum_{t=0}^{\infty} \delta^t u(c_t)$. The single asset in the economy is traded in a competitive market at price p_t, and pays an exogenous dividend flow $\{d_t\}$. The consumer also receives exogenous labor income $\{w_t\}$. The Euler equation for the maximization problem of the consumer is

  $$u'(c_t)p_t = \delta \; E_t[\; u'(c_{t+1}) \; (p_{t+1} + d_{t+1})]. \tag{8}$$

  To map this model into the above framework we let $z_t = (p_t, c_t, d_t)$, and $u_t = (d_t, w_t)$; the system of equations corresponding to (1) is given by (8) and the market clearing condition $c_t = d_t + w_t$. The function $\phi(z)$ is given by $u'(c)\;(p + d)$. A time-invariant solution for the asset price can be found for which $x_t \equiv (d_t, w_t)$ is a sufficient set of state variables.

- **Example 1.2 (Simple Stochastic Growth Model)**

  Consider the simple growth model where an agent maximizes $E_0 \sum_{t=0}^{\infty} \delta^t u(c_t)$ subject to

  $$\begin{aligned} &c_t + i_t \leq k_{t-1}^{\alpha}\theta_t \\ &k_t = k_{t-1}\mu + i_t \;, \qquad k_{-1} \text{ given,} \end{aligned} \tag{9}$$

  where c_t denotes consumption, k_t is the capital stock, i_t is investment, and θ_t is an exogenous stochastic productivity shock, Markov of order one. The first order condition for optimality is

  $$u'(c_t) = \delta \; E_t[\; u'(c_{t+1}) \; (k_t^{\alpha-1}\alpha\theta_{t+1} + \mu)] \tag{10}$$

  To map this model into the above framework, set $z_t = (c_t, k_t, k_{t-1}, i_t, \theta_t)$, $u_t = \theta_t$, and $x_t = (k_{t-1}, \theta_t)$. The function g is given by the resource constraints (9) and the Euler equation (10). The function $\phi(z)$ is given by

$u'(c)\ (k^{\alpha-1}\alpha\theta+\mu)$. Standard results from dynamic programming guarantee that the solution is characterized by a time-invariant conditional expectation function.

- **Example 1.3 (Simple Growth Model with Lower Bounds on Investment)**

  This example shows how inequality constraints can easily be handled by PEA. Suppose we add a non-negativity constraint to the simple growth model of example 1.2:

$$i_t \geq 0 \tag{11}$$

  With this restriction, the first order condition (10) is replaced by the Kuhn-Tucker conditions

$$u'(c_t) - \lambda_t - \delta\ E_t[\ u'(c_{t+1})\ (k_t^{\alpha-1}\alpha\theta_{t+1}+\mu) - \mu\lambda_{t+1}] = 0 \tag{12}$$

$$i_t\lambda_t = 0 \tag{13}$$

$$\lambda_t \geq 0, i_t \geq 0, \tag{14}$$

  where λ_t denotes the Lagrange multiplier associated with constraint (11). Mapping this model into framework (1) is accomplished by setting $z_t = (c_t, k_t, k_{t-1}, i_t, \lambda_t, \theta_t)$, $u_t = \theta_t$. The system g is given by (9), the Kuhn-Tucker conditions (12) and (13), and the negative parts of the inequalities in (14). Note that, in this case, $\phi(z) \equiv u'(c)\ (k^{\alpha-1}\alpha\theta + \mu) - \mu\lambda$. Again, the model can be shown to be recursive with standard dynamic programming techniques.[3]

2.2 Definition and Calculation of an Approximate PEA Solution

System (1) and explicit formulas for g and ϕ are easy to find in many models. Unfortunately, finding the solution analytically in any generality is not possible. The difficulty is that $\mathcal{E}$ cannot be determined unless the process

[3]The application of PEA with inequality constraints was developed by Marcet and Singleton [1991]. For a more detailed discussion of how to impose the Kuhn-Tucker conditions the reader is referred to that paper.

for z_t is known, but that process can not be backed out from (6) unless the conditional expectation $\mathcal{E}$ is known. The PEA algorithm addresses this difficulty by replacing $\mathcal{E}$ with an approximating function ψ which is chosen to resemble $\mathcal{E}$ in a manner to be made precise.

Formally, let P denote a class of functions that is dense in $\left\{h : R^l \rightarrow R^m\right\}$. It is assumed that each element of P is characterized by a parameter vector β with a finite number of nonzero elements. For example, P can be taken as the set of polynomials, splines, neural networks or finite elements. Then, β would be, respectively, the coefficients of the polynomial, the parameters characterizing each partition for the spline, the elements of the neural net, or the parameters characterizing the finite elements. The element of P characterized by parameter vector β is denoted either by $\psi(\beta, \cdot)$ or by the simpler notation ψ_β. The *set of admissible parameters* with at most ν non-zero elements is denoted D_ν. That is, $D_\nu \subset \left\{\beta \in R^\infty : i^{th} \text{ element of } \beta \text{ is zero if } i > \nu\right\}$.

In PEA, the function $\mathcal{E}$ is approximated by choosing an element of P that satisfies a property close to (6) and (5). More precisely, for a given positive integer ν, $\beta \in D_\nu$, and $\{u_t\}_{t=0}^{\infty}$, define the process $\{z_t(\beta), u_t\}_{t=0}^{\infty}$ as the solution for all t of the system

$$g\left(\psi(\beta, x_t(\beta)), z_t(\beta), z_{t-1}(\beta), u_t\right) = 0 \tag{15}$$

$$x_t(\beta) = f(z_{t-1}(\beta), u_t)$$

Now, let

$$G_\nu(\beta) = \arg\min_{\xi \in D_\nu} E \parallel \phi(z_{t+1}(\beta)) - \psi(\xi, x_t(\beta)) \parallel^2 \tag{16}$$

We wish to choose $\beta \in D_\nu$ as close as possible to $G_\nu(\beta_\nu)$, so the approximate PEA solution is characterized by a fixed point, denoted β_ν, as follows:

$$\beta_\nu = G_\nu(\beta_\nu). \tag{17}$$

For the sake of simplicity we will assume in the remainder of this section that G_ν is a well-defined mapping, and that a fixed point of G_ν exists; these issues will be addressed formally in section 4. With the appropriate substitutions, one can see that ψ_{β_ν} plays the role of $\mathcal{E}$ in (6), so our approximate solution is the process $\{z_t(\beta_\nu), u_t\}_{t=0}^{\infty}$.

From this point on, we address the problem of approximating the solution at the stationary and ergodic set. This is an interesting case, since stationarity and ergodicity are the basis of time series econometrics or calibration exercises. PEA has been applied to solutions outside of the stationary distribution by replacing the long simulation in Step 2 below with many repeated short-run simulations.[4] Therefore, the claim in Judd [1992] (page 447) that PEA can not solve for equilibria outside the ergodic set and it can not be used for policy comparisons is incorrect.

We now give an operational algorithm for computing the approximate PEA solution for a fixed ν:

- Step 1: Write the system g in (1) so that it is invertible with respect to its second argument. Find a set of state variables x that satisfies (3). If a unique solution exists, g and x should be selected so that (1) and (3) are satisfied only by the unique solution. Replace the true conditional expectation by the finitely-parameterized function $\psi(\beta, \cdot)$ for $\beta \in D_\nu$ to obtain (15).

- Step 2: Fix z_0. For a given $\beta \in D_\nu$ and for T large enough, draw a sample of size T of the exogenous stochastic shock and recursively calculate$\{z_t(\beta), u_t\}_{t=0}^T$ using (15).

- Step 3: Find the sample version $G_{\nu,T}$ of G_ν. More precisely, $G_{\nu,T}$ is defined by:

$$G_{\nu,T}(\beta) = \arg\min_{\xi \in D_\nu} \frac{1}{T} \sum_{t=0}^T \| \phi(z_{t+1}(\beta)) - \psi(\xi, x_t(\beta)) \|^2 \qquad (18)$$

  This minimization is easy to perform by computing a non-linear least squares regression with the sample $\{z_t(\beta), u_t\}_{t=0}^T$, taking $\phi(z_{t+1}(\beta))$ as the dependent variable and $\psi(\cdot, x_t(\beta))$ as the explanatory function.

- Step 4: Find the fixed point

$$\beta_{\nu,T} = G_{\nu,T}(\beta_{\nu,T}) \qquad (19)$$

  We will discuss the numerical calculation of $\beta_{\nu,T}$ in the section 5.

[4] See Marshall [1988] and [1993], Marcet and Marimon [1992], Rojas [1993] and Christiano and Fisher [1994] for applications. Also, see Marcet and Marshall [1992] for a proof of convergence when short-run simulations are used.

We end this subsection by discussing how this algorithm would be applied to the three examples presented in section 2.1.

- **Example 1.1 (Lucas [1978] asset pricing model)**

  The only non-trivial endogenous variable is the stock price p_t. In this and the remaining examples, ϕ takes only positive values, so it is appropriate to choose P as a set of functions that can take only positive values. Step 2 is easily accomplished by solving for the price in each period from

  $$u'(c_t)\, p_t(\beta) = \delta\, \psi(\beta, d_t, w_t)$$

  which is the (15) version of the Euler equation. The fixed point can be found iteratively[5]. Notice that our choice of state variables guarantees that we are approximating the unique stationary (non-bubble) solution. Bubble solutions could be allowed by adding today's price to the list of state variables.

- **Example 1.2 (Simple Stochastic Growth Model)**

  Steps 1 and 2 follow similar considerations to the previous example. Since the solution is a time-invariant function of (k_{t-1}, θ_t) and it satisfies the Euler equation, we can be certain that the solution approximates the unique solution of the model. Notice that the (suboptimal and meaningless) solutions to the Euler equation that violate the transversality condition will never be approximated; in those solutions c_0 is fixed, and k_t is a time-invariant function of (k_{t-1}, θ_t) and c_t, so that the transversality condition is imposed by our choice of state variables and the fact that initial consumption is not fixed.

- **Example 1.3 (Simple Growth Model with Lower Bounds on Investment)**

  We show how to use the Kuhn-Tucker conditions to find $\{c_t(\beta), i_t(\beta), k_t(\beta)\}$ while imposing inequalities (14). Let us write the parameterized version of (12) and (14):

  $$u\prime(c_t(\beta)) - \lambda_t(\beta) = \delta\, \psi(\beta; k_{t-1}(\beta), \theta_t) \qquad (20)$$

[5] For this model, the mapping G_ν as well as the fixed point β_ν can be found analytically. See Marcet [1988]

$$\lambda_t(\beta) \geq 0 \quad \text{and} \quad i_t(\beta) \geq 0 \tag{21}$$

Notice that, for a fixed value of $\psi(\beta; k_{t-1}(\beta), \theta_t)$, the left side of (21) is increasing in $i_t(\beta)$.

We can then proceed as follows: for each t:

(a) compute $(c_t(\beta), i_t(\beta))$ from (20) under the conjecture that $\lambda_t(\beta) = 0$. If the corresponding investment turns out to be negative, then

(b) set $i_t(\beta) = 0$, find $c_t(\beta)$ from the feasibility constraint, and then compute $\lambda_t(\beta)$ from (20).

It is clear that if step (a) delivers a negative investment, consumption will be lower when we go to step (b), so $u\prime(c_t(\beta))$ will be higher (relative to (a)). This insures that (20) delivers a positive λ, so (21) is satisfied.

This strategy requires that the Kuhn-Tucker condition be written in a way that the function g is envertible. For example, the Kuhn-Tucker condition could be expressed as,

$$u'(c_t)c_t - \lambda_t c_t = \delta \; E_t \left[c_t \; u'(c_{t+1}) \; (k_t^{\alpha-1}\alpha\theta_{t+1} + \mu) - \mu c_t \lambda_{t+1} \right] \tag{22}$$

However, if steps (a) and (b) were applied to (22), there may be states in which the function $u'(c)c$ is decreasing in i. In that case, step (b) would deliver a negative value for the multiplier λ. Therefore, (22) would be an inappropriate choice for g if PEA is used as a solution algorithm. In terms of the step-by-step description of the algorithm, (22) violates step 1, since it implies a representation for g that may not be invertible with respect to its second element.

3 Relation of PEA to the Literature

3.1 Approximate PEA Solution and Least Squares Learning

In section 4, below, we show that the PEA solution can approximate the rational expectations equilibrium arbitrarily well by letting ν and T go to

infinity. In this subsection we give an alternative interpretation to the PEA solution for fixed ν as the equilibrium of an economy with boundedly rational agents.

Suppose it were infeasible for the agents to compute the true conditional equilibrium function $\mathcal{E}$. For example, agents may not know the correct functional form for $\mathcal{E}$, or they may face computation constraints. Instead, agents are constrained to use some function ψ_β to forecast of $\phi(z_{t+1})$. In principle, they can choose any parameter $\beta \in D_\nu$, but they are restricted to a fixed degree ν. The process $\{z_t(\beta)\}$ can be interpreted as the vector of endogenous variables that would be generated by such an economy. If these agents were to choose an arbitrary parameter vector $\beta \in D_\nu$, the economy would not be in equilibrium: the best forecaster of $\phi(z_{t+1})$ within the set D_ν would in fact be given by $\psi(G_\nu(\beta), x_t(\beta))$, rather than by $\psi(\beta, x_t(\beta))$, the function used by the agents ex ante. Presumably, agents would eventually recognize their systematic errors and would update the parameters in their forecast function. The equilibrium of this boundedly rational economy corresponds to β_ν, the fixed point of G_ν defined in equation (17): if the boundedly rational agents use β_ν they will eventually realize that this is their best alternative, given that they are restricted to staying in D_ν.

For this model of boundedly rational agents to be meaningful economically, it should be locally stable: if agents start at $t = 0$ using some beliefs β_0 near the fixed point β_ν, and update β as new information received at each period (so their best forecast at time t is given by $\psi(\beta_t, x_{t-1})$), their forecast coefficients should converge to β_ν. One way to model how agents update their forecasts is to assume a least squares learning model (LSL), in which β_t is generated by a non-linear least squares recursive algorithm (see Appendix 2). Adapting some results in Ljung [1975], under certain regularity conditions, it is possible to show the following results:

- For $\overline{\beta} \in D_\nu$, if $\overline{\beta} \neq \beta_\nu$, $Prob(\beta_t \to \overline{\beta}) = 0$. That is, LSL will almost surely not converge to a parameter vector that is not a fixed point of G_ν.
- $Prob(\beta_t \to \beta_\nu) = 1$ locally, if and only if the differential equation

$$\dot{\beta} = G_\nu(\beta) - \beta \tag{23}$$

is locally stable at β_ν. That is, local stability of equation (23) is equivalent to local stability of the equilibrium of boundedly rational agents under least squares learning.

The second result will be used in section 5.1 to suggest some algorithms for computing the approximate solution.

To prove the convergence result in Proposition 1, we must impose sufficient conditions to insure that if a law of motion H° is close to the true solution H in (7), then $\{z_t^\circ\}$ (the process generated by H°) is close to $\{z_t\}$ (This property is implied by assumptions 4* and 4, in section 4.) Not surprisingly, this property is one of the regularity conditions that must be imposed for least squares learning to converge. Intuitively, if two H's are close but generate z's very far apart, agents will be chasing a moving target; their observations on z will not tell them how to update their forecast function. A condition like this is likely to be needed for *any* learning scheme to converge, as long as the learning scheme is only based on observations of the realized process.

3.2 Relation to Other Algorithms.

PEA has developed from other solution approaches used previously in economics. It is closely related to the backsolving procedure of Sims [1985], Novales [1991] and Ingram [1990]. In this approach, conditional expectations are fixed by assuming a process for $\{\phi(z_{t+1})\}$. The remaining variables, including exogenous processes, are then solved from a system like (1) using this assumption. Backsolving can be understood as PEA without Steps 3 and 4, with particular assumptions on the whole process for $\phi(z_{t+1})$. Another predecessor to PEA can be found in Townsend [1983] in a linear model with private information; that paper uses $\nu = 1$ and replaces Steps 2 and 3 with an explicit calculation of G_1 based on spectral densities, in which the linearity of the model is exploited. The idea of calculating the mapping G_1 with long-run simulations was previously discussed in Marcet and Sargent [1989].

Other authors have used approximations of Euler equations with flexible functional forms more recently. Examples can be found in Coleman (1988), Judd (1992), Baxter (1991) and McGratten (1993); some of these methods are based on numerical methods used in other sciences such as engineering

or physics. We now relate these methods to PEA in terms of the framework and the notation laid out in section 2. Define $\mathcal{E}(x;\beta)$ as the analogue to $\mathcal{E}$ for the β process. More precisely,

$$\mathcal{E}(x;\beta) \equiv E[\ \phi(z_{t+1}(\beta))|\ x_t(\beta) = x\ ], \tag{24}$$

so, if $F(\cdot \mid u)$ denotes the distribution of u_{t+1} conditional on $u_t = u$, (24) implies

$$\mathcal{E}(f(z,u);\beta) = \int^{U} \phi\left[H\left(H(z,u;\beta),u';\beta\right)\right]\ dF(u'|u).$$

Assuming that $x \equiv z$, these solution procedures can be summarized as follows:

- Choose a class of functions P and a degree of approximation ν. Parameterize the law of motion by an element $\pi \in P$; this is equivalent to parameterizing expectations as

$$\psi(\beta,z) = g_1^{-1}(\pi(\beta,z,u),z,u) \tag{25}$$

  where g_1^{-1} represents the inverse of g with respect to its first argument, evaluated at $g = 0$.

- Fix a grid of p points $(x^1,x^2,...,x^p)$ in the state space R^l and a set of weights $w^1,w^2,...,w^p$.

- For a given β, calculate $\mathcal{E}(x^j;\beta)$ numerically for each j. As we can see from (24), this involves calculating p integrals over R^s (recall that s is the number of exogenous shocks in u_t).

- Find β_ν that solves

$$\beta_\nu = \arg\min_{\{\beta\in D_\nu\}} \ \sum_{j=1}^{p}\ w_j \parallel \mathcal{E}(x^j;\beta) - \psi(\beta,x^j) \parallel^2 . \tag{26}$$

The difference $\|\mathcal{E}(x^j;\beta) - \psi(\beta,x^j)\|$ is often called the 'residual' of the Euler equation at grid-point j, and a generic name for these methods is 'Minimum Weighted Residuals'. The methods differ in the way the approximating class P, the grid-points, and the weights are chosen, how the system

g is written, the way iterations are performed to find the above minimum, and in the method for computing the integrals involved in $\mathcal{E}(x;\beta)$.

It is clear that the steps involved in PEA are similar to the steps in MWR methods. In particular, (26) is analogous to (18). There are, however, several fundamental differences:

1. In MWR, the grid-points and weights are exogenous in the sense that they are independent from the approximating parameter β. Notice that, in PEA, the sum in (18) is evaluated at the generated series, so that $\{x_t(\beta)\}_{t=1}^{T}$ plays the role of the grid, and the weights are given by the empirical probability of each value of x in the simulated series; hence, the grid and weights are chosen by the algorithm, as a function of β. This is known as *endogenous oversampling.*

2. The conditional expectation involved in $\mathcal{E}$ is never calculated explicitly in PEA. Instead, the approximation in (18) tries to find ψ_ξ close to $\phi(z_{t+1})$, so that the calculation of the *residual* (a fundamental step in MWR methods), is entirely bypassed.

3. all integrals are calculated by Monte-Carlo instead of quadrature.

The endogenous oversampling feature implies that PEA only pays attention to those points that actually happen in the solution. This avoids the problem that researchers face when they must choose grid-points without knowing which points are likely in equilibrium. With endogenous oversampling, only the economically relevant region of the state space is explored, so no computer time is spent on states that never happen in equilibrium. Furthermore, the resulting approximation fits more closely at those states that happen more frequently. This is one reason why PEA has been successful in models with a large number of state variables. On the other hands, methods based on exogenous state-space grids are impractical for such models, since the number of points in the grid increases exponentially with the number of state variables.

In numerical analysis, the grid $(x^1, x^2, ..., x^p)$ is sometimes adapted in the course of the algorithm as a result of the iterations on β. This is a form of endogenous oversampling. Unfortunately, these procedures have not yet been used by economists applying MWR. Indeed, some economists have

claimed that endogenous oversampling is a bad feature of an algorithm[6] . In fact, endogenous oversampling is an important feature of many algorithms in numerical analysis, and it is likely to be an essential element in any algorithm for solving models with many state variables.

By eliminating the integral computations involved in the calculation of the residual, PEA further reduces computation time, since standard quadrature integration is often very costly even in two-dimensional problems. (Notice that, strictly speaking, this makes PEA a non-MWR method.) The use of Monte-Carlo integration means that the integrals can be calculated even in models with a large number of stochastic shocks.

PEA parameterizes the conditional expectations directly, while economists using MWR often parameterize the laws of motion π. Formally, the two alternatives are equivalent: we showed in section 2 how to find a law of motion consistent with ψ_β; equation (25) shows how to find ψ_β consistent with a proposed law of motion π. However, parameterizing conditional expectations often has some practical advantages: *i)* sometimes we know that x contains fewer variables than z, so that β contains fewer parameters; *ii)* when the degree of the polynomial is increased, one can see if some higher-order elements need to be introduced beforehand by testing the predictive power of those elements[7]; *iii)* if the shocks have a continuous distribution, the conditional expectation is obtained by integrating over H, so that $\mathcal{E}$ is likely to be smoother than H and easier to approximate with low degree polynomials; *iv)* inequality constraints are often easier to impose when the conditional expectation function is parameterized[8].

The method in Heaton [1993][9] fits into the framework of Section 2 except that the approximate solution satisfies

$$\frac{1}{T}\sum_{t=0}^{T}\left[\phi(z_{t+1}(\beta_{\nu,T}))-\psi(\beta_{\nu,T},x_t(\beta_{\nu,T}))\right]\frac{\partial\pi(\beta_{\nu,T},x_t(\beta_{\nu,T}))}{\partial\beta}=0$$

Simple algebra shows that this is exactly equivalent to PEA with the parameterization (25) if the objective function defining G_ν in (9) is modified

[6]See, for example, Judd [1992].

[7]See den Haan and Marcet (1994) for a full description of this idea.

[8]Christiano and Fisher (1994) make this point in their comparison of solution techniques in a model close to our Example 1.3.

[9]Also used in Bekaert [1993].

to $E \| [\phi(z_{t+1}(\beta)) - \psi(\xi, x_t(\beta))]\ h(z_t(\beta)) \|^2$, where $h \equiv \left|\partial g_1^{-1}/\partial \pi\right|^{-1/2}$. In other words, Heaton's method is a special case of our algorithm where Step 3 is modified to be a *weighted* non-linear least squares minimization, and where h is used as a weighting function. Heaton's method performs endogenous oversampling, it uses Monte-Carlo integration, and it does not calculate any residual. A small modification of our convergence proof accounts for this case, but is not included in this paper.[10]

4 Convergence to the Approximate Solution as $\nu \longrightarrow \infty$

Convergence will be proved in the strong sense that the approximate law of motion H_{β_ν}, defined in (28) below, converges uniformly to the true law of motion H on the support of the stationary distribution. Proposition 1 establishes the convergence result when ν and T are chosen sufficiently large. Proposition 2 provides a partial converse to proposition 1.

A formal proof of convergence is essential for any proposed approximation method. It is not sufficient to show that ψ_β can, in principle, approximate any fixed function arbitrarily well. In PEA, the approximating function ψ_β is required to approximate a conditional expectation function $\mathcal{E}(\cdot;\beta)$ which itself depends on β through the simulated process $\{z_t(\beta), x_t(\beta)\}$. In other words, $\mathcal{E}(\cdot;\beta)$ is a moving target.

Furthermore, one can easily construct examples of plausible approximation schemes which fail to converge to the function being approximated. For example, Judd (1994) points to a standard non-convergence example where, if a polynomial of degree ν was fitted exactly at ν equally-spaced grid-points, the approximation becomes arbitrarily bad as ν becomes larger. He suggests that PEA may suffer from a similar problem of nonconvergence; Proposition 1 formally disposes of Judd's criticism.[11]

[10]Details on the algebra and the convergence proof for the weighted case are in Marcet and Marshall [1992].

[11]In any case, this example does not apply to PEA for two reasons: *i)* in the example, the function is evaluated at ν points, while PEA evaluates the function at T points, where T is much larger than ν; *ii)* the ν points in the example cited by Judd are chosen independently of β, while in PEA this task is endogenous. If anything, this non-convergence example may be indicative of convergence problems for algorithms, including some suggested in Judd

To demonstrate convergence of the approximate solution, several conceptual issues must be addressed: *i)* the class P has to be dense in the appropriate space. *ii)* Compactness of the function space P is needed in order to guarantee that a sequence of approximators contains a convergent subsequence. *iii)* In general, $G_{\nu,T}$ need not have a fixed point, so the definition of an approximate solution must be generalized to insure that a solution exists for all ν and T, and that a fixed point exists approximately for large ν and T. *iv)* It must be shown that $G_{\nu,T}$ is a good approximation to G_ν, for large T. *v)* Simulations are generated endogenously in Step 2, so we must insure that $\{z_t(\beta)\}$ is well behaved. In particular, we must show that the series does not explode, and that the effect of the initial conditions chosen for z dies out sufficiently rapidly. Point *v)* and the choice of P are related problems since one must insure that P is dense in the set of laws of motion that generate well-behaved series.

The first set of assumptions are regularity conditions on the functions defining the equilibrium (1). Let g_i^{-1} be the implicit function that defines the i^{th} argument of g; for example,

$$g(\alpha, g_2^{-1}(\alpha, z, u), z, u) \equiv 0 \tag{27}$$

The process $\{z_t(\beta), x_t(\beta)\}$ defined by equation (15) is a first-order Markov process satisfying $z_0(\beta) = z_0$ and

$$z_t(\beta) = H(z_{t-1}(\beta), u_t; \beta) \equiv g_2^{-1}(\psi(\beta, x_t(\beta)), z_{t-1}(\beta), u_t) \tag{28}$$

For notational convenience we will let H_β denote the function $H(\cdot, \cdot; \beta)$. Note that H_β is analogous to H.[12]

Assumption 1 *(a) Functions ϕ, g, and f are uniformly[13] Lipschitz-continuous in all arguments, and are differentiable a.e.*

(1992), where $n = \nu$ and where the grid is exogenous. To our knowledge, no convergence proof for such methods is yet available.

[12]We use "H" to denote both the Markov operator defined in (7) and that defined in (28). It will be clear which usage is intended since the operator defined in (28) has β as an argument, while the true solution does not.

[13]A function with multiple arguments is said to be uniformly Lipschitz-continuous if the Lipschitz coefficient with respect to the i^{th} argument does not depend on the value of the j^{th} argument.

(b) For all $(\alpha, z, u) \in \phi(Z) \times Z \times U$, g_2^{-1} *satisfying (27) exists and is uniformly Lipschitz-continuous in its second argument.*

(c) The true conditional expectation $\mathcal{E}$ *is Lipschitz-continuous a.e.*

where 'a.e.' is with respect to the Lebesgue measure. Notice that parts (a) and (b) can be checked directly, while part (c) can not be verified directly from functions ϕ, g, and f. It should be kept in mind, however, that in most models parts (a) and (b) imply part (c).[14]

Assumption 1(b) insures that H can be derived from knowledge of g and $\mathcal{E}$; similarly, it insures that H_β is well defined and that it can be derived from knowledge of g and ψ_β. The assumption that g_2^{-1} is Lipschitz-continuous insures that a small change in z_{t-1} does not necessitate an arbitrarily large change in z_t to maintain equilibrium condition (1). If g is differentiable, this assumption requires the partial derivative of g with respect to its second argument to be uniformly bounded away from zero. Assumption 1 implies that H is Lipschitz-continuous.

The next two assumptions insure that the conditional expectation terms in (1) can be well approximated by some function ψ. Approximating classes of functions, such as polynomials. exponentiated polynomials, or splines, approximate a given function arbitrarily well only over a compact set. Therefore, we impose

Assumption 2 *$Z \times U$ is a compact set.*

Assumption 2 implies that both exogenous and endogenous state variables have compact support. Compact support for endogenous processes can be implied by economic models in a number of ways. In models with capital accumulation, depreciation often implies a bounded capital stock if U is bounded. Asset pricing models are only well defined if a lower bound on asset-holdings is imposed on agents; this is usually assumed through short-sale constraints. Finally, if the model is stationary, bounded support can be achieved by directly imposing exogenous constraints on the model's variables at levels which will be attained with very small probability. For example, in the simple growth model with zero depreciation the capital stock can be

[14]See, for example, Santos's [1991] proof that $\mathcal{E}$ is differentiable in dynamic programming models.

arbitrarily large. However, a very high level of the capital stock would be achieved in equilibrium with extremely low probability, so the technology of the model can be modified by imposing a very large upper bound on the level of the capital stock. Presumably, the difference between the behavior of the original model and this modified model will be negligible.

The next set of assumptions describes the class of approximating functions, $\{\psi : D_\nu \times \mathbf{R}^p \to \mathbf{R}^m\}$, where D_ν is a compact subset of the space of sequences with at most the first ν elements non-zero.[15] We will construct the sequence $\{D_\nu\}_{\nu=1}^{\infty}$ such that $D_\nu \subset D_{\nu+1}, \forall\ \nu$, and we define $D \equiv \bigcup_{\nu \geq 1} D_\nu$.

Assumption 3 *(a)* $\forall\ x \in X$, *for each* ν, *the restriction of* $\psi(\cdot, x)$ *to* D_ν *satisfies a Lipschitz condition uniformly in* x.

(b) $|\psi(\beta, \cdot)| \leq M,\ \forall \beta \in D$, *where* $\sup_{z \in Z} \phi(z) < M < \infty$.

(c) $\psi(\beta, \cdot)$ *is continuous, differentiable almost everywhere, and* $\exists\ K < \infty$ *such that* $\left| \frac{\partial \psi(\beta, x)}{\partial x} \right| < K,\ \forall \beta \in D,\ x \in X$ *where the derivative exists.*

(d) For any continuous function $q : D_\nu \times \mathbf{R}^p \to \mathbf{R}^m$ *such that* $| q | \leq K' < K$, *there exists a sequence* $\{\overline{\beta}_\nu\}_{\nu=1}^{\infty}$, $\overline{\beta}_\nu \in D_\nu$, *such that*

$$\lim_{\nu \to \infty} \sup_{x \in X} \left| \frac{\partial \psi(\overline{\beta}_\nu, x)}{\partial x} - q(x) \right| = 0.$$

Assumption 3(b) is nonrestrictive, since ψ is only used to approximate the conditional expectation of ϕ. Notice that in assumption 3(d) we assume that the *derivatives* of the approximating function sequence approximate any continuous function. In Lemma 1 of Appendix 1 we show that assumption 3(d) implies that any absolutely continuous function can be uniformly approximated by some sequence $\{\psi_{\overline{\beta}_\nu}\}_{\nu=1}^{\infty}$. In practice, assumption 3(d) is not very restrictive. For most commonly-used approximation functions, such as polynomials or splines, the derivatives are themselves a class of approximating function. (Step functions are one exception.)

[15]The restriction that D_ν be compact is without loss of generality. In the case that the parameters may have to be arbitrarily large for obtaining an approximation (as it may be the case, for example, with polynomials), the $\{D_\nu\}$ sequence is constructed so that there is a bound on all elements of $\{D_\nu\}$ but this bound goes to infinity as ν grows.

There is a popular misconception that the derivatives of polynomial approximators must diverge from the derivatives of the target function when the approximators become uniformly close to the target. This is true only for certain ways of constructing polynomial approximators; for example, if the approximation is required to fit a ν^{th} order polynomial to the target function exactly on ν points. It is not true, however, if the approximating sequence is chosen to minimize other criteria, such as the L^2 distance, that take into account the fit of the function at many points. For example, polynomials would fit Assumption 3, since the derivative of a polynomial is itself a polynomial, and our Lemma 1 shows that an approximating sequence can be chosen with bounded derivatives.

In this paper, we restrict our attention to stationary and ergodic processes.

Definition $\{z_t^\infty(\beta)\}$ and $\{z_t^\infty\}$ are stationary and ergodic processes satisfying equations (28) and (7), respectively, for $t = 0, \pm 1, \pm 2, ...$

That is, $\{z_t\}$ and $\{z_t(\beta)\}$ denote processes generated respectively by (25) and (12) starting from a fixed initial condition z_0, while $\{z_t^\infty\}_{t=-\infty}^{\infty}$ and $\{z_t^\infty(\beta)\}_{t=-\infty}^{\infty}$ denote stationary and ergodic processes. The processes $\{z_t(\beta)\}$ and $\{z_t\}$ are, in general, non-stationary, since the initial condition is fixed, rather than a draw from the stationary distribution.

There is a final set of issues to be addressed. In common with virtually all solution methods discussed in section 3, PEA delivers an approximation, H_β, to the equilibrium law of motion H. This approximate law of motion generates a time series $\{z_t(\beta)\}_{t=1}^{T}$, starting from some arbitrary initial condition, which is used to obtain inferences about the stochastic properties of the true stationary equilibrium process z_t^∞. For this procedure to be valid, three conditions must hold: *i)* $\{z_t^\infty(\beta)\}$ and $\{z_t^\infty\}$ must exist; *ii)* the effect of the initial condition must decline as t grows; *iii)* if H_β is close to H, then $z_t^\infty(\beta)$ must be close to z_t^∞. These three conditions are closely related; most processes either satisfy all three conditions or violate all of them. None of these conditions hold, for example, if $\{z_t\}$ were an explosive process or a random walk.

Insuring that these conditions hold is not only a problem for PEA. They must hold for *any* proposed approximation procedure if the approximator is

to converge to the true stationary equilibrium process. For example, if H were explosive, an approximation method that used a fixed grid $(x_1, ..., x_p)$ could never converge to the true equilibrium, since the endogenous process would eventually explode out of the pre-specified discrete grid.

These considerations require additional regularity conditions both on the true equilibrium and on the space of admissible PEA approximators. We propose two distinct sets of conditions, either of which is sufficient to prove proposition 1. In the first, more general, approach, we simply assume the needed conditions directly. Formally, for an arbitrary law of motion H^α (where α is an index) let $\{ z_t^\alpha \}$ denote the process generated by H^α starting at fixed initial condition z_0, and let $\{z_t^{\infty\alpha}\}$ denote the stationary process associated with H^α.

Definition S is a closed set (in the sup norm) of laws of motion such that, for all $H^\alpha \in S$

(a) a stationary process $\left\{z_{t-1}^{\infty\alpha}, u_t\right\}$ with support in the set $\overline{ZU}^\alpha$ exists

(b) if initial condition $(z_0, u_1) \in \overline{ZU}^\alpha$, $\sup_\alpha |z_t^\alpha - z_t^{\infty\alpha}| \to 0$ almost surely as $t \to \infty$, uniformly in the initial condition z_0.

(c) for any sequence of functions $\left\{H^k\right\}$ such that $H^k \to H^\alpha$ in the sup norm as $k \to \infty$, we have $z_t^{\infty k} \to z_t^{\infty\alpha}$ almost surely as $k \to \infty$, uniformly in t.

Note that condition (b) allows for a process with several disjoint ergodic sets. All that is required is that, once the process is in the ergodic set, the effect of the initial conditions disappears.

Assumption 4* *The law of motion H is in the interior of S.*

This assumption can not be verified directly from knowledge of system (1). There is a large literature on how to verify parts (a) and (b) of the definition of S analytically.[16] Part (c) in this definition is essentially a robustness condition; if it did not hold, one can make the case that the model at hand is not a particularly interesting one, since the time-series properties of the

[16]For example, Marimon (1989) shows existence of a stationary and ergodic solution of the growth model under very general assumptions. Santos [1991] verifies part b) for dynamic programming models.

model's endogenous variables would be highly sensitive to small deviations from fully rational behavior. Alternatively, one can verify assumption 4* by using proposition 2 below: if assumption 4* is the only assumption to fail, this will be detected with PEA by the absence of an asymptotic fixed point.

Under assumptions 1, 2, 3, and 4* , the approximate PEA solution converges to the true stationary equilibrium, in the sense of Proposition 1 below, if the set D_ν of admissible parameters is restricted to the subset $\mathbb{B}_\nu^S$, defined by

$$\mathbb{B}_\nu^S \equiv \{\beta \in D_\nu : H_\beta \in \mathcal{S}\}. \tag{29}$$

A potential problem with using assumption 4* is that it is difficult to check formally if $\beta \in \mathbb{B}_\nu^S$. In principle, $\beta \in \mathbb{B}_\nu^S$ can be checked informally, as follows: part (a) can be tested by observing if the solution settles around a stationary distribution; part (b) can be tested by re-doing the calculations with different z_0; finally, if part (c) were not satisfied, the series will not settle down even with small changes in β, and the user will notice that the iterations to find the fixed point will not converge.

Our second set of regularity conditions avoids this problem: the needed restrictions on the β_ν's can be checked formally. The cost is that this second set of regularity conditions is somewhat less general than assumption 4*. The alternative conditions uses the following version of Duffie and Singleton's (1993) asymptotic unit circle (AUC) condition:

Assumption 4 *(a) H has a stationary and ergodic solution $\left\{z_{t-1}^{\infty}, u_t\right\}$ with support in the set $\overline{ZU}$.*

(b) There exists a sequence of positive random variables $\{\rho(u_t)\}$ satisfying

$$E\log[\rho(u_t)] \equiv \alpha < 0 \text{ a. s.} \tag{30}$$

such that $H(\cdot, u)$ has Lipschitz coefficient $\rho(u)$ in $\overline{ZU}$.

Assumption 4 is a nonlinear analogue to the unit circle condition in linear time series models. While this is more restrictive than 4*, it is still satisfied in most models of interest. First, the Lipschitz constants on H need to be imposed only on the ergodic set of the true solution. Second, as long as the expectation in (30)is less than zero, the condition permits Lipschitz constants greater than unity for a subset of U with positive (possibly substantial)

probability measure. Finally, the assumption is silent about the particular transformation of the endogenous variables used to construct the z_t series. For example, the simple growth model of Example 1.2 for $\mu = 0$ is known to satisfy this condition if we write the law of motion in terms of $\log(k_t)$, even though this may not be true if H is written in terms of k_t directly.

Under assumption 4, we must impose an analogous AUC condition on the set of admissible approximators, which will be denoted $\mathbf{B}_\nu^{AUC}$:

Definition: For all ν, $\mathbf{B}_\nu^{AUC}$ is a closed subset of D_ν with the property that, $\forall \beta \in \mathbf{B}_\nu^{AUC}$, there exist psitive constants δ_β and positive random variables $\{\rho_\beta(u_t)\}$ satisfying

$$E \log[\rho_\beta(u_t)] \leq \alpha < 0 \text{ a.s.} \tag{31}$$

such that, for all $\| \beta' - \beta \| \leq \delta_\beta$, $H(\cdot, u_t; \beta')$ has Lipschitz coefficient $\rho_\beta(u_t)$.

The expectation in (31) is with respect to the stationary distribution of u_t. For any given β, condition (31) can be verified numerically: set $\rho_\beta(u_t)$ equal to the maximum derivative of $H(\cdot, u_t; \beta')$, and integrate numerically over the $u's$.[17]

Under either assumption 4* or assumption 4, we must restrict the approximate solution $\beta_{\nu,T}$ to a subset $\mathbf{B}_\nu \in D_\nu$, where $\mathbf{B}_\nu$ equals $\mathbf{B}_\nu^S$ or $\mathbf{B}_\nu^{AUC}$ depending on which assumption is used. Under this restriction it is not easy to guarantee that $G_{\nu,T}$ has a fixed point, since $G_{\nu,T}$ now maps $\mathbf{B}_\nu$ into a larger set $\mathbf{D}_\nu$. One way to guarantee the existence of a fixed point is to restrict the minimization that defines $G_{\nu,T}$ in (18) to the set $\mathbf{B}_\nu$. The function $G_{\nu,T}$ would then map $\mathbf{B}_\nu$ into itself, and existence of a fixed point would follow from Brouwer's theorem. It turns out that this strategy does not deliver a proof of Proposition 1: we must use the fact that $G_{\nu,T}$ minimizes the

[17]Another consequence of using assumption 4 and $\mathbf{B}_\nu^{AUC}$ is that the proof becomes much more involved than if $\mathbf{B}_\nu^S$ and assumption 4* were used. The reason is that the proof needs to show that H can be arbitrarily well approximated by a sequence of $H_{\overline{\beta}}$'s with $\overline{\beta} \in \mathbf{B}_\nu^{AUC}$. In other words, we need that $P \cap \{\psi_\beta : \beta \in \cup_{\nu=1}^\infty \mathbf{B}_\nu^{AUC}\}$ is dense in the subset of $\{h : R^l \to R^m\}$ containing $\mathcal{E}$. Now, since $\cup_{\nu=1}^\infty \mathbf{B}^{AUC}$ has an empty interior, this does not follow immediately from the fact that P is dense in $\{h : R^l \to R^m\}$. On the other hand, denseness of $P \cap \{\psi_\beta : \beta \in \cup_{\nu=1}^\infty \mathbf{B}_\nu^S\}$ in the subset of $\{h : R^l \to R^m\}$ containing $\mathcal{E}$ follows immediately from assumption 4* and the definition of this set.

mean square error among *all* admissible $\beta \in D_\nu$. (See, for example, Lemma 8, equation 62). Instead, we use the following definition of an approximate solution, which generalizes equation (19):

Definition: An approximate solution of order ν, sample size T, is a parameter vector $\beta_{\nu,T}$ satisfying

$$\beta_{\nu,T} = \arg\min_{\beta \in \mathbf{B}_\nu} \frac{1}{T}\sum_{t=1}^{T} |\psi(\beta, x_t(\beta)) - \psi(G_{\nu,T}(\beta), x_t(\beta))|^2 \qquad (32)$$

Under assumptions 1 to 3 an approximate solution always exists by continuity and boundedness of the objective function. If $G_{\nu,T}$ had a fixed point in the set $\mathbb{B}_\nu$, the two definitions would obviously coincide. In practice, a fixed point usually can be found. Even if no fixed point exists for finite ν and T, lemma 9 in the appendix implies that as $\nu, T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} |\psi(\beta_{\nu,T}, x_t(\beta_{\nu,T})) - \psi(G_{\nu,T}(\beta_{\nu,T}), x_t(\beta_{\nu,T}))|^2 \ \to 0.$$

This makes precise the sense in which β and $G_{\nu,T}(\beta)$ can be made arbitrarily close. In that sense, the approximate solution (32) represents a fixed point asymptotically.

The minimization problem in the above definition is over a restricted set. Finding constrained minima numerically in non-linear setups is often difficult. In practical applications of PEA, one can avoid imposing this restriction directly by solving an unconstrained minimization along a homotopy path for which the constraint does not bind. More details are given at the end of next section.

Notice that, either under assumption 4 or under 4*, we may have non-uniqueness. This can happen because system (1) may be satisfied for several laws of motion H, or because a given law of motion has several ergodic sets $\overline{ZU}$. Our discussion of examples 1.1, 1.2 and 1.3 shows how to pick out unique solutions in some cases.

Now we present the fundamental proposition in this section.[18]

[18]It is understood that the double limits in the paper are defined as:

$$\lim_{\nu,T\to\infty} \zeta_{\nu,T} = \lim_{\nu\to\infty}\left(\lim_{T\to\infty} \zeta_{\nu,T}\right)$$

Proposition 1 (Convergence of Approximate Solutions): *Assume that there is a unique solution with a unique stationary and ergodic distribution; denote the support as* $\overline{ZU}$. *Under assumptions 1, 2, and 3, if , in addition,* either *assumption 4* holds and* $\mathbf{B}_\nu = \mathbf{B}_\nu^S$ or *assumption 4 holds and* $\mathbf{B}_\nu = \mathbf{B}_\nu^{AUC}$, *then*

$$\lim_{\nu,T\to\infty}\left(\sup_{(z,u)\in\overline{ZU}} \mid H(z,u;\beta_{\nu,T}) - H(z,u) \mid \right) = 0, \text{ a.s.}$$

$$\lim_{\nu,T\to\infty}\left(\sup_{x\in f(\overline{ZU})} \mid \psi(\beta_{\nu,T},x) - \mathcal{E}(x) \mid\right) = 0, \text{ a.s.}$$

(All proofs are in Appendix 1.) Notice that convergence obtains in the strong sense of uniform convergence. Also notice that convergence is guaranteed only in the ergodic set $\overline{ZU}$. In case that the solution is non-unique or has several ergodic sets, a trivial modification of the proof of Proposition 1 would show that the approximate solution becomes arbitrarily close to the set of ergodic solutions

The following corollary asserts that all properties of the model that are of interest in time series applications are appropriately approximated by PEA.

Corollary 1 *Under the conditions of Proposition 1, if* $(z_0,u_1) \in \overline{ZU}$, *we have*

(a) (Simulated Solution Paths converge)

$$z_t(\beta_{\nu,T}) \to z_t \qquad a.s., \text{ uniformly in } t,$$

(b) (Equilibrium Conditions are satisfied in the limit)

$$g\left(E\left[\phi(z_{t+1}(\beta_{\nu,T}))|x_t(\beta_{\nu,T})\right], z_t(\beta_{\nu,T}), z_{t-1}(\beta_{\nu,T}), u_t\right) \to 0 \qquad a.s.$$

(c) (Simulated Sample Moments converge): If $d : Z \to \mathbf{R}^q$ *is any Lipschitz function, then*

$$\frac{1}{T}\sum_{t=1}^{T} d(z_t(\beta_{\nu,T})) \to E(d(z_t^\infty)) \qquad a.s.$$

as $\nu, T \to \infty$.

We now provide a partial converse to Proposition 1. It shows how the conclusion of Proposition 1 depends on assumption 4* or 4, and provides a way to check these assumptions in the limit. Specifically, if $\mathbb{B}_\nu = \mathbb{B}_\nu^S$, part *(a)* says that if H does not satisfy assumption 4*, this can be detected because $\min \left\| \psi_\beta - \psi_{G_\nu(\beta)} \right\|^2$ is eventually bounded away from zero; part *(b)* tells us that those equilibria that fail to satisfy assumption 4* will not be approximated by PEA. (An analogous interpretation holds for $\mathbb{B}_\nu = \mathbb{B}_\nu^{AUC}$.) For example, this implies that rational expectations bubble equilibria in Example 1.1 will not be approximated by PEA with the choice of state variables proposed in section 2.

Proposition 2 *Under assumptions 1, 2 and 3,*

(a) if $\mathbb{B}_\nu = \mathbb{B}_\nu^S$ *and assumption 4* is not satisfied* or *if* $\mathbb{B}_\nu = \mathbb{B}_\nu^{AUC}$ *and assumption 4 is not satisfied, then there exist* $\gamma > 0$ *and* $N > 0$ *such that* $\forall \nu > N$

$$\min_{\beta \in B_\nu} E[\psi(\beta, x_t^\infty(\beta)) - \psi(G_\nu(\beta), x_t^\infty(\beta))]^2 > \gamma$$

(b) if $\mathbb{B}_\nu = \mathbb{B}_\nu^S$ *and there exists an equilibrium* $\tilde{H}$ *that does not satisfy Assumption 4*,* or *if* $\mathbb{B}_\nu = \mathbb{B}_\nu^{AUC}$ *and there exists an equilibrium* $\tilde{H}$ *that does not satisfy Assumption 4, then* $H_{\beta_{\nu,T}}$ *does not converge to* $\tilde{H}$.

Proposition 2 can be used to verify assumption 4* or 4 by calculating

$$\frac{1}{T} \sum_{t=1}^{T} \| \psi(\beta, x_t(\beta)) - \psi(G_\nu(\beta), x_t(\beta)) \|^2$$

for ν and T arbitrarily large. If $\mathbb{B}_\nu = \mathbb{B}_\nu^S$, assumption 4* is satisfied if and only if this infinite sum can be made arbitrarily close to zero. (An analogous verification of assumption 4 holds for $\mathbb{B}_\nu = \mathbb{B}_\nu^{AUC}$.) This is important because verifying assumption 4* or 4 analytically may be difficult if one simply inspects the equilibrium conditions (1).

5 Some Practical Issues.

5.1 Simple Algorithms for finding $\beta_{\nu,T}$

In order to calculate the fixed point of Step 4 (or the arg min in equation (32)) one can use standard hill-climbing algorithms for solving non-linear systems of equations. Nevertheless, this may not always be the best alternative. First, calculation of the gradient of $G_{\nu,T}$ can become very expensive in models with many coefficients. Second, $G_{\nu,T}(\beta)$ is only well defined if $\beta \in B_\nu$.

We will discuss the second issue in the subsection on homotopy, below. In order to avoid the first problem, the following algorithm based on modified successive approximations has been used successfully in many applications

$$\beta_{\nu,T}(\tau+1) = (1-\lambda)\beta_{\nu,T}(\tau) + \lambda G_{\nu,T}(\beta_{\nu,T}(\tau)) \tag{33}$$

for some $\lambda > 0$. It is clear that no gradient has to be calculated, so that the iterations are extremely easy to program, and each iteration is done very quickly. This algorithm needs more iterations to converge than a gradient algorithm, but there is a trade-off between ease of performing each iteration and number of iterations needed to converge. In practice, the above algorithm is often at least as fast as gradient algorithms.

Although one can construct examples where iterations on (33) are locally unstable, this has not been the case in most practical applications up to date. Furthermore, since the least-squares learning model is locally stable if and only if the differential equation (23) is stable, iterations on (33) with λ small diverge only if the model is locally unstable under learning. If this were the case, the model would be uninteresting from an economic standpoint.

Another fast algorithm that avoids calculating the gradient is to simulate directly the model under least squares learning. More precisely, Steps 2, 3 and 4 are substituted by

- Step 2': For T large enough, draw a sample of size T of the exogenous stochastic shock and calculate $\{\tilde{z}_t, u_t\}_{t=0}^{T}$ using the least squares learning model of Appendix 2.

Again, if the model is stable under learning β_t will converge to the fixed point.

5.2 Speed of Computation, Initial Conditions and Homotopy.

The design of Steps 1 to 4 in section 2 is describes the simplest possible application of PEA. A number of elementary modifications can be used to speed up computations, for example: one iteration on the algorithm for running the non-linear regressions of Step 3 is sufficient to deliver the fixed point; g can be often rewritten in a way that solving for z_t in Step 2 is very simple; the class of functions P can be chosen to match some properties of the conditional expectation (for example, it can be set to take only positive values); and, as suggested by the definition in equation (32), it is better to place the convergence criterion used to decide that the algorithm has arrived at the fixed point on the values of ψ_β, instead of on the values of β.

Suppose a researcher has calculated the solution of order ν and wants to calculate the solution of order $\nu+1$. In this case, using $\beta_{\nu,T}$ as initial condition for the iterations to find the $\nu+1$ approximation is not a good alternative, since the elements of the higher degree will usually be correlated with those of a lower degree; instead, it is best to start the algorithm at $G_{\nu+1,T}(\beta_{\nu,T})$. Furthermore, it is not always necessary to introduce all elements of degree $\nu+1$. Only those higher-order elements that have some incremental predictive power for $\phi(z_{t+1}(\beta))$ need be included.

It is likely that the introduction of textbook techniques from numerical analysis will be useful for finding minima, computing efficient Monte-Carlo integrals, introducing alternative flexible functional forms, and setting homotopy paths. One has to be careful not to introduce these techniques unless there is a good reason; oftentimes, the simplest approach will be sufficient for solving the model of interest. In many cases, a simple algorithm is also the fastest alternative.

Many successful pplications of PEA make use of homotopy techniques; this is a simple way of obtaining starting values for the algorithm and of keeping the simulations in the stable set B_ν. Along a homotopy, the desired solution is obtained by moving slowly from a known solution to the solution of the model we are interested in. For example, den Haan and Marcet (1990) calculate the simple growth model of example 1.2 with partial depreciation by starting out at the solution for the case $\mu = 0$ (a case for which we know the analytic solution), and then solving a series of models increasing μ gradually. At each step along the homotopy, the solution for the previous

step is used as initial condition. In this way, one insures that all iterations stay close to the true law of motion and, therefore, away from the boundary of $\mathbf{B}_\nu$. Therefore, there is no need to use *constrained* minimization routines in order to solve the minimization problem in (32).

PEA has turned out to be a fast and simple method in many applications. A comparison of speed of convergence, accuracy and convenience in a highly non-linear model has been made by Christiano and Fisher (1994); they use a simple growth model with irreversible investment (our example 1.3) and assume that the shock θ_t is a discrete Markov chain. Their model is particularly unsuitable to PEA, since it only has one continuous state variable and the stochastic shock can take only two possible values: in this particular model, endogenous oversampling is not very useful, and Monte-Carlo integration is very inefficient, since integrals $\mathcal{E}(x^j;\beta)$ are given by a simple formula that is exploited in the other algorithms but not in PEA. The computation times are higher for PEA but, even for their model, only by factors of five or seven. Furthermore, of the several algorithms they tried, PEA was the only one where the computation time increased only slightly when a non-negativity constraint was introduced. Finally, Christiano and Fisher find that PEA delivers a very accurate answer for low-degree polynomials and it is, by far, the easiest method to implement.

The claim in Judd (1992) that MWR methods are 'hundreds of times faster' than PEA is based on comparing the computation times of den Haan and Marcet (1990) with his own computations. These computation times are not comparable because different starting values were used: den Haan and Marcet (1990) (who were unaware that they were in a race) deliberately started the algorithm at a very incorrect initial condition to illustrate the use of homotopy. The solutions reported in Judd (1992) use the non-stochastic *steady-state solution* as starting value, which happens to be very close to the stochastic solution. Despite their model choice, the comparison done in Christiano and Fisher (1994) is more informative, since they start out all the algorithms at equivalent initial conditions.

6 Conclusion

We have presented the PEA algorithm for solving dynamic stochastic non-linear models with rational expectations. This approach is highly flexible,

and it has been used successfully in many applications with suboptimal equilibria, strong non-linearities and inequality constraints. It is particularly fast in those models with a large number of state variables and stochastic shocks, and is quite easy to implement.

We prove that, for models with continuous laws of motion and under some mild regularity assumptions, the approximate solution converges to the true solution with arbitrary accuracy as the approximation is refined. Convergence has been proved for the case of the stationary distribution. Similar results for other techniques solving Euler equations with finite dimensional approximations are not yet available.

A very general condition requiring, essentially, ergodicity of the true process, is enough for convergence. We have also proved convergence under the more restrictive assumption that Duffie and Singleton's AUC condition has to hold in the support of the ergodic distribution in order to guarantee that the stability condition on the approximate process can be verified formally. Future research might deliver similar convergence results for PEA under discontinuous laws of motion, but certain technical details, such as a condition for compactness of the class of approximating functions and an equivalent stability condition, need to be addressed. Our restriction to the support of the stationary distribution is non-essential: a result that guarantees convergence of the law of motion outside the ergodic set can be found in the predecessor to this paper when the algorithm uses short-run simulations.

The convergence theorem in this paper represents a first step in studying the properties of PEA and other numerical methods for solving systems of Euler equation. Additional work needs to be done. For example, it would be useful to understand the optimal rate at which ν and T should be allowed to grow. A more formal method of selecting the higher degree elements to include when increasing ν would be valuable. Finally, the use of more sophisticated homotopy techniques can improve the speed of convergence in large models, and it can provide a basis for a theorem proving that the iterations to find the approximate solution converge globally from the initial conditions. We believe that this area of research will pay substantial dividends in expanding the range of economic models accessible to quantitative analysis.

References

- Baxter, M. (1991); "Approximation of Suboptimal Dynamic Equilibrium: An Euler Equation Approach", *Journal of Monetary Economics*, October, pp 173-200.

- Bansal, R. (1990); "Can Nonseparabilities Explain Exchange Rate Movements and Risk Premia?" Ph.D. thesis, Carnegie-Mellon University.

- Bansal, R., R. Gallant, R. Hussey and G. Tauchen (1994); "Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data", forthcoming in *Journal of Econometrics*.

- Bekaert, G. (1993); "The Time-Variation of Risk and Return in Foreign Exchange Markets: A General Equilibrium Approach," Working Paper No. 1276, Stanford University, Graduate School of Business.

- Christiano, L. and J. Fisher (1994); "Prototype Algorithms for Solving Dynamic Models with Occasionally Binding Constraints", working paper, Northwestern University and University of Western Ontario.

- Coleman, W.J. (1988) "Equilibrium in a Production Economy with an Income Tax", working paper, Board of Governors of the Federal Reserve System.

- den Haan, W. (1990); "The Optimal Inflation Path in a Sidrausky Model with Uncertainty," *Journal of Monetary Economics* vol 25, 389-410.

- den Haan, W. (1991); "The Term Structure of Interest Rates in Real and Monetary Production Economies" working paper, Carnegie Mellon University.

- den Haan, W. (1993); "Aggregate Uncertainty, Heterogeneous Agents, and Short-term Interest Rates", working paper, UC San Diego.

- den Haan, W. and Marcet A. (1994); "Accuracy in Simulations", *Review of Economic Studies*, January.

- den Haan and Marcet (1990); "Solving a Simple Growth model by Parameterizing Expectations" *Journal of Business and Economic Statistics*, January, pp 31-34.
- Duffie, D., and K.J. Singleton (1993); "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica,* vol. 61, 929-952.
- Hansen, L. P. (1982); "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, vol. 50, 1029-1054.
- Heaton, J. (1993); "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," Working Paper, Alfred P. Sloan School of Management, Massachusetts Institute of Technology.
- Ingram, Beth (1990); "Equilibrium Modelling of Asset Prices: Rationality Versus Rules of Thumb", *Journal of Business and Economic Statistics.*
- Judd, K. (1992); "Projection Methods in Optimal Growth Models", *Journal of Economic Theory*, December, pp 410-453. Appeared as a working paper (1989) "Minimum Weighted Residual Methods in Dynamic Economic Models", Hoover Institution.
- Judd, K. (1993); 'Comments on Marcet, Pakes and Rust', discussion, forthcoming in *Advances in Econometrics*, ed. C.A. Sims, Cambridge University Press.
- Judd, K. (1994); 'Numerical Analysis in Economics', mimeo.
- Ketterer and Marcet (1989); "Introducing Derivative Securities: A General Equilibrium Approach". Preliminary Draft, Carnegie Mellon U.
- Kolmogorov and Fomin (1970); *Introductory Real Analysis*, Dover Publications, Inc.
- Ljung, L. (1975); "Theorems for the Asymptotic Analysis of of Recursive Stochastic Algorithms", Report 7522, Department of Automatic Control, Lund Institute of Technology, Sweden.

- Lucas, R. E. (1978); "Asset Prices in an Exchange Economy", *Econometrica*, vol. 46, 1426-1446.
- Marcet, A. (1988); "Solving Non-linear Stochastic Models by Parameterizing Expectations", working paper, Carnegie Mellon University.
- Marcet, A. (1990) "Simulation Analysis of Stochastic Dynamic Models: Applications to Theory and Estimation", Invited Symposium of the VI World Congress of the Econometric Society. Forthcoming in Advances in Econometrics, ed. by C.A. Sims.
- Marcet, A. and Marimon, R. (1992); "Communication, Commitment and Growth", *Journal of Economic Theory*, December, pp 219-250.
- Marcet, A. and D.A. Marshall (1992); "Convergence of Approximate Model Solutions to Rational Expectations Equilibria Using the Method of Parameterized Expectations," Working Paper No. 73, Department of Finance, Kellogg Graduate School of Management, Northwestern University.
- Marcet, A. and T. J. Sargent (1989); "Convergence of Least Squares Learning in Environments with Hidden State Variables and Private Information", *Journal of Political Economy*, vol 97, 1306-1322.
- Marcet, A. and K.J. Singleton (1990) "Equilibrium Asset Prices and Savings of Heterogeneous Agents with Portfolio Constraints" Working Paper, Carnegie Mellon University.
- Marimon, R. (1989); "Stochastic Turnpike Property and Stationary Equilibrium", *Journal of Economic Theory*, pp. 287-306.
- Marshall, D. (1988); "Inflation and Asset Returns in a Monetary Economy with Transaction Costs". Ph.D. thesis, Carnegie Mellon University.
- Marshall, D. (1992); "Inflation and Asset Returns in a Monetary Economy". *Journal of Finance*, vol. 47, 1315-1342.
- McCurdy T. and Ricketts, N. (1992); "An International Economy with Country-Specific Money and Productivity Growth Processes", working paper, Queen's University.

- McGratten, E. (1993); "Solving the Stochastic Growth Model by a Finite Element Method", Staff Report, Federal Reserve Bank of Minneapolis.
- Novales, A. (1990):"A Stochastic Monetary Equilibrium Model of the Interest Rate" *Econometrica*, vol.58, 93-112.
- Otker (1991); "The Welfare Implications of Fiscal Policy with Distortionary Taxation; a Numerical Analysis", working paper, Carnegie Mellon University.
- Rojas, G. (1991); "Optimal Taxation in a Model with Public Capital", working paper, Universitat Pompeu Fabra.
- Santos, M. (1991); "Smoothness of the Policy Function in Discrete Time Economic Models", *Econometrica*, vol. 59, 1365-1382.
- Sims, C. A. (1985); "Solving Non-Linear Stochastic Equilibrium Models 'Backwards"', manuscript, University of Minnesota.
- Townsend, R. (1983); "Forecasting the Forecasts of Others", *Journal of Political Economy*, vol 91, 546-565.
- Wheeden and Zygmund (1977); *Measure and Integral*, Marcel Dekker Inc.

APPENDIX 1
PROOFS AND LEMMAS

For ease of exposition and notation, the propositions and lemmas are proved for the case where G_ν and $G_{\nu,T}$ are well-defined functions (that is, the least-squares minimizers in (16), (18), (32) and , (54) are unique). This avoids dealing explicitly with convergence of correspondences. All propositions are valid for the general case where these objects may be non-unique. First, suppose $G_{\nu,T}(\beta)$ were multi-valued. For any $\xi \in G_{\nu,T}(\beta)$, $\psi(\xi, x)$ takes the same value, so the minimand in (32) is well-defined. If the minimizer in (32) were non-unique, there would be multiple equivalent PEA approximators for each value of ν and T. In that case, one could still construct a doubly-indexed sequence $\{\beta_{\nu,T}\}$ by selecting any one of the equivalent PEA approximators for each ν and T. Any such sequence converges to the true equilibrium, in the sense of proposition 1. Finally, throughout this appendix, $||\cdot||$ denotes the L^2 norm, and $\mathrm{B} \equiv \cup_{\nu \geq 1} \mathrm{B}_\nu$.

The first lemma proves that P is dense in the space of conditional expectations. If $\mathcal{E}$ was differentiable, the lemma would follow almost trivially from the fundamental theorem of integral calculus. Most of the derivations in the Lemma are to handle nondifferentiabilities in $\mathcal{E}$.

Lemma 1 *There exists a sequence $\overline{\beta}_\nu \in \mathrm{D}_\nu$ such that*

(a)

$$\lim_{\nu \to \infty} \frac{\partial \psi(\overline{\beta}_\nu, x)}{\partial x} = \frac{\partial \mathcal{E}(x)}{\partial x}, \text{ a.e. in } X.$$

(b)

$$\lim_{\nu \to \infty} \sup_{x \in X} \left| \psi(\overline{\beta}_\nu, x) - \mathcal{E}(x) \right| = 0$$

(c)

$$limsup_{\nu \to \infty} \sup_{x \in X} \left| \frac{\partial \psi(\overline{\beta}_\nu, x)}{\partial x} \right| = \sup_{x \in X} \left| \frac{\partial \mathcal{E}(x)}{\partial x} \right| .$$

Proof:

To accommodate points of non-differentiability, our proof strategy involves smoothing the function $\mathcal{E}$. We first define and characterize kernel approximator functions. For $k = 1, 2, \cdots$, let $\kappa^k : \mathbb{R}^p \to \mathbb{R}^m$ be a continuous function for which $\int_{N^k} \kappa^k(t)dt = 1$, and let $A \subset X$ be a set of Lebesgue measure zero. Let $q : (X - A) \to \mathbb{R}^m$ be any arbitrary bounded continuous function. The kernel approximator function q^k is defined as follows:

$$q^k(x) \equiv \int_{N^k} q(x+t)\kappa^k(t)dt$$

In a technical appendix, available upon request, we prove that

$$\limsup_{k\to\infty} \sup_{x\in X} \mid q^k(x) \mid \; = \sup_{x\in X} \mid q(x) \mid . \tag{34}$$

Property (34) will be used in proving part (c) of the lemma.

To prove the lemma, choose $\overline{x} \in X$. By theorem (7.29) in Wheeden and Zygmund (1977)), since $\mathcal{E}$ is absolutely continuous, the derivative of $\mathcal{E}$ exists a.e., it is uniformly bounded, and we have

$$\int_{\overline{x}}^{x} \frac{\partial \mathcal{E}(x')}{\partial x} dx' = \mathcal{E}(x) - \mathcal{E}(\overline{x})$$

for all $x \in X$. Let $\partial\mathcal{E}^k$ denote the kernel approximator to $\frac{\partial \mathcal{E}(x)}{\partial x}$ and let $\mathcal{E}^k$ be defined as follows:

$$\mathcal{E}^k(x) \equiv \int_{\overline{x}}^{x} \partial\mathcal{E}^k(x')dx' + M^k, \tag{35}$$

where the scalar sequence $\{M^k\}$ is chosen so that $\mathcal{E}^k(\overline{x}) = \mathcal{E}(\overline{x})$ for all k. Clearly,

$$\partial\mathcal{E}^k(x) = \frac{\partial \mathcal{E}^k(x)}{\partial x},$$

so, in particular, $\partial\mathcal{E}^k$ is differentiable everywhere and,

$$\lim_{k\to\infty} \left| \frac{\partial \mathcal{E}(x)}{\partial x} - \frac{\partial \mathcal{E}^k(x)}{\partial x} \right| \; = \; 0 \tag{36}$$

pointwise, except on a set of Lebesgue measure zero.

We first prove that

$$\lim_{k\to\infty} \sup_{x\in X} \left|\mathcal{E}^k(x) - \mathcal{E}(x)\right| = 0. \tag{37}$$

Given our construction (35), it is enough to show that the integrals of $\partial\mathcal{E}^k$ converge uniformly to the integrals of $\frac{\partial\mathcal{E}}{\partial x}$.

Let $\widehat{K}$ denote the Lipschitz coefficient on $\mathcal{E}$, which exists by virtue of assumption 1(c). It follows that $\widehat{K}$ is a bound on $\frac{\partial\mathcal{E}(x)}{\partial x}$, and (by virtue of (34) on $\frac{\partial\mathcal{E}^k(x)}{\partial x}$. Using Egorov's theorem (see Wheeden and Zygmund (1977) page 57), we know that convergence a.e. of measurable functions implies uniform convergence except in a set of arbitrarily small measure. Therefore, given any $\epsilon > 0$, we can find $A \subset X$ such that $\mid A \mid < \epsilon/(4\ \widehat{K})$ (where, $\mid A \mid$ represents the Lebesgue measure of the set A) and

$$\sup_{x\in X-A} \left|\frac{\partial\mathcal{E}^k(x)}{\partial x} - \frac{\partial\mathcal{E}(x)}{\partial x}\right| \to 0 \quad \text{as} \quad k \to \infty$$

In particular, there exists a $\widehat{k}$ such that, for all $k > \widehat{k}$, $\sup_{x\in X-A} \left|\frac{\partial\mathcal{E}^k(x)}{\partial x} - \frac{\partial\mathcal{E}(x)}{\partial x}\right| < \epsilon/(2\,|X|)$. For such a k, Egorov's theorem implies that,

$$\sup_{x\in X} \left|\mathcal{E}^k(x) - \mathcal{E}(x)\right| = \sup_{x\in X} \left|\int_{\overline{x}}^{x} \frac{\partial\mathcal{E}^k(x')}{\partial x}dx' - \int_{\overline{x}}^{x} \frac{\partial\mathcal{E}(x')}{\partial x}dx'\right|$$

$$\leq \sup_{x\in X} \left(\int_{[\overline{x},x]-A} \left|\frac{\partial\mathcal{E}^k(x')}{\partial x} - \frac{\partial\mathcal{E}(x')}{\partial x}\right| dx' + \int_A \left|\frac{\partial\mathcal{E}^k(x')}{\partial x} - \frac{\partial\mathcal{E}(x')}{\partial x}\right| dx'\right)$$

$$\leq \sup_{x\in X} \left(\int_{[\overline{x},x]-A} \epsilon/(2\ \mid X \mid)dx' + \mid A \mid 2\widehat{K}\right) \leq \mid X \mid \epsilon/(2\ \mid X \mid) + 2\widehat{K}\epsilon/(4\ \widehat{K}) = \epsilon\,,$$

This completes the proof of (37).

Parts (a), (b), and (c) of the Lemma can now be demonstrated. Fix k. By Assumption 3 there exists a sequence $\overline{\beta}_\nu^k \in D_\nu$ such that

$$\lim_{\nu\to\infty} \sup_{x\in X} \left|\frac{\partial\psi(\overline{\beta}_\nu^k, x)}{\partial x} - \frac{\partial\mathcal{E}^k(x)}{\partial x}\right| = 0. \tag{38}$$

Equation (38) implies

$$\lim_{\nu\to\infty}\sup_{x\in X}\left|\psi(\overline{\beta}_\nu^k,x)-\mathcal{E}^k(x)\right|=0. \tag{39}$$

and

$$\lim_{\nu\to\infty}\sup_{x\in X}\left|\frac{\partial\psi(\overline{\beta}_\nu^k,x)}{\partial x}\right|=\sup_{x\in X}\left|\frac{\partial\mathcal{E}^k(x)}{\partial x}\right| \tag{40}$$

We know from the properties of kernel approximators that $\mathcal{E}^k\to\mathcal{E}$ as $k\to\infty$ and that, for each k; from (39) we have that $\psi_{\overline{\beta}_\nu^k}\to\mathcal{E}^k$ uniformly as $\nu\to\infty$. Furthermore, (34) implies that $\sup_{x\in X}\left|\frac{\partial\mathcal{E}^k(x)}{\partial x}\right|\to\sup_{x\in X}\left|\frac{\partial\mathcal{E}(x)}{\partial x}\right|$ as $k\to\infty$. Therefore, a sequence $\{\overline{\beta}_\nu\}$ can be constructed by taking the appropriate elements from the doubly indexed sequence $\{\overline{\beta}_\nu^k\}$ so that parts (a) , (b), and (c) of the Lemma are satisfied. QED

The next lemma proves that restricting the approximation to stay in the set of well-behaved laws of motion $\mathbb{B}$ does not preclude an arbitrarily good approximation. Formally, the lemma could be stated as saying that P is dense in $\mathbb{B}$.

Lemma 2 *Let $\left\{\overline{\beta}_\nu\right\}$ be as in lemma 1, then $\overline{\beta}_\nu\in\mathbb{B}_\nu$ for ν sufficiently large.*

Proof:

In the case where assumption 4* is made and $\mathbb{B}_\nu=\mathbb{B}_\nu^S$, the lemma follows immediately from part (c) of the definition of S, since H is in the interior of S. The remainder of the proof is for the case where assumption 4 is made and $\mathbb{B}_\nu=\mathbb{B}_\nu^{AUC}$. We use the following result:

$$\sup_{x\in X}\left|\frac{\mathcal{E}(x)-\mathcal{E}(\widehat{x})}{x-\widehat{x}}\right|=\sup_{x\in X}\left|\frac{\partial\mathcal{E}(x)}{\partial x}\right| \tag{41}$$

A proof of equation (41) is in a technical appendix available from the authors. Assumptions 1(a) and 3(a) imply that $H(z,u;\cdot)$ is continuous in β. Therefore, to prove the lemma, it is sufficient to show that, for each ν, the minimal Lipschitz coefficient of $H(\cdot,u;\overline{\beta}_\nu)$ (which we will denote $\rho_{\overline{\beta}_\nu}(u)$) satisfies

$$\limsup_{\nu\to\infty}\rho_{\overline{\beta}_\nu}(u)<\rho(u), \tag{42}$$

where $\rho(u)$ is the Lipschitz coefficient on H defined in Assumption 4. Under Assumptions 1(a) and 1(b), the functions g, g_2^{-1}, and f are uniformly Lipschitz in all arguments. Therefore, the following coefficients are finite:

$$K_1(u) = \max_{\alpha,\widehat{\alpha},z',z} \left| \frac{g(\alpha,z',z,u) - g(\widehat{\alpha},z',z,u)}{\alpha - \widehat{\alpha}} \right|$$

$$K_3(u) = \max_{\alpha,\widehat{z},z',z} \left| \frac{g(\alpha,z',z,u) - g(\alpha,z',\widehat{z},u)}{z - \widehat{z}} \right|$$

$$\overline{K}(u) = \max_{\alpha,\widehat{z},z} \left| \frac{g_2^{-1}(\alpha,z,u) - g_2^{-1}(\widehat{\alpha},\widehat{z},u)}{z - \widehat{z}} \right|$$

$$K_f(u) = \max_{\widehat{z},z} \left| \frac{f(z,u) - f(\widehat{z},u)}{z - \widehat{z}} \right|$$

where z, $\widehat{z}$, $z' \in Z$, and $\widehat{\alpha} \in \phi(Z)$. Assumption 4 then implies that

$$\left[K_1(u)K_f(u)\widehat{K} + K_3(u)\right] \overline{K}(u) \leq \rho(u) \tag{43}$$

where $\widehat{K} = \sup_{x \in X} \left| \frac{\partial \mathcal{E}(x)}{\partial x} \right|$. Let K_β be defined as follows:

$$K_\beta = \sup_{x \in X} \left| \frac{\partial \psi(\beta,x)}{\partial x} \right|$$

It follows that

$$\rho_\beta(u) = [K_1(u)K_f(u)K_\beta + K_3(u)]\,\overline{K}(u) \tag{44}$$

Equations (43), (44), and Lemma 1(c) immediately imply (42). **QED.**

The next Lemma collects a number of results that will be used later. Let $\mathcal{E}^*$ denote an arbitrary function defined on X; let $H^* : Z \times U \to Z$ be defined by

$$g(\mathcal{E}^*(f(z,u)), H^*(z,u), z, u) \equiv 0,$$

so H^* is the law of motion consistent with $\mathcal{E}^*$; let $\rho^*(u)$ be the Lipschitz coefficient of $H^*(\cdot,u)$.

If assumption 4* is made, we will assume that $H^* \in S$. If assumption 4 is made, we will assume that H^* satisfies the AUC condition in (31), so that

$$E\left[\log[\rho^*(u_t)]\right] \leq \alpha^* < 0 \quad \text{a.s.} \tag{45}$$

Let $z_t^{\infty,*}$ be a stationary and ergodic process satisfying $z_t^{\infty,*} = H^*(z_{t-1}^{\infty,*}, u_t)$. For any given $\bar{z} \epsilon Z$, define $\{z_t(\tau, \bar{z}; \beta)\}_{t=-\tau}^{\infty}$ recursively as follows:

$$z_{-\tau}(\tau, \bar{z}; \beta) = \bar{z}$$

$$z_t(\tau, \bar{z}; \beta) = H(z_{t-1}(\tau, \bar{z}; \beta), u_t; \beta), \quad t > -\tau,$$

so $z_t(\tau, \bar{z}; \beta)$ is obtained with the law of motion H_β, starting at $\bar{z}$ at date $-\tau$. Similarly, $\{z_t^*(\tau, \bar{z})\}_{t=-\tau}^{\infty}$ is obtained with the law of motion H^* and initial condition $z_{-\tau}(\tau, \bar{z}) = \bar{z}$. Note that $z_t(\tau, \bar{z}; \beta)$ and $z_t^*(\tau, \bar{z})$ depend only on $\{u_j\}_{j=-\tau+1}^{t}$, while the stationary processes $z_t^{\infty,*}$ and $z_t^{\infty}(\beta)$ depend on $\{u_j\}_{j=-\infty}^{t}$. Finally, let $\{\beta_\nu^*\}$ be a sequence such that

$$\lim_{\nu \to \infty} \sup_{x \in X} |\psi(\beta_\nu^*, x) - \mathcal{E}^*(x)| = 0$$

(Note that, according to Lemma 1, $\{\overline{\beta}_\nu\}$ plays the role of this sequence if $\mathcal{E}^* = \mathcal{E}$).

Lemma 3 *(a) $z_t^{\infty,*}$ exists, and $\lim_{\tau \to \infty} | z_t^*(\tau, \bar{z}) - z_t^{\infty,*} | = 0$ a.s.*

(b) For all $\beta \in \mathrm{B}$, $z_t^{\infty}(\beta)$ exists, and $\lim_{\tau \to \infty} | z_t(\tau, \bar{z}; \beta) - z_t^{\infty}(\beta) | = 0$ a.s.

(c) For any ν and any t, the restriction of $z_t^{\infty}(\beta)$ to $\mathbb{B}_\nu$ is continuous in β a.s.

(d) $\lim_{\nu \to \infty} \sup_{z \in Z, u \in U} | H(z, u; \beta_\nu^*) - H^*(z, u) | = 0.$

(e) $\lim_{\nu \to \infty} \sup_{\bar{z} \in Z, \forall \tau \geq 1} | z_t(\tau, \bar{z}; \beta_\nu^*) - z_t^*(\tau, \bar{z}) | = 0$ a.s.

Proof:

Under assumption 4* with $\mathrm{B}_\nu = \mathrm{B}_\nu^S$, parts (a) and (b) follow immediately from parts (a) and (b) of the definition of S. Under assumption 4 with $\mathrm{B}_\nu = \mathrm{B}_\nu^{AUC}$, parts (a) and (b) follow directly from Duffie and Singleton's [1993] Lemma 3. Part (c) is proven as follows: Let $\epsilon > 0$ be given. It is sufficient to show that $\exists \delta > 0$ such that

$$| \beta - \hat{\beta} | < \delta \text{ implies that } | z_t^{\infty}(\beta) - z_t^{\infty}(\hat{\beta}) | < \epsilon \quad a.s.$$

For any $\overline{z} \in Z$ and any τ,

$$\begin{aligned} \mid z_t^{\infty}(\hat{\beta}) - z_t^{\infty}(\beta) \mid \; \leq \; & \mid z_t^{\infty}(\beta) - z_t(\tau, \overline{z}; \beta) \mid \\ + \mid z_t(\tau, \overline{z}; \beta) - z_t(\tau, \overline{z}; \hat{\beta}) \mid \; + \; & \mid z_t(\tau, \overline{z}; \beta) - z_t^{\infty}(\hat{\beta}) \mid \end{aligned} \tag{46}$$

According to part (b) of this lemma, τ can be chosen big enough so the first and third terms on the right-hand side of (46) are less than $\epsilon/3$ almost surely. Furthermore, $H(z,u;\cdot) : \mathbf{B} \rightarrow Z$ is continuous, so $z_t(\tau, \overline{z}; \cdot) : \mathbf{B} \rightarrow Z$ is continuous. It follows that there exists $\delta > 0$ such that $\mid \beta - \hat{\beta} \mid < \delta$, and the second term on the right-hand side of (46) is less than $\epsilon/3$.

Part (d) of the lemma is proved as follows. Assumption 1(b) insures that g_2^{-1} is uniformly Lipschitz. Since

$$H^*(z,u) \equiv g_2^{-1}(\mathcal{E}^*(f(z,u)), z, u) \text{ and}$$

$$H(z,u;\beta_\nu^*) \equiv g_2^{-1}(\psi(\beta_\nu^*, f(z,u)), z, u)$$

we have, for some $K < \infty$,

$$\sup_{z,u} \mid H^*(z,u) - H(z,u;\beta_\nu^*) \mid$$

$$\leq \frac{1}{\epsilon} K \sup_{z,u} \mid \mathcal{E}^*(f(z,u)) - \psi(\beta_\nu^*, f(z,u)) \mid \; \rightarrow 0$$

as $\nu \rightarrow \infty$.

We now turn to part (e) of the Lemma. Under assumption 4* with $\mathbf{B}_\nu = \mathbf{B}_\nu^S$, part (e) follows immediately from part (d) of the lemma along with part (c) of the definition of S. The remainder of this proof treats the case where assumption 4 is made and where $\mathbf{B}_\nu = \mathbf{B}_\nu^{AUC}$. In this part of the proof, it is useful to notate explicitly the dependence of the z's on ω, the draw from the sample space, so we use the notation " $z_t(\tau, \overline{z}, \omega; \beta)$" and "$z_t^*(\tau, \overline{z}, \omega)$".

According to part (d) of this lemma, for any $\delta > 0$ $\exists$ $N(\delta) < \infty$ such that $\sup_{z,u} \mid H^*(z,u) - H(z,u;\beta_\nu^*) \mid \; \leq \delta, \forall \nu > N(\delta)$. Therefore, for date $-\tau + 1$,

$$\left| z_{-\tau+1}(\tau, \overline{z}, \omega; \beta_\nu^*) - z_{-\tau+1}^*(\tau, \overline{z}, \omega) \right| \; =$$

$$\mid H(\overline{z}, u_{-\tau+1}(\omega); \beta_\nu^*) - H^*(\overline{z}, u_{-\tau+1}(\omega)) \mid \; \leq \delta,$$

$\forall \nu > N(\delta)$. Proceeding recursively forward to date t, one obtains:

$$\mid z_t(\tau, \overline{z}, \omega; \beta_\nu^*) - z_t^*(\tau, \overline{z}, \omega) \mid \; \leq \; \delta \left[1 + \sum_{j=1}^{t+\tau-1} \prod_{k=1}^{j} \rho^*(u_{t-k+1}(\omega)) \right]$$

$\forall \nu > N(\delta)$. Since delta can be made arbitrarily small, part (e) of the lemma follows if $M(t,\omega,\tau) \equiv \sum_{j=1}^{t+\tau-1} \prod_{k=1}^{j} \rho^*(u_{t-k+1}(\omega))$ is bounded for almost all ω, with the bound uniform in τ. (Notice that $M(t,\omega,\tau)$ does not depend on $\overline{z}$.) Since $\log(\rho^*(u_t(\omega))$ is stationary and ergodic, the sample mean of $\log[\rho^*(u_t(\omega))]$ converges almost surely to its population mean. Therefore, equation (45) there exists a $J(t,\omega) < \infty$ such that, for almost all ω, $\frac{1}{j}\sum_{k=1}^{j} \log[\rho^*(u_{t-k+1}(\omega))] < \alpha^*, \forall j \geq J(t,\omega)$, and

$$\begin{aligned} M(t,\omega,\tau) &\leq \sum_{j=1}^{\infty} \prod_{k=1}^{j} \rho^*(u_{t-k+1}(\omega)) \\ &\leq \sum_{j=1}^{J(t,\omega)} \prod_{k=1}^{j} \rho^*(u_{t-k+1}(\omega)) + \sum_{j=1}^{\infty} [e^{\alpha^*}]^j \\ &= \sum_{j=1}^{J(t,\omega)} \prod_{k=1}^{j} \rho^*(u_{t-k+1}(\omega)) + e^{\alpha^*}/(1-e^{\alpha^*}) \equiv K(t,\omega) < \infty, \end{aligned}$$

for almost all ω. This proves part (e) of the lemma. QED.

The following lemma insures compactness of P in the supnorm.

Lemma 4 *The family of functions $P = \{\psi(\beta,\cdot), \beta \in D\}$ is equicontinuous.*

Proof:

Assumption 3 states that the derivative of elements of P is uniformly bounded, so

$$| \psi(\beta,\overline{x}) - \psi(\beta,x) | = \left| \int_{\overline{x}}^{x} \frac{\partial \psi(\beta,(x')}{\partial x} dx' \right| \leq K \mid x - \overline{x} \mid$$

QED.

The next lemma is a uniform strong law of large numbers for the mean square prediction error. For $\beta, \xi \in D$ let $h_t(\beta,\xi) \equiv \phi(z_{t+1}(\beta)) - \psi(\xi, x_t(\beta))$, let $h_t^{\infty}(\beta,\xi) \equiv \phi(z_{t+1}^{\infty}(\beta)) - \psi(\xi, x_t^{\infty}(\beta))$ and let $\gamma_{\nu,T}$ be defined by:

$$\gamma_{\nu,T}^{*} \equiv \sup_{\beta \in B_\nu, \xi \in D_\nu} \left\{ \mid \frac{1}{T} \sum_{t=1}^{T} [h_t(\beta,\xi)]^2 - E[h_t^{\infty}(\beta,\xi)]^2 \mid \right\}. \tag{47}$$

Lemma 5 $\lim_{T\to\infty} \gamma_{\nu,T}^{*} = 0$ a.s.

Proof:

The proof is sketched.

$$\begin{aligned} \gamma^*_{\nu,T} \;\leq\; & \sup_{\beta\in B_\nu,\xi\in D_\nu} \left|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} h_t(\beta,\xi)^2 - \tfrac{1}{T}\sum_{t=1}^{T} h_t^\infty(\beta,\xi)^2\right| \\ + \; & \sup_{\beta\in B_\nu,\xi\in D_\nu} \left|\tfrac{1}{T}\textstyle\sum_{t=1}^{T} h_t^\infty(\beta,\xi)^2 - E\; h_t^\infty(\beta,\xi)^2\right| \end{aligned} \tag{48}$$

$h_t^\infty(\beta,\xi)$ satisfies the conditions of Hansen's uniform strong law of large numbers for stationary and ergodic processes. (See Hansen [1982], Lemma 4.5.) It follows that the second term on the right-hand side of (48) converges to zero almost surely as $T \to \infty$. Under assumption 4* with $\mathbf{B}_\nu = \mathbf{B}_\nu^S$, the first term on the right-hand side of (48) converges to zero almost surely as an immediate implication of part (b) of the definition of S. On the other hand, if assumption 4 is made and $\mathbf{B}_\nu = \mathbf{B}_\nu^{AUC}$, it can be shown that the first term on the right-hand side of (48) also converges to zero almost surely by using a slight modification of the proof of Lemma 4 in Duffie and Singleton (1993). **QED.**

The following lemma insures that the non-linear regressions converge to the population least squares minimizer.

Lemma 6 $\lim_{T\to\infty} \sup_{\beta\in\mathbf{B}_\nu} \mid G_{\nu,T}(\beta) - G_\nu(\beta) \mid \;=\; 0, \quad a.s.$

Proof:

Let $\epsilon > 0$ be given. Let set $Q(\epsilon,\beta) \in D_\nu$ be the set of elements of D_ν which are bounded away from $G_\nu(\beta)$. Formally:

$$Q(\epsilon,\beta) \equiv \{\xi \in D_\nu \text{ s.t. } \mid \xi - G_\nu(\beta) \mid \;\geq \epsilon\} \tag{49}$$

Define $\pi(\epsilon,\beta) > 0$ by

$$\pi(\epsilon,\beta) = \inf_{\xi\in Q(\epsilon,\beta)} \left\{E[h_t^\infty(\beta,\xi,\omega)^2] - E[h_t^\infty(\beta,G_\nu(\beta),\omega)]^2\right\} \tag{50}$$

and let $\pi(\epsilon)$ be defined by

$$\pi(\epsilon) = \inf_{\beta\in\mathbf{B}_\nu} \pi(\epsilon,\beta) \tag{51}$$

Compactness of D_ν implies that

$$\pi(\epsilon) > 0 \tag{52}$$

Equation (52) and Lemma 5 imply that $\exists\ \Lambda\ \in\ \Omega$ with prob $(\Lambda) = 1$ and $\exists\ T(\omega, \epsilon)$ such that, for $\forall\ \omega\ \epsilon\ \Lambda$ and $\forall\ T > T(\omega, \epsilon)$,

$$0 \leq \gamma^{\circ}_{\nu,T}(\omega) < \frac{\pi(\epsilon)}{2} \tag{53}$$

where $\gamma^{\circ}_{\nu,T}(\omega)$ is defined in (47). Choose some $\omega\ \epsilon\ \Lambda, T > T(\omega, \epsilon)$, and arbitrary $\beta \in \mathbb{B}_{\nu}$.

¿From the definition of $G_{\nu,T}$,

$$\begin{aligned}
0 &\leq E[h^{\infty}_t(\beta, G_{\nu,T}(\beta,\omega),\omega)]^2 - E[h^{\infty}_t(\beta, G_{\nu}(\beta),\omega)]^2 \\
&\leq E[h^{\infty}_t(\beta, G_{\nu,T}(\beta,\omega),\omega)]^2 - \frac{1}{T}\sum_{t=1}^{T}[h_t(\beta, G_{\nu,T}(\beta),\omega)]^2 \\
&+\frac{1}{T}\sum_{t=1}^{T}[h_t(\beta, G_{\nu,T}(\beta),\omega)]^2 - \frac{1}{T}\sum_{t=1}^{T}[h_t(\beta, G_{\nu}(\beta),\omega)]^2 \\
&+\frac{1}{T}\sum_{t=1}^{T}[h_t(\beta, G_{\nu}(\beta),\omega)]^2 - E[h^{\infty}_t(\beta, G_{\nu}(\beta),\omega)]^2 \\
&\leq \left|E[h^{\infty}_t(\beta, G_{\nu,T}(\beta,\omega),\omega)]^2 - \frac{1}{T}\sum_{t=1}^{T}[h_t(\beta, G_{\nu,T}(\beta),\omega)]^2\right| \\
&+\left|\frac{1}{T}\sum_{t=1}^{T}[h_t(\beta, G_{\nu}(\beta),\omega)]^2 - E[h^{\infty}_t(\beta, G_{\nu}(\beta),\omega)]^2\right| \\
&\leq 2\gamma^{\circ}_{\nu,T}(\omega) < \pi(\epsilon) \leq \pi(\epsilon,\beta)
\end{aligned}$$

where the third inequality follows from the definition of $G_{\nu,T}(\beta,\omega)$, the fourth inequality follows from (53), and the fifth inequality follows from (51). This implies that $G_{\nu,T}(\beta,\omega) \notin Q(\beta,\epsilon)$, proving the lemma. **QED.**

Lemma 7 states that the approximate solution for large enough T will be arbitrarily close to the "population" approximate solution of order ν :

$$\beta_{\nu} = \arg\min_{\beta\in\mathbf{B}_{\nu}} \parallel \psi(\beta, x^{\infty}_t(\beta)) - \psi(G_{\nu}(\beta), x^{\infty}_t(\beta)) \parallel \tag{54}$$

Lemma 7 *Let $\beta_{\nu,T}$ be the approximate solution, as in equation (32). Then, given ν, we have*

$$\lim_{T\to\infty} \mid \beta_{\nu,T} - \beta_{\nu} \mid \ = \ 0, \text{a.s.}$$

Proof:

We first show that the minimand in (54) converges to the minimand in (18) uniformly in β as $T \to \infty$:

$$\sup_{\beta \in \mathbf{B}_\nu} \left| \left| \frac{1}{T} \sum_{t=1}^{T} \psi(\beta, x_t(\beta)) - \psi(G_{\nu,T}(\beta), x_t(\beta)) \right|^2 - E \mid \psi(\beta, x_t^\infty(\beta)) - \psi(G_\nu(\beta), x_t^\infty(\beta)) \mid^2 \right|$$

$$\leq \frac{1}{T} \sum_{t=1}^{T} \sup_{\beta \in \mathbf{B}_\nu} \Big| \mid \psi(\beta, x_t(\beta)) - \psi(G_{\nu,T}(\beta), x_t(\beta)) \mid^2 - |\psi(\beta, x_t(\beta)) - \psi(G_\nu(\beta), x_t(\beta))|^2 \Big| \tag{55}$$

$$+ \sup_{\beta \in \mathbf{B}_\nu} \left| \frac{1}{T} \sum_{t=1}^{T} |\psi(\beta, x_t(\beta)) - \psi(G_\nu(\beta), x_t(\beta))|^2 - E \mid \psi(\beta, x_t^\infty(\beta)) - \psi(G_\nu(\beta), x_t^\infty(\beta)) \mid^2 \right|$$

We first show that the first term on the right-hand side of (55) converges to zero. According to assumption 3(a), $\psi(\cdot, x)$ is Lipschitz, uniformly in $x \in X$. Furthermore, the $\psi : \mathbf{B}_\nu \times X \to \mathbf{R}^m$ is a continuous function of a compact set, so is bounded. These conditions imply that $|\psi(\beta, x_t(\beta)) - \psi(\xi, x_t(\beta))|^2$ is Lipschitz in $\xi \in \mathbf{B}_\nu$, uniformly in $\beta \in \mathbf{B}_\nu$, so $\exists K < \infty$ such that

$$\Big| \mid \psi(\beta, x_t(\beta)) - \psi(G_{\nu,T}(\beta), x_t(\beta)) \mid^2 - |\psi(\beta, x_t(\beta)) - \psi(G_\nu(\beta), x_t(\beta))|^2 \Big|$$

$$\leq K \mid G_{\nu,T}(\beta) - G_\nu(\beta) \mid \tag{56}$$

Lemma 6 implies that the right-hand side of (56) converges uniformly to zero as $T \to \infty$, so, given $\epsilon > 0, \exists\ T(\epsilon)$ such that $\forall\ T > T(\epsilon)$,

$$\sup_{\beta \in \mathbf{B}_\nu} \mid G_{\nu,T}(\beta) - G_\nu(\beta) \mid \leq \frac{\epsilon}{K} \tag{57}$$

Equations (56) and (57) imply that the first term on the right-hand side of (55) is less than ϵ for all $T > T(\epsilon)$, which proves that term converges to zero as $T \to \infty$.

We now complete the proof. By an argument analogous to that used in the proof of Lemma 5, $\mid \psi(\beta, x_t(\beta)) - \psi(G_\nu(\beta), x_t(\beta)) \mid^2$ satisfies a uniform strong law of large numbers. This implies that the second term on the right-hand side of (55) converges to zero as $T \to \infty$. Thus, the left-hand side of

(55) converges to zero, so the minimand in (18) converges uniformly to the minimand of (54). Since the minimization is over the compact set $\mathbb{B}_\nu$, this implies the conclusion of the lemma. QED.

Lemma 8 *Let* $\{\tilde{\beta}_\nu\}_{\nu=1}^{\infty}$ *be a sequence such that* $\tilde{\beta}_\nu \in \mathbb{B}_\nu, \forall \nu$. *Then*

$$\lim_{\nu \to \infty} \left\| \mathcal{E}(x_t^\infty(\tilde{\beta}_\nu), \tilde{\beta}_\upsilon) - \psi(G_\nu(\tilde{\beta}_\nu), x_t^\infty(\tilde{\beta}_\nu)) \right\| = 0 \tag{58}$$

Proof:

According to Lemma 4, $\{\psi(\beta, \cdot), \beta \in B\}$ is an equicontinuous family of functions. Assumptions 1(a) and 1(b) then imply that $\{\mathcal{E}(\cdot, \beta), \beta \in B\}$ is also equicontinuous. Furthermore, since ϕ is a bounded function, it follows that $\{\mathcal{E}(\cdot, \beta), \beta \in B\}$ is uniformly bounded and Arzela's theorem (Kolmogorov and Fomin (1970, p. 102)) then implies that there exists a subsequence ν_k and a continuous function $\widehat{\mathcal{E}}$ such that

$$\lim_{k \to \infty} \sup_{x \in X} \mid \mathcal{E}(x, \tilde{\beta}_{\nu_k}) - \widehat{\mathcal{E}}(x) \mid = 0. \tag{59}$$

The function $\widehat{\mathcal{E}}$ is the uniform limit of a sequence of uniformly Lipschitz functions, so $\widehat{\mathcal{E}}$ satisfies a Lipschitz condition. By a proof analogous to Lemma 1(b), there exists a sequence $\left\{\widehat{\beta}_{\nu_k}\right\}_{k=1}^{\infty}$ with $\widehat{\beta}_{\nu_k} \in D_{\nu_k}$ such that

$$\lim_{k \to \infty} \sup_{x \in X} \mid \widehat{\mathcal{E}}(x) - \psi(\widehat{\beta}_{\nu_k}, x) \mid = 0. \tag{60}$$

(Notice that this sequence is different from $\left\{\tilde{\beta}_{\nu_k}\right\}$ in the statement of the lemma)

Equation (58) is now demonstrated along subsequence ν_k. The definition of G implies that

$$G_{\nu_k}(\tilde{\beta}_{\nu_k}) = \arg \min_{\xi \in D_{\nu_k}} \left\| \mathcal{E}(x_t^\infty(\tilde{\beta}_{\nu_k}), \tilde{\beta}_{\nu_k}) - \psi(\xi, x_t^\infty(\tilde{\beta}_{\nu_k})) \right\|, \tag{61}$$

since the expectational error is orthogonal to the least squares predictor.

The following chain of inequalities hold:

$$
\begin{aligned}
&\| \mathcal{E}(x_t^{\infty}(\tilde{\beta}_{\nu_k}), \tilde{\beta}_{\nu_k}) - \psi(G_{\nu_k}(\tilde{\beta}_{\nu_k}), x_t^{\infty}(\tilde{\beta}_{\nu_k})) \| \\
&\quad \leq \| \mathcal{E}(x_t^{\infty}(\tilde{\beta}_{\nu_k}), \tilde{\beta}_{\nu_k}) - \psi(\hat{\beta}_{\nu_k}, x_t^{\infty}(\tilde{\beta}_{\nu_k})) \| \\
&\quad \leq \sup_{x \in X} \mid \mathcal{E}(x, \tilde{\beta}_{\nu_k}) - \psi(\hat{\beta}_{\nu_k}, x) \mid \\
&\quad \leq \sup_{x \in X} \mid \mathcal{E}(x, \tilde{\beta}_{\nu_k}) - \hat{\mathcal{E}}(x) \mid + \sup_{x \in X} \mid \hat{\mathcal{E}}(x) - \psi(\hat{\beta}_{\nu_k}, x) \mid \rightarrow 0
\end{aligned}
\tag{62}
$$

where the first inequality follows from (61) and the third inequality follows from the triangle inequality. The first term on the right-hand side of (62) converges to zero as $k \rightarrow \infty$ by (59). The second term on the right-hand side of (62) converges to zero as $k \rightarrow \infty$ by (60).

Since (62) holds for any arbitrary convergent subsequence, this result implies the conclusion of the lemma. QED.

The next and last lemma shows that β_ν gets arbitrarily close to $G_\nu(\beta_\nu)$ as ν grows. In this sense, we say that an asymptotic fixed point exists.

Lemma 9

$$
\lim_{\nu \rightarrow \infty} \left\{ \min_{\beta \in \mathbf{B}_\nu} \| \psi(\beta, x_t^{\infty}(\beta)) - \psi(G_\nu(\beta), x_t^{\infty}(\beta)) \| \right\} = 0
$$

Proof.

According to Lemma 1(b), a sequence $\{\overline{\beta}_\nu\}_{\nu=1}^{\infty}$ exists such that

$$
\lim_{\nu \rightarrow \infty} \sup_{x \in X} \mid \mathcal{E}(x) - \psi(\overline{\beta}_\nu, x) \mid = 0 \tag{63}
$$

Recall that $\mathcal{E}$ is the true conditional expectation. According to Lemma 2, $\overline{\beta}_\nu \in \mathbf{B}_\nu$, so the minimand in the statement of the lemma is dominated by

$$
\left\| \psi(\overline{\beta}_\nu, x_t^{\infty}(\overline{\beta}_\nu)) - \psi(G_\nu(\overline{\beta}_\nu), x_t^{\infty}(\overline{\beta}_\nu)) \right\| \tag{64}
$$

and it is enough to show that (64) goes to zero. Now,

$$
\begin{aligned}
&\left\| \psi(\overline{\beta}_\nu, x_t^{\infty}(\overline{\beta}_\nu)) - \psi(G_\nu(\overline{\beta}_\nu), x_t^{\infty}(\overline{\beta}_\nu)) \right\| \\
&\quad \leq \left\| \psi(\overline{\beta}_\nu, x_t^{\infty}(\overline{\beta}_\nu)) - \mathcal{E}(x_t^{\infty}(\overline{\beta}_\nu)) \right\| + \left\| \mathcal{E}(x_t^{\infty}(\overline{\beta}_\nu)) - \mathcal{E}(x_t^{\infty}(\overline{\beta}_\nu); \overline{\beta}_\nu) \right\| \\
&\quad + \left\| \mathcal{E}(x_t^{\infty}(\overline{\beta}_\nu); \overline{\beta}_\nu) - \psi(G_\nu(\overline{\beta}_\nu), x_t^{\infty}(\overline{\beta}_\nu)) \right\|
\end{aligned}
\tag{65}
$$

The first term in the right side goes to zero as $\nu \to \infty$ by the choice of $\left\{\overline{\beta}_\nu\right\}$. The second term is bounded by

$$sup_{z\times u\in\overline{ZU}}\left|\int \phi(H(z,u'))\, dF(u'|u) - \int \phi(H(z,u';\overline{\beta}_\nu))\, dF(u'|u)\right|$$

where $F(\cdot|u)$ is the conditional distribution of u_{t+1} given $u_t = u$. The above expression converges to zero by Lemma 3 (d) and the Lipschitz-continuity of ϕ. Finally, the third term goes to zero by Lemma 8. QED.

Proof of Proposition 1

In this proof, convergence of a function of (z,x) in the supnorm is taken over the ergodic set of the true process $\overline{ZU} \subset Z \times U$.

Let $\psi_\beta = \psi(\beta,\cdot)$ and $H_\beta = H(\cdot,\cdot;\beta)$. We first show that ψ_{β_ν} converges to $\mathcal{E}$ pointwise.

Since P is a uniformly bounded, equicontinuous class of functions, it is a compact set in the sup norm. Therefore, there exists a subsequence (indexed by k) and a limiting function ψ^* such that $\psi_{\beta_{\nu_k}} \to \psi^*$ in the supnorm. It is enough if we show that, for any such convergent subsequence, $\psi^* = \mathcal{E}$ a.e. in $\overline{ZU}$.

Let H^* be the law of motion consistent with ψ^*, and $\mathcal{E}^*$ the true conditional expectation consistent with H^*. We have $H_{\beta_{\nu_k}} \to H^*$ uniformly, by an argument analogous to that used in the proof of Lemma 3(d). Note that

$$\mathcal{E}_{\beta_{\nu_k}}(f(z,u)) \equiv \int_U \phi\left[H_{\beta_{\nu_k}}\left(H_{\beta_{\nu_k}}(z,u),u'\right)\right]\, dF(u'|u). \tag{66}$$

All the functions inside the integral are continuous and bounded, so Lebesgue Dominated Convergence implies

$$\mathcal{E}_{\beta_{\nu_k}}(f(z,u)) \to \int_U \phi\left[H^*\left(H^*(z,u),u'\right)\right]\, dF(u'|u) \equiv \mathcal{E}^*(f(z,u)) \tag{67}$$

as $k \to \infty$.

We will eventually show that $H = H^*$; to this end, we first have to prove that H^* satisfies assumption 4* (if $\mathrm{B}_\nu = \mathrm{B}_\nu^S$) or assumption 4 (if $\mathrm{B}_\nu = \mathrm{B}_\nu^{AUC}$). In the former case, $H_{\beta_{\nu_k}} \in S, \forall k$, by the definition of B_ν^S. Since S is closed under the sup norm, and $H_{\beta_{\nu_k}} \to H^*$ uniformly, it follows that $H^* \in S$.

We now turn to the latter case. It is sufficient to show that, if $\mathbf{B}_\nu = \mathbf{B}_\nu^{AUC}$, H^* has a Lipschitz coefficient satisfying (45). Let

$$\rho^*(u) \equiv \sup_{z,\bar{z}\in Z} \left| \frac{H^*(z,u) - H^*(\bar{z},u)}{z - \bar{z}} \right|$$

Fix $\epsilon > 0$, and any $u \in U$. By definition of $\rho^*(u)$, there exist z and $\bar{z}$ such that

$$\rho^*(u) \leq \frac{|H^*(z,u) - H^*(\bar{z},u)|}{|z - \bar{z}|}(1+\epsilon)$$

$$\leq \frac{\left|H^*(z,u) - H_{\beta_{\nu_k}}(z,u)\right| + \left|H_{\beta_{\nu_k}}(z,u) - H_{\beta_{\nu_k}}(\bar{z},u)\right| + \left|H_{\beta_{\nu_k}}(\bar{z},u) - H^*(\bar{z},u)\right|}{|z - \bar{z}|}(1+\epsilon)$$

$$\left[\frac{\left|H^*(z,u) - H_{\beta_{\nu_k}}(z,u)\right| + \left|H_{\beta_{\nu_k}}(\bar{z},u) - H^*(\bar{z},u)\right|}{|z - \bar{z}|} + \rho_{\beta_{\nu_k}}(u)\right](1+\epsilon) \tag{68}$$

where $\rho_{\beta_{\nu_k}}(u)$ is the function defined in (31) Now, fix $z, \bar{z}$; take logs on both sides of (68). Since inequality (68) holds for all k, we can take the $\liminf_k$ on both sides and use the uniform convergence of $H_{\beta_{\nu_k}} \to H^*$ to obtain

$$\log\ \rho^*(u) \leq \left(\liminf_{k\to\infty}\ \log\ \rho_{\beta_{\nu_k}}(u)\right) + \log(1+\epsilon) \tag{69}$$

which holds for all $\epsilon > 0$ and all u. Now, given any ϵ, taking expectations over u we get

$$E\left[\log\ \rho^*(u)\right] \leq E\left[\liminf_{k\to\infty}\ \log\ \rho_{\beta_{\nu_k}}(u)\right] + \log(1+\epsilon)$$

$$\leq \liminf_{k\to\infty}\ \left[E \log\ \rho_{\beta_{\nu_k}}(u)\right] + \log(1+\epsilon) \leq \alpha + \log(1+\epsilon)$$

where the second inequality follows from Fatou's lemma, and the last inequality follows from the fact that $\beta_{\nu_k} \in B_\nu$. Since the above inequality holds for all $\epsilon > 0$, we conclude

$$E\ \log\ \rho^*(u) \leq \alpha$$

This completes the demonstration that H° satisfies (45) when $\mathbb{B}_\nu = \mathbb{B}_\nu^{AUC}$.

We have shown that H° satisfies assumption 4* or 4 (depending on the definition of $\mathbb{B}_\nu$ being used). It follows that a stationary and ergodic process $\{z_t^{\circ\infty}\}$ exists satisfying $z_t^{\circ\infty} = H^\circ(z_{t-1}^{\circ\infty}, u_t)$, so the L^2 norm of $\{z_t^{\circ\infty}\}$ is well defined. We can write,

$$\begin{aligned}
&\|\psi^\circ(x_t^{\circ\infty}) - \mathcal{E}^\circ(x_t^{\circ\infty})\| \;\leq \\
&\|\psi^\circ(x_t^{\circ\infty}) - \psi(\beta_{\nu_k}, x_t^{\circ\infty})\| \;+\; \|\psi(\beta_{\nu_k}, x_t^{\circ\infty}) - \psi(\beta_{\nu_k}, x_t^{\infty}(\beta_{\nu_k}))\| \\
&\quad + \|\psi(\beta_{\nu_k}, x_t^{\infty}(\beta_{\nu_k})) - \psi(G_{\nu_k}(\beta_{\nu_k}), x_t^{\infty}(\beta_{\nu_k}))\| \\
&\quad + \|\psi(G_{\nu_k}(\beta_{\nu_k}), x_t^{\infty}(\beta_{\nu_k})) - \mathcal{E}(x_t^{\infty}(\beta_{\nu_k}), \beta_{\nu_k})\| \\
&\|\mathcal{E}(x_t^{\infty}(\beta_{\nu_k}), \beta_{\nu_k}) - \mathcal{E}(x_t^{\circ\infty}, \beta_{\nu_k})\| \;+\; \|\mathcal{E}(x_t^{\circ\infty}, \beta_{\nu_k}) - \mathcal{E}(x_t^{\circ\infty})\| \,. \qquad (70)
\end{aligned}$$

We can now apply Lemma 3, parts (a), (b), and (e), to conclude

$$|x_t^{\circ\infty} - x_t^{\infty}(\beta_{\nu_k})| \to 0 \text{ a.s.} \qquad (71)$$

The limsup of the right side of (70) equals zero, as follows: The limsup of the first term is equal to zero by the definition of ψ°; the second term goes to zero by (71), along with the continuity of $\psi_{\beta_{\nu_k}}$, and Lebesgue Dominated Convergence; the third term goes to zero by Lemma 9; the fourth term goes to zero by Lemma 8, the fifth term goes to zero by (71), along with the continuity of $\mathcal{E}(\cdot, \beta_{\nu_k})$, and Lebesgue Dominated Convergence; the sixth term goes to zero by (67).

We have proved that the left side of (70) is equal to zero, which implies $\psi^\circ(x_t^{\circ\infty}) = \mathcal{E}^\circ(x_t^{\circ\infty})$ almost surely in the support of $x_t^{\circ\infty}$. We can summarize this derivation as follows:

$$\psi(\beta_{\nu_k}, x) \to \psi^\circ(x) = E\left(\phi(z_{t+1}^{\circ\infty}) \mid x_t^{\circ\infty} = x\right) \qquad (72)$$

This is an important result in itself; it says that the limit of the parameterized expectations ψ° is equal to the true expectation of the process generated by ψ°.

Finally, since

$$g\left(\psi(\beta_{\nu_k}, x_t(\beta_{\nu_k})), z_t(\beta_{\nu_k}), z_{t-1}(\beta_{\nu_k}), u_t\right) = 0, \;\; \forall k$$

by construction, and g is continuous, (72) and lemma 3(e) imply

$$g\left(E\left(\phi(z_{t+1}^*) \mid x_t^*\right), z_t^*, z_{t-1}^*, u_t\right) = 0 ,$$

with probability one, which implies that H^* is a rational expectations equilibrium or, equivalently, that $\mathcal{E}^* \equiv \mathcal{E}$ and $H \equiv H^*$. To summarize, we have shown that any convergent subsequence of H_{β_ν} goes to H in the supnorm, which implies that the same holds for the whole sequence.

To complete the proof, we have to show the same type of convergence for the sample version of the approximate solution. It is enough to show

$$\underset{T \to \infty}{\text{Lim}} \quad \sup_{x \in X} \; |\psi(x, \beta_{\nu,T}) - \psi(x, \beta_\nu)| \; = \; 0. \tag{73}$$

Lemma 7 and continuity of ψ_β in β imply $\psi_{\beta_{\nu,T}} \to \psi_{\beta_\nu}$ pointwise. Since P is equicontinuous, by Lemma 4, and bounded it is compact in the sup norm, which implies uniform convergence as in (73). Finally, (72), (73), and Lemma 3(d) imply that $H_{\beta_{\nu,T}}$ converges uniformly to H^*. This completes the proof of Proposition 1.

Proof of Corollary

a) Follows from lemma 3(e) and the uniform convergence of $H_{\beta_{\nu,T}} \to H$ as in Proposition 1.

b) Lemmas 8 and 9 imply $\| \mathcal{E}(x_t^\infty(\beta_{\nu,T}), \beta_{\nu,T}) - \psi(\beta_{\nu,T}, x_t(\beta_{\nu,T})) \| \to 0$. Part b) of the corollary follows from part a) and continuity of g.

c) Follows from proposition 1 and part (b) of the definition of S (if assumption 4* is imposed and $\mathrm{B}_\nu = \mathrm{B}_\nu^S$) or from the AUC condition, Proposition 1 and Lemma 4 in Duffie and Singleton (if assumption 4 is imposed and $\mathrm{B}_\nu = \mathrm{B}_\nu^{AUC}$).

Proof of Proposition 2

We give the proof for the case where $\mathrm{B}_\nu = \mathrm{B}_\nu^{AUC}$. The proof for the case where $\mathrm{B}_\nu = \mathrm{B}_\nu^S$ is analogous, and is therefore omitted.

- We first prove that if

$$\min_{\beta \in B_\nu} E[\psi(\beta, x_t^\infty(\beta)) - \psi(G_\nu(\beta), x_t^\infty(\beta))]^2 \to 0 \tag{74}$$

then assumption 4 is satisfied. Consider any convergent subsequence of $\{\psi_{\beta_{\nu,T}}\}$; as in the proof of proposition 1, this subsequence has a limit ψ° for which the corresponding law of motion H° satisfies the AUC condition (30). As in the proof of proposition 1, (70) holds for this ψ° and the corresponding conditional expectation $\mathcal{E}^\circ$ (defined as in (67)) . Now, consider the right side of (70); with the exception of the third term, all terms go to zero as $k \to \infty$ if Assumptions 1, 2 and 3 are satisfied[19]. The third term goes to zero because of (74). This proves the existence of an equilibrium H° that satisfies AUC and that Assumption 4 is satisfied.

To prove part *(a)* assume, towards a contradiction, that

$$\liminf_{\nu\to\infty} \min_{\beta\in B_\nu} E[\psi(\beta, x_t^\infty(\beta)) - \psi(G_\nu(\beta), x_t^\infty(\beta))]^2 = 0.$$

Then there would be a subsequence $\{\beta_{\nu_k}\}$ for which the min goes to zero and we can find a sub-subsequence $\{\beta_{\nu_{k_i}}\}$ such that $H_{\beta_{\nu_{k_i}}} \to H^\circ$. By the argument given above, this would imply existence of H° that satisfies Assumption 4, which is a contradiction.

- We now prove part *(b)*. Let $Q \equiv$ { laws of motion for z that satisfy AUC conditi By an argument analogous to that used in proving that H° of Proposition 1 satisfies AUC, we can prove that Q is a closed set. Since each $H_{\beta_{\nu,T}} \in Q$ and $\tilde{H} \notin Q$, part *(b)* follows immediately.

QED

[19]Notice that assumption 4 has only been used in results leading to Lemma 9, which causes the third term in (70) to go to zero.

APPENDIX 2
NON-LINEAR LEAST SQUARES LEARNING

The recursive least squares estimator to predict $\phi(\tilde{z}_{t+1})$ can be written as

$$\begin{aligned} \beta_t &= \beta_{t-1} + \alpha_t \, R_t^{-1} \left(\frac{\partial \psi(\beta_t, \tilde{x}_{t-1})}{\partial \beta} \left[\phi(\tilde{z}_t) - \psi(\beta_t, \tilde{x}_{t-1}) \right] \right) \\ R_t &= R_{t-1} + \alpha_t \, \frac{\partial \psi(\beta_t, \tilde{x}_{t-1})}{\partial \beta} \, \frac{\partial \psi(\beta_t, \tilde{x}_{t-1})}{\partial \beta'} \\ g\left(\psi(\beta_t, \tilde{x}_t), \tilde{z}_t, \tilde{z}_{t-1}, u_t\right) &= 0 \quad . \end{aligned} \tag{75}$$

Notice that if we fixed β, replaced $\tilde{z}_t$ by $z_t(\beta)$, and for the particular choice $\alpha_t = (1/t)$, β_t is the non-linear least squares estimator to predict $\phi(z_t(\beta))$ with $\psi(\cdot, x_{t-1}(\beta))$, where the series z are not affected by β_t, except that the residuals are calculated recursively.

In the system defined by (75), however, the $\tilde{z}$'s depend on the estimate β_t. In words, (75) describes an economy where agents use recursive non-linear least squares to form their expectations about the future, and where equilibrium is generated by such expectations; the first equation says that today's beliefs are updated according with last period's beliefs and the prediction error made this period, the second equation gives the gradient for how the prediction error affects the beliefs β_t, and the last equation defines the equilibrium $\tilde{z}_t$ in terms of the expectations of the agents. This is a non-linear version of economies studied Marcet and Sargent (1989), where the non-linearities appear both in the equation generating the observations and in the expectational rules.

Using notation

$$D\psi_t(\overline{\beta}) \equiv \left| \partial \psi(\beta, x_{t-1}(\overline{\beta})) / \partial \beta \right|_{\beta = \overline{\beta}} \qquad \text{and} \qquad R(\overline{\beta}) \equiv E(\; D\psi_t(\overline{\beta}) D\psi_t(\overline{\beta})' \;)$$

and adapting results from Ljung [1975], it is possible to show that the only β's where least squares learning can converge with positive probability are $\overline{\beta} \in D_\nu$ that satisfy

$$R(\overline{\beta})^{-1} E\left(\; D\psi_t(\overline{\beta}) \; \left[\phi(z_t(\overline{\beta})) - \psi(\overline{\beta}, x_{t-1}(\overline{\beta})) \right] \right) = 0. \tag{76}$$

This equation is obtained by taking the object multiplying α_t in the recursive algorithm, fixing $\beta = \overline{\beta}$, and taking the expectation with respect to the

stationary distribution of $\{z_t(\beta)\}$. Now, since the first order conditions for the maximization problem of (16) imply that

$$E\left(D\psi_t(\bar{\beta})\ [\phi(z_t(\beta)) - \psi(G_\nu(\beta), x_{t-1}(\beta))]\right) = 0.$$

it is clear that (76) is satisfied, precisely, at the fixed point β_ν.

Further results in Ljung [1975] guarantee that $\beta_t \to \beta_\nu$ if the differential equation

$$\frac{\partial\beta(\tau)}{\partial\tau} = R(\beta(\tau))^{-1}E\left(\ D\psi_t(\beta(\tau))\ [\phi(z_t(\beta(\tau))) - \psi(\beta(\tau), x_{t-1}(\beta(\tau)))]\right) \tag{77}$$

is stable. To analyze this differential equation, notice that the right side of (77) is equal to

$$R(\beta)^{-1}E\left(\ D\psi_t(\beta)\ [\psi(G_\nu(\beta), x_{t-1}(\beta)) - \psi(\beta, x_{t-1}(\beta))]\right)$$

by the definition of G_ν. By the mean value theorem, in a neighborhood of β_ν this expression can be written as

$$R(\beta)^{-1}E\left(\ D\psi_t(\tilde{\beta})\ D\psi_t(\tilde{\beta})'\ (G_\nu(\beta) - \beta)\right)$$

for some $\tilde{\beta}$ in this neighborhood. Therefore, $\beta_t \to \beta_\nu$ locally if and only if the differential equation

$$\frac{\partial\beta(\tau)}{\partial\tau} = R(\beta(\tau))^{-1}E\left(\ D\psi_t(\beta(\tau))\ D\psi_t(\beta(\tau))'\right)\ (G_\nu(\beta(\tau)) - \beta(\tau)) = G_\nu(\beta(\tau)) - \beta(\tau)$$

is locally stable at β_ν.

RECENT WORKING PAPERS

1. Albert Marcet and Ramon Marimon
Communication, Commitment and Growth. (June 1991)
[Published in *Journal of Economic Theory* Vol. 58, no. 2, (December 1992)]

2. Antoni Bosch
Economies of Scale, Location, Age and Sex Discrimination in Household Demand. (June 1991)
[Published in *European Economic Review* 35, (1991) 1589-1595]

3. Albert Satorra
Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures. (June 1991)
[Published in *Sociological Methodology* (1992), pp. 249-278, P.V. Marsden Edt. Basil Blackwell: Oxford & Cambridge, MA]

4. Javier Andrés and Jaume Garcia
Wage Determination in the Spanish Industry. (June 1991)
[Published as "Factores determinantes de los salarios: evidencia para la industria española" in J.J. Dolado et al. (eds.) *La industria y el comportamiento de las empresas españolas (Ensayos en homenaje a Gonzalo Mato)*, Chapter 6, pp. 171-196, Alianza Economia]

5. Albert Marcet
Solving Non-Linear Stochastic Models by Parameterizing Expectations: An Application to Asset Pricing with Production. (July 1991)

6. Albert Marcet
Simulation Analysis of Dynamic Stochastic Models: Applications to Theory and Estimation. (November 1991), 2d. version (March 1993)
[Forthcoming in *Advances in Econometrics* invited symposia of the Sixth World Congress of the Econometric Society (Eds. JJ. Laffont i C.A. Sims). Cambridge University Press]

7. Xavier Calsamiglia and Alan Kirman
A Unique Informationally Efficient and Decentralized Mechanism with Fair Outcomes. (November 1991)
[Forthcoming in *Econometrica*]

8. Albert Satorra
The Variance Matrix of Sample Second-order Moments in Multivariate Linear Relations. (January 1992)
[Published in *Statistics & Probability Letters* Vol. 15, no. 1, (1992), pp. 63-69]

9. Teresa Garcia-Milà and Therese J. McGuire
Industrial Mix as a Factor in the Growth and Variability of States'Economies. (January 1992)
[Forthcoming in *Regional Science and Urban Economics*]

10. Walter Garcia-Fontes and Hugo Hopenhayn
Entry Restrictions and the Determination of Quality. (February 1992)

11. Guillem López and Adam Robert Wagstaff
Indicadores de Eficiencia en el Sector Hospitalario. (March 1992)
[Published in *Moneda y Crédito* Vol. 196]

12. Daniel Serra and Charles ReVelle
The PQ-Median Problem: Location and Districting of Hierarchical Facilities. Part I (April 1992)
[Published in *Location Science*, Vol. 1, no. 1 (1993)]

13. Daniel Serra and Charles ReVelle
The PQ-Median Problem: Location and Districting of Hierarchical Facilities. Part II: Heuristic Solution Methods. (April 1992)
[Forthcoming in *Location Science*]

14. Juan Pablo Nicolini
Ruling out Speculative Hyperinflations: a Game Theoretic Approach. (April 1992)

15. Albert Marcet and Thomas J. Sargent
Speed of Convergence of Recursive Least Squares Learning with ARMA Perceptions. (May 1992)
[Forthcoming in *Learning and Rationality in Economics*]

16. Albert Satorra
Multi-Sample Analysis of Moment-Structures: Asymptotic Validity of Inferences Based on Second-Order Moments. (June 1992)
[Forthcoming in *Statistical Modelling and Latent Variables* Elsevier, North Holland. K.Haagen, D.J.Bartholomew and M. Deistler (eds.)]

Special issue Vernon L. Smith
Experimental Methods in Economics. (June 1992)

17. Albert Marcet and David A. Marshall
Convergence of Approximate Model Solutions to Rational Expectation Equilibria Using the Method of Parameterized Expectations.

18. M. Antònia Monés, Rafael Salas and Eva Ventura
Consumption, Real after Tax Interest Rates and Income Innovations. A Panel Data Analysis. (December 1992)

19. Hugo A. Hopenhayn and Ingrid M. Werner
Information, Liquidity and Asset Trading in a Random Matching Game. (February 1993)

20. Daniel Serra
The Coherent Covering Location Problem. (February 1993)

21. Ramon Marimon, Stephen E. Spear and Shyam Sunder
Expectationally-driven Market Volatility: An Experimental Study. (March 1993)
[Forthcoming in *Journal of Economic Theory*]

22. Giorgia Giovannetti, Albert Marcet and Ramon Marimon
Growth, Capital Flows and Enforcement Constaints: The Case of Africa. (March 1993)
[Published in *European Economic Review* 37, pp. 418-425 (1993)]

23. Ramon Marimon
Adaptive Learning, Evolutionary Dynamics and Equilibrium Selection in Games. (March 1993)
[Published in *European Economic Review* 37 (1993)]

24. Ramon Marimon and Ellen McGrattan
On Adaptive Learning in Strategic Games. (March 1993)
[Forthcoming in *A. Kirman and M. Salmon eds.*"Learning and Rationality in Economics" Basil Blackwell]

25. Ramon Marimon and Shyam Sunder
Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence. (March 1993)
[Forthcoming in *Econometrica*]

26. Jaume Garcia and José M. Labeaga
A Cross-Section Model with Zeros: an Application to the Demand for Tobacco. (March 1993)

27. Xavier Freixas
Short Term Credit Versus Account Receivable Financing. (March 1993)

28. Massimo Motta and George Norman
Does Economic Integration cause Foreign Direct Investment? (March 1993)
[Published in *Working Paper University of Edinburgh* 1993:I]

29. Jeffrey Prisbrey
An Experimental Analysis of Two-Person Reciprocity Games. (February 1993)
[Published in *Social Science Working Paper 787* (November 1992)]

30. Hugo A. Hopenhayn and Maria E. Muniagurria
Policy Variability and Economic Growth. (February 1993)

31. Eva Ventura Colera
A Note on Measurement Error and Euler Equations: an Alternative to Log-Linear Approximations. (March 1993)

32. Rafael Crespí i Cladera
Protecciones Anti-Opa y Concentración de la Propiedad: el Poder de Voto. (March 1993)

33. Hugo A. Hopenhayn
The Shakeout. (April 1993)

34. Walter Garcia-Fontes
Price Competition in Segmented Industries. (April 1993)

35. Albert Satorra i Brucart
On the Asymptotic Optimality of Alternative Minimum-Distance Estimators in Linear Latent-Variable Models. (February 1993)

36. Teresa Garcia-Milà, Therese J. McGuire and Robert H. Porter
The Effect of Public Capital in State-Level Production Functions Reconsidered. (February 1993)

37. Ramon Marimon and Shyam Sunder
Expectations and Learning Under Alternative Monetary Regimes: an Experimental Approach. (May 1993)

38. José M. Labeaga and Angel López
Tax Silumlations for Spain with a Flexible Demand System. (May 1993)

39. Daniel Serra and Charles ReVelle
Market Capture by Two Competitors: The Pre-Emptive Location Problem. (May 1993)
[Forthcoming in *Journal of Regional Science*]

40. Xavier Cuadras-Morató
Commodity Money in the Presence of Goods of Heterogenous Quality. (July 1993)
[Forthcoming in *Economic Theory*]

41. M. Antònia Monés and Eva Ventura
Saving Decisions and Fiscal Incentives: A Spanish Panel Based Analysis. (July 1993)

42. Wouter J. den Haan and Albert Marcet
Accuracy in Simulations. (September 1993)
[Forthcoming in *Review of Economic Studies*]

43. Jordi Galí
Local Externalities, Convex Adjustment Costs and Sunspot Equilibria. (September 1993)
[Forthcoming in *Journal of Economic Theory*]

44. Jordi Galí
Monopolistic Competition, Endogenous Markups, and Growth. (September 1993)
[Forthcoming in *European Economic Review*]

45. Jordi Galí
Monopolistic Competition, Business Cycles, and the Composition of Aggregate Demand. (October 1993)
[Forthcoming in *Journal of Economic Theory*]

46. Oriol Amat
The Relationship between Tax Regulations and Financial Accounting: a Comparison of Germany, Spain and the United Kingdom. (November 1993)

47. Diego Rodríguez and Dimitri Vayanos
Decentralization and the Management of Competition. (November 1993)

48. Diego Rodríguez and Thomas M. Stoker
A Regression Test of Semiparametric Index Model Speciication. (November 1993)

49. Oriol Amat and John Blake
Control of the Costs of Quality Management: a Review or Current Practice in Spain. (November 1993)

50. Jeffrey E. Prisbrey
A Bounded Rationality, Evolutionary Model for Behavior in Two Person Reciprocity Games. (November 1993)

51. Lisa Beth Tilis
Economic Applications of Genetic Algorithms as a Markov Process. (November 1993)

52. Ángel López
The Comand for Private Transport in Spain: A Microeconometric Approach. (December 1993)

53. Ángel López
An Assessment of the Encuesta Continua de Presupuestos Familiares (1985-89) as a Source of Information for Applied Reseach. (December 1993)

54. Antonio Cabrales
Stochastic Replicator Dynamics. (December 1993)

55. Antonio Cabrales and Takeo Hoshi
Heterogeneous Beliefs, Wealth Accumulation, and Asset Price Dynamics. (February 1993, Revised: June 1993)

56. Juan Pablo Nicolini
More on the Time Inconsistency of Optimal Monetary Policy. (November 1993)

57. Lisa B. Tilis
Income Distribution and Growth: A Re-examination. (December 1993)

58. José María Marín Vigueras and Shinichi Suda
A Model of Financial Markets with Default and The Role of "Ex-ante" Redundant Assets. (January 1994)

59. Angel de la Fuente and José María Marín Vigueras
Innovation, "Bank" Monitoring and Endogenous Financial Development. (January 1994)

60. Jordi Galí
Expectations-Driven Spatial Fluctuations. (January 1994)

61. Josep M. Argilés
Survey on Commercial and Economic Collaboration Between Companies in the EEC and Former Eastern Bloc Countries. (February 1994)

62. German Rojas
Optimal Taxation in a Stochastic Growth Model with Public Capital: Crowding-in Effects and Stabilization Policy. (September 1993)

63. Irasema Alonso
Patterns of Exchange, Fiat Money, and the Welfare Costs of Inflation. (September 1993)

64. Rohit Rahi
Adverse Selection and Security Design. (February 1994)

65. Jordi Galí and Fabrizio Zilibotti
Endogenous Growth and Poverty Traps in a Cournotian Model. (November 1993)

66. Jordi Galí and Richard Clarida
Sources of Real Exchage Rate Fluctuations: How Important are Nominal Shocks?. (October 1993, Revised: January 1994)
[Forthcoming in *Carnegie-Rochester Conference in Public Policy*]

67. John Ireland
A DPP Evaluation of Efficiency Gains from Channel-Manufacturer Cooperation on Case Counts. (February 1994)

68. John Ireland
How Products' Case Volumes Influence Supermarket Shelf Space Allocations and Profits. (February 1994)

69. Fabrizio Zilibotti
Foreign Investments, Enforcement Constraints and Human Capital Accumulation. (February 1994)

70. Vladimir Marianov and Daniel Serra
Probabilistic Maximal Covering Location Models for Congested Systems. (March 1994)

71. Giorgia Giovannetti.
Import Pricing, Domestic Pricing and Market Structure. (August 1993, Revised: January 1994)

72. Raffaela Giordano.
A Model of Inflation and Reputation with Wage Bargaining. (November 1992, Revised March 1994)

73. Jaume Puig i Junoy.
Aspectos Macroeconómicos del Gasto Sanitario en el Proceso de Convergencia Europea. (Enero 1994)

74. Daniel Serra, Samuel Ratick and Charles ReVelle.
The Maximum Capture Problem with Uncertainty (March 1994)

75. Oriol Amat, John Blake and Jack Dowds.
Issues in the Use of the Cash Flow Statement-Experience in some Other Countries (March 1994)

76. Albert Marcet and David A. Marshall.
Solving Nonlinear Rational Expectations Models by Parameterized Expectations: Convergence to Stationary Solutions (March 1994)