COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 984

DYNAMIC STRUCTURAL MODELS: PROBLEMS AND PROSPECTS. MIXED CONTINUOUS DISCRETE CONTROLS AND MARKET INTERACTIONS.

by

Ariel Pakes

July 1991

Revised May, 1991

Dynamic Structural Models: Problems and Prospects. Mixed Continuous Discrete Controls and Market Interactions.

by Ariel Pakes[1]

This paper reviews dynamic structural econometric models with both continuous and discrete controls, and those with market interactions. Its goal is to highlight techniques which enable researchers to obtain estimates of the parameters of models with these characteristics, and then use the estimates in subsequent descriptive and policy analysis. In an attempt to increase the accessibility of structural modelling, emphasis has been laid on estimation techniques which, though consistent with the underlying structural model, are computationally simple. The extent to which this is possible depends on the characteristics of the applied problem of interest, so the paper ends up coverring more than one topic. To help the reader who has more focussed interests, we now provide an outline of what can be found in the various subsections of the paper.

Section II introduces the examples used to illustrate the points made in the paper. We begin with single agent problems involving continuous, as well as discrete, controls, and later place the agent explicitly into a market setting. The availability of continuous controls raises the possibility of using stochastic Euler equations to estimate some of the

---

[1]This is one part of a two part tour of dynamic structural modelling prepared for the Sixth World Congress of the Econometric Society, Barcelona, 1990. The other part, by John Rust(1991), is on discrete decision processes. Both parts are self contained. I have benefited from discussions with many individuals in the course of writing this paper, among them; Don Andrews, Steve Berry, Gary Chamberlain, Sam Kortum, Jim Heckman, Ken Judd, John Rust, Chris Sims and Stephen Zeldes. Special thanks go to John Rust who read over several previous drafts and made very helpful comments. The paper also borrows liberally from my previous work with Rick Ericson, and with Steve Olley, and therefore, also owes them a special debt. All errors, of course, remain my own responsibility. The research reported here was funded, in part, by the National Science Foundation, through grant number SES-882172.

parameters of the model, and section III begins the substantive discussion of the paper by considering this possibility.

We first show that the fact that there are discrete, as well as continuous, controls does not destroy our ability to generate stochastic Euler equations off of perturbations to the continuous controls, and that, provided the data is handled with care, these equations can be used to generate computationally simple estimators of the form developed by Hansen and Singleton (1982). Next we show that modified Euler equations can also be developed to analyze situations in which there are boundaries on the choice of the continuous control that are binding with positive probability (III.1; we consider both the case of boundaries whose values evolve endogenously, and the case of exogenously set boundaries), and to analyze certain special cases of models with unobserved state variables (III.2). Finally we consider conditions for the use of Euler equations when the impact of the controls on future values of the state variables are stochastic (section III.3), and when the realizations of the state variables determining behavior are not necessarily conditionally independent across agents (conditional on past history; section III.4).

For long enough panels probably the most troubling aspect of the assumptions needed to generate desirable estimators out of Euler equations, is the fact that Euler equation techniques can only accommodate very limited forms of serial correlation, or dependence, in unobserved state variables. Moreover, the problem of obtaining consistent estimators of the parameters of models with serially correlated unobserved state variables persists when we consider estimators based on the complete solution to the control problem, and not limit ourselves to the restrictions embodied in Euler equations. As a result we devote Section IV to the problem of incorporating serially correlated unobserved state variables into the structural models we estimate.

That section begins by noting that the problem of serially correlated unobserved state variables can be reduced to an "initial conditions" problem almost identical to the problem discussed in Heckman's(1981) analysis of estimation in discrete state, discrete time, stochastic processes. Section IV.1 provides some comments on the applicability of the results from previous analysis of the initial conditions problem to estimating structural economic models with serially correlated unobserved state variables.

Section IV.2 provides an alternative method for dealing with serially correlated unobserved state variables that arises naturally in certain economic models with continuous (as well as discrete) controls, and provides proofs of its validity for the two examples used extensively in this paper. Where applicable, the alternative can often be combined with semiparametric estimation techniques to enable one to derive computationally simple estimators for problems that are inherently very complex (such as those that allow for market interactions). We illustrate with an empirical example taken from Olley and Pakes (1990). It derives estimates of a Cobb–Douglas production function in the presence of a serially correlated unobserved productivity shock which generates both a simultaneity and a selection problem (the first because of the endogeneity of input choices, and the second because of the fact that firms which draw better productivity sequences are more likely to survive).

Section V of the paper begins by making the agent's payoff in a given period depend on the state variables of other agents in that period, thereby formally incorporating market interactions into the problem. Once we do this we are faced with the issue of specifying the nature of the equilibria established among the various (potential and actual) actors. We limit ourselves to a discussion of the estimation and computational issues that arise in models with Markov–Perfect Nash equilibria (see Maskin and Tirole, 1987, 1988a, 1988b).

Our attitude toward the empirical analysis of dynamic models with market interactions is to separate the problem of estimation from the problem of computing the equilibrium implications of the parameters estimated. The estimation problem is broken into smaller parts, each of which is both consistent with the overall dynamic equilibrium framework, and provides an estimator (with desirable properties) for a subvector of the model's parameters. We assume estimation will have to proceed in this fashion because, for most problems of current interest, neither our computational resources, nor our data, are rich enough to allow us to estimate all the model's parameters in a single unified iterative estimation algorithm. This procedure also turns out to be helpful in providing a framework which enables us to separate out and empirically analyze single primitives from richer, and hopefully more realistic, economic environments.

Once we have estimated the primitives of the problem, we will want to compute and analyze the distribution of equilibrium, or market, responses to policy and environmental changes. Section V.1 turns to this computational problem. It begins with an algorithm for computing Markov Perfect Nash equilibria. We use this algorithm to compute a differentiated products version of the Ericson–Pakes model of industry dynamics (Ericson and Pakes,1989), and then use the output of this computation to illustrate the many aspects of reality that can be captured by the current generation of structural models.

The example also has the useful property that the Markov process that defines its equilibrium lives on a finite set of points. So we can calculate the value functions and policies it generates to any desired degree of precision and then compare the true values to the values obtained from alternative approximation methods. We show that for exact calculations the number of grid points that need to be evaluated at every iteration of the recursive fixed point calculation grows polynomially in the (least upper bound to the) number of agents (ever) active in the market. The time per grid point evaluation grows as a polynomial of lower order. So we turn to an examination of the possibility of using polynomial expansions to approximate the value function at each iteration of the computational algorithm; an idea which, in somewhat different form, has been used extensively in a variety of recent research (see Judd,1990, the chapter by Marcet in this volume, and the literature cited in those articles).

The major analytic result in this section shows that provided the value function of a given agent is symmetric (more precisely exchangeable) in the state vectors of its competitors, the number of polynomial coefficients one needs to determine for a given order of approximation is independent of the number of agents active in the market. This implies that the number of grid points which we need to use at each iteration of the fixed point calculation will also become independent of the number of agents active in the market – a result which may enable us to devise relatively straightforward algorithms for computing equilibria for large markets. The computational part of the paper concludes by fitting the actual value functions for our example to the exchangeable basis of polynomials, and then examining the quality of the fit from the polynomial approximation.

There is a also short concluding section to the paper. It provides a more personal view on the use of structural economic modelling, and its role in helping to interpret data.

Notation, and the Role and Choice of Examples.

The following notation will be augmented at various points in the paper.

s is the state variable, assumed to be an element of some metric space S. In this chapter $s=(y,\hat{s})$, where y is understood to be the vector of state variables whose values are agent specific, and $\hat{s}$ consists of the y vectors of all other agents operating in the market.

$C_s = Y_s \times \Gamma_s$: where $Y(\cdot)$ specifies a finite (but strictly positive) number of feasible alternatives for the discrete control (χ), and $\Gamma(\cdot)$ specifies a compact valued continuous correspondence which provides the feasible values for the continuous control (x).

$d = (\chi,x)$: where d is the decision vector, χ is the discrete control, and x is the continuous control.

In all our examples the one period return function will be written as $\pi(s,d)-c(s,d)$, where $\pi(\cdot)$ is the single period "profit" function and $c(\cdot)$ provides the cost of the chosen policy [note also that the feasible levels of d depend on s through Γ_s].

There are, of course, many possible state variables whose values do not differ across agents (examples include prices, technology, and regulatory rules). It will be understood here (for notational convenience) that these individual invariant variables are also included in y. Under standard regularity conditions the agent's optimal policy solves the Bellman equation

$$V(s) = \sup_{\{d\in C_s\}} \{\pi(s,d) - c(s,d) + \beta\int V(s')P(ds'\,|\,s,d)\}, \quad (1).$$

with V(s) given by the unique solution to the implied contraction mapping. We let $\{d_s, s\in S\}$ be the associated stationary optimal policy.[1]

[1]Uniqueness of the policy function for the continuous control is often more difficult to obtain; see Benveniste Scheinkman,1979, Blume Easely and O'Hare,1982, the discussion in Stockey, Lucas, and Prescott,1989,and for the case where the impact of the continuous control on the family of measures $\{P(\cdot\,|\,\cdot,x)\}$ is sufficiently smooth in x, Ericson and Pakes, 1989. At the very least nonuniqueness generates nonuniqueness in behavioral responses to policy and/or environmental changes, and this becomes a problem for policy

At this level of generality the notation hides distinctions which become important in both choosing specifications which are appropriate for different applied problems, and for determining the availability and properties of alternative estimators. Rather than cataloging special cases in an abstract way, we carry particular examples along in the various subsections of the paper. The examples have been chosen for their ability to allow us to illustrate the issues we thought were important in as simple a setting as possible. Using examples in this way has the additional advantage that it allows us to comment on some of the more detailed specification issues that arise in choosing appropriate assumptions for certain classes of applied problems.

The first example we deal with is a production–investment model, similar to those which have been used extensively in both the macro (see Stockey, Lucas, and Prescott,1989) and in the industrial organization (see Tirole,1989) literatures. Section II, which allows for both continuous and discrete controls but not market interactions, considers a monopolist who accumulates physical capital according to a deterministic law of motion but faces a stochastic environment. The monopolist makes two decisions in each period; whether or not to exit, and if not, how much to invest in capital accumulation. The second example is also a production investment model; but this time one that allows for stochastic accumulation. This model is also somewhat more detailed, and we use it more intensively in the later sections of the paper where we revoke the monopoly assumption and consider estimation and computation in models that allow for market interactions.

II.An Introductory Example

We begin with what is probably the simplest model with both continuous and discrete controls that one would attempt to take to panel data. It has one continuous control (investment), a choice between two discrete alternatives (remaining active or exiting), an

---

analysis. It may also generate an additional set of estimation problems (see Jovanovic,1989, for a discussion of the related problem of estimation in models with multiple equilibria). Note, however, that once the value function in (1) is computed for a given set of primitives one can simply inspect the solution for uniqueness of the policy.

exogenous state variable which evolves stochastically, and no market interactions. For simplicity we will take the exogenous stochastic state variable to be unobserved, but, in general it could be a vector process with an unobserved component. The unobservable is needed to rationalize the heterogeneity in both the outcome paths and the investment choices observed in the data. Also, in a more general framework we would want to allow for separate disturbances affecting the value of all but one of the discrete alternatives (this to rationalize the discrete choices in the data). Provided the discrete state specific disturbances are included in an additive fashion and are serially independent, as in the discussion of the chapter by Rust(1991), they have no substantive effects on the points to be made using this simple example, and therefore, have been omitted from the discussion [2]

In terms of our previous notation, we make the following assumption.

Assumption 2

$$y = (k,\omega) \in K \times \Omega \subset \mathbb{R}^2_+,$$

where it is understood that we only observe y for firms that are active at the beginning of the period (χ=1), while if $d=(\chi_t,x_t) = (1,x)$,

$$k_{t+1} = k_t\,(1-\delta) + x,$$

with probability one, and the distribution of ω_{t+1} conditional on y_t is determined by the family

---

[2]To add them back in simply assume that ϕ, the exit value in the discussion that follows, is random. Provided the distribution of the discrete state specific disturbance is sufficiently rich, the saturation condition discussed in Rust(1991) will amount to the condition that the observed combinations of the continuous control and observable component of the state vector can be generated by the primitives of the model and the alternative possible values of the unobservable state. A partial discussion of this issue can be found in the related literature on continuous choice using extremal processes; see Cosslett,1988, Dagsvick, 1988, Resnick and Roy,1989.

of distributions,

$$\mathbb{P}_{\omega} = \{P(\cdot\,|\,\omega), \omega\epsilon\Omega\},$$

which are assumed to have densities w.r.t some dominating measure, to be stochastically increasing in ω, and to possess the property that if $h(\cdot)$ is continuous and bounded then

$$\int h(\omega')\,(d\omega'\,|\,\omega)$$

is a continuous function of ω.

We also assume that $\pi(y;\theta)$ is bounded, increasing in both its arguments, differentiable (with bounded derivative) and concave in k, and has $\lim_{k\to 0} \frac{\partial\Pi}{\partial k}(\cdot)=\infty$ while $\lim_{k\to\infty} \frac{\partial\Pi}{\partial k}(\cdot)=0$, for each $\omega\epsilon\Omega$; that $c(x,s;\theta) = c(x;\theta)$ which is increasing, differentiable (with bounded derivative), and convex in x and that both $\partial c()/\partial x$, and $\partial\pi()/\partial k$, are differentiable in θ (a.s.). •

The assumption that the cost of adjustment depends only on the amount of investment (and not on the capital in place) is made solely for expositional convenience. Remark 1 following Theorem 27 generalizes the results in this section to the case where the cost of adjustment depends also on k.

There are, however, at least two aspects of these assumptions that are more problematic. First (2) assumes that the accumulation relationship between the continuous control and the state variable is deterministic. Though this has become a traditional assumption in the literature on the accumulation of physical capital, it is a special case of a more general model in which the impact of investment is stochastic. One might argue the relevance of the special deterministic case for investment in physical capital, but it seems much less appropriate for the accumulation of the "intangible" capital stocks that emanate from a firm's investment in research and exploration, or ir advertising and goodwill. Here the randomness in the outcome from the investment activities often seem to both have strikingly large variances, and to underlie many of the "simultaneity" and "selection" issues that generated the interest in structural modeling of the phenomena of interest in the first place. [3] Similar distinctions occur among the different types of stocks accumulated by households (compare, for example, investments in health, to investments in consumer durables).

There are several differences between models with deterministic and stochastic accumulation which are important for the discussion which follows. First, when a state variable evolves deterministically knowledge of past investments implies knowledge of the current stock (at least up to an initial condition and the parameters describing the decay process). So deterministically controlled state variables are generally assumed to be observed by the econometrician. In contrast, unless there is a separate reading on the outcome of (in contrast to the input into) the investment process, a stock that accumulates stochastically will be unobserved, and will therefore generate a serially correlated unobserved state variable (of course, there may well be serially correlated unobserved state variables in models with deterministic accumulation also; section IV discusses estimation in the presence of serially correlated unobserved state variables). Also, though one can derive "Euler equations" for some models with stochastic accumulation, both their form, and the assumptions needed to justify them, differ from those needed for models with deterministic accumulation (see section III.3). On the plus side, models which allow for stochastic accumulation, but presume smoothness in the relationship between the continuous choice and the transition probabilities, generate first order conditions with relative ease. This, in turn, both enables more detailed analytic treatment of optimal policies, and simplifies computational issues (section V). Our second example is a model with stochastic accumulation, and it will be used to illustrate these points.

[3]For an early model with stochastic accumulation see Roberts and Weizman, 1981. Tirole, 1989 chapter,10, and Ericson and Pakes, 1989, discuss some of the more recent contributions to the literature.

The assumption of a convex and differentiable cost of investment function can also be problematic. Nonconvexities can often be handled by adding additional dimensions to the set of discrete alternatives (see Das, forthcoming). In models of capital accumulation one often worries about the differentiability of c(x) at the point x=0, as this is the point at which small movements carry with them the difference between selling, and purchasing and installing, units of the stock. A count of the number of observations at which x is exactly zero in the data ought to provide some indication of whether this is likely to be an important problem in any given application (and it often is).

With assumption 2 the Bellman equation for our problem (equation 1) reduces to

$$V(\omega,k) = \max\ \{\Phi, \sup_{[x\epsilon\Gamma(\omega,k)]}[\pi(\omega,k) - c(x) + \beta\int V[k(1-\delta) + x, \omega')(d\omega' \mid \omega)], \qquad (1'),$$

where Φ is the return to closing down the firm (the return to $\chi=0$) and transferring its salvageable assets to another activity[4] and, here, and below, it is understood that all functions are indexed by θ_0.

We pause here to fill in two details. First, to complete the specification of the model we need to specify $\Gamma(y)$, the choice set for x. This is a primitive of the problem and we will want to consider different assumptions on it and investigate their implications below. For starters assume there are no restrictions on $\Gamma(y)$, or that $\Gamma(y) = \mathbb{R}$. Now note that boundedness of $\pi(\cdot)$ together with the fact that $\beta < 1$ implies boundedness of the expected discounted value of any feasible program, and, therefore, that there exists a (finite) $\bar{x}$ such that $\sup_y x(y) < \bar{x}$. Similarly the fact that $\lim_{k\to 0} \partial\pi()/\partial k = \infty$, implies that it will never be optimal for an active firm to derive its capital stock to zero, so for all k $\inf_\omega x(\omega,k) > -k$. So $\omega.\ell.$o.q we take, $\Gamma(y) = [-k,\bar{x}]$, and note that

[4]One could make this return depend on y, but then, to preserve the form of the optimal stopping policy below, we would need to insure that $\Phi(y)$ does not increase at as rapid a rate in y as the return from staying in operation does.

$$x(\omega,k) \in \mathring{\Gamma}(y), \qquad (3a),$$

for every $(\omega,k)\epsilon K x \Omega$ [here $\mathring{\Gamma}(y)$ is notation for the interior of the set $\Gamma(\cdot)$]. That is, investment will never be at a "corner" of its choice set (we come back to the problems generated by corners below).

Second, note that policies for this problem are couples, (χ,x), where $\chi: K x \Omega \to [0,1]$, provides the exit decision, and $x: K x \Omega \to \mathbb{R}$ determines investment. Given out assumptions it is straightforward to show that $V(\cdot)$ is nondecreasing in both its arguments, and that, as a consequence, there is a function $\underline{\omega}(\cdot)$, which is decreasing in k, such that $V(\omega,k) \leq \Phi$ if and only if $\omega \leq \underline{\omega}(k)$. So the exit policy is a stopping rule, $\underline{\omega}(\cdot): K \to \Omega$, with the property that

$$\chi_{t+1} = 0 \text{ iff } \omega \leq \underline{\omega}(k), \quad k \epsilon K, \qquad (3b).$$

III. Euler Equations From Mixed Continuous–Discrete Choice Models

Having added a continuous alternative to the discrete choice problem, it is natural to begin with the question of whether we can go back to the computationally simple estimation techniques based on stochastic Euler equations to estimate at least some of the parameters of this mixed continuous–discrete choice model. To this end we compare the discounted cash flow earned from the optimal policy to that from the starred alternatives to that policy given by

$$\chi^*(t+\tau)=\chi(y_{t+\tau}), \text{ for all } \tau,$$

(4)

$$x^*(t,\epsilon)=x(y_t)-\epsilon, \quad x^*(t+1,\epsilon)=x(y_{t+1})+(1-\delta)\epsilon,$$

$$x^*(t+\tau+1,\epsilon)= x(y_{t+\tau+1}), \text{ for } \tau\geq 1,$$

and alternative values of ϵ.

An important property of the alternative programs in (4) is that they all hold the discrete choice the same as what the discrete choice would have been in the optimal program (no matter the realized state of the world), and they only perturb the continuous control only in periods t and t+1. As a result all programs are constructed so that $y^*_{t+\tau} = y_{t+\tau}$ for all $\tau \geq 2$ with probability one, and the difference between the cash flows generated by the ϵ–alternative and the optimal policy is only a function of the costs of adjustment and profits in periods t and t+1. Now note that boundedness of the return function and the fact that $\beta < 1$ insure that the difference in cash flows is well defined, while the feasibility of the ϵ–alternative program for sufficiently small $|\epsilon|$ [see (3a)] together with the optimality of the original program, insure that this difference must be nonpositive in a neighborhood of $\epsilon=0$. Thus, provided this difference is differentiable, its derivative must be 0 at $\theta=\theta_0$. Differentiability follows from the assumptions that c() and π() have bounded derivatives, the fact that χ() is differentiable almost everywhere (see 3b), and the Lebesgue Dominated Convergence Theorem. Taking that derivative, and evaluating it at $\epsilon=0$, we get the Euler equation in Lemma 5.

5.Lemma.

A necessary condition for a policy couple $\{\chi(y),x(y)\}$ to be optimal is that

$$-\partial c(x)/\partial x+\beta\int \chi[\omega',k(1-\delta)+x]\{\partial\pi[\omega',k(1-\delta)+x]/\partial k+(1-\delta)\partial c[x(\omega',k(1-\delta)+x)]/\partial x]P(d\omega'|\omega)=0,$$

at $\theta=\theta_0$.

Lemma 5 makes it clear that the presence of discrete, as well as continuous, controls does not destroy our ability to generate stochastic Euler equations. We simply substitute the observed value of the discrete control into the return functions and then proceed in precisely the same way we did in the continuous problem. The Euler equation formed in this way can then be used as a basis for forming moment conditions which can be fed into a method of moments estimation algorithm of the form developed by Hansen and Singleton (1982; for further notes on the estimation algorithm see the discussion at the end of section III).. Note, however, that if we were to use an equation such as (5) as a basis for estimation we would have to select the observations used to form the sample moments on variables which are measurable date t. If an agent satisfies the selection criteria in year t and transfers discrete states in year t+1, the agent must be kept in sample for the purpose of the period t Euler equation, even if the relevant date t+1 variables are not "reported". In our example then, when a firm actually does exit in period t+1, we use the period t observation by substituting the actual x_t in for the first term for that observation in (5), and setting the second term equal to zero, as would be the case if $\chi_{t+1}=0$.

There are, of course, several problems that actually can destroy our ability to use stochastic Euler equations, at least those as simple as the one in lemma 5, as a basis for estimating parameters of dynamic models from panel data, but they have little to do with the addition of discrete alternatives to the choice set. We now move on to a brief review of some of them. Section III.1 considers the possibility of binding constraints on the choice set, section III.2 considers unobserved state variables, III.3 considers stochastic (in contrast to deterministic) accumulation, and III.4 considers cases in which the realizations of the state variables are not conditionally independent across agents in the panel. Since the extent to which any of these problems are likely to be important will vary with the characteristics of the economic model and of the data one is analyzing, we have tried to insure that each subsection is self–contained.[5]

---

[5]Throughout we derive our Euler equations by constructing a set of alternative feasible policies and checking for differentiability of the implied perturbations to the value function . This makes the problems that arise in constructing Euler equations in applications with a single continuous control transparent. An alternative would be to nest a system of random lagrange multipliers into the control problem, and derive the Euler equations from their properties; see Kushner, 1965a, and 1965b. This latter technique is more detailed notationally, but would have advantages in applications with a system of continuous controls in which case we might want to use the relationships between the various constraint sets to help structure estimation.

III.1. Binding Constraints on the Choice Set.

We begin with the possibility that the optimal choice is not in the interior of the feasible set, i.e. of constraints in the correspondence defining the feasible choices for the continuous control that are binding for a set of values for the state variables that have positive probability. In this case condition (3a) is not always satisfied and, as a result, it is not always possible to construct the alternative program in (4) for all $|\epsilon|$ less than some positive κ.

There are at least two types of characterizations of the economic environment that have lead to binding boundaries in empirical work. In the first the location of the relevant boundary is both, known to the econometrician, and independent of the actions of the agent. This would have occurred in our example if we had not assumed a market for used capital goods, as this would change the choice set for x from $[-k,\bar{x}]$ to $[0,\bar{x}]$, and we could not rule out the possibility that the optimal investment choice is 0 for a set of $y=(k,\omega)$ values with positive probability. A similar problem occurs when one of the primitive functions is nondifferentiable at zero (or at any other prespecified point), and the modification of the Euler equation in (5) developed below will cover this case also. Both these cases generate zero investment with positive probability, and when $x=0$ the x choice need not satisfy a first order condition (like the Euler equation) with equality.

In the second type of problem with binding boundaries the location of the boundary is endogenous, being determined, in part, by the actions of the agent in previous periods. This latter case has been studied extensively in both the consumption and the investment literatures under the heading of liquidity and/or financial constraints (see Hubbard and Kashyap,1989,Gilchrist,1989, and Himmelburg,1989, and the literature cited there for the investment problem, and Hayashi,1987, Keane, 1983, Zeldes,1989, and the literature cited in these articles for the consumption problem).[6] This problem would have occurred in our example

[6]This heading does not adequately describe the richness of the issues at hand. These are not so much a result of any notion of the illiquidity of assets, as they are a result of the incompleteness of markets for future income streams. Moreover, different formulations for market opportunities lead to different budget constraints, and the precise formulation of the budget constraint will generally effect the properties of alternative estimators; see, for example, the discussion in section 5 of Hayashi, 1987, and the literature referred to below.

were we to have introduced another state variable for the firm, its financial assets in period t,or A(t), and assumed that the firm cannot borrow more than A(t). This would restrict the choice set for x to equal [−k,A], and we could not insure that x<A for all possible y.

We begin with binding boundaries of the first kind using our example with the additional restriction that $\Gamma(y)=[0,\bar{x}]$. Note first that the conditions we required for the proof of Lemma 5 can now be violated in one of two different ways. First it is possible that $x(y_t)=0$. Second, even if $x(y_t)>0$, it is possible that there is a set of y_{t+1} that has positive probability conditional on y_t, for which $x(y_{t+1})=0$. In either one of these cases the alternative program in (4) will not be feasible for all values of $|\epsilon| \leq \kappa$ (and some $\kappa > 0$), and, as a result, the logical basis for constructing the Euler equation in (5) breaks down. Note,however, that provided x(y) is observable the first requirement, i.e. that $x(y_t)>0$, does not, in itself, destroy our ability to obtain parameter estimates from an Euler equation for a selected sample. That is, since, x is measurable date t, we can select only those observations with $x(y_t)>0$, consider the sample analog to the restriction in (5) for this sub sample, and base estimation on the fact that the expected value of the Euler equation for the sub sample will equal zero at the true θ_0 . So provided x is observable, the difficulty in deriving Euler equations that rely on compensating perturbations in adjacent periods, like those in (4) above, when there is a possibility of binding corners, is that we do not directly observe $x(y_{t+1})$ for all values of y_{t+1} that have positive probability conditional on period t information.

One way around this problem is to relax the requirement that we base our Euler equations on perturbations in adjacent periods.[7] As we now show this will allow us to push the compensating ϵ−perturbation forward to some (random) future period in which the agent is not at a corner, thereby making it feasible for all values of $|\epsilon|\leq\kappa$ with probability one. With this in

[7]Another possibility is to impose restrictions which make it possible to develop a semiparametric estimator which uses the information in the data to select out a subsample for which both $x(y_t) > 0$ and $x(y_{t+1}) > 0$ with probability one, and then generate an Euler equation from this sub sample that allows one to obtain consistent parameter estimates.

mind define

$$\tau^* = \min_{\tau\geq 1}\{\chi(y_{t+\tau})=0 \text{ or } x(y_{t+\tau})\geq\kappa\}, \qquad (6a).$$

τ^* is a positive, integer–valued, random variable determined either by the first time the agent's program calls for positive investment, or the period in which the agent exits, whichever comes first [formally if $F_{t+i}=\sigma(y_t,\ldots,y_{t+i})$, the σ–algebra generated by $(y_t,\ldots,y_{t+i})$, then τ^* is a stopping time with respect to the sequence $\{F_{t+i}\}$]. Now select out a sub sample with $x(y_t)\geq\kappa$, and for that sub sample consider the alternative programs

$$\chi^*(y_{t+\tau},\epsilon) = \chi(y_{t+\tau}), \text{ for all } \tau \text{ and } \epsilon,$$

and

$$\text{(6b)} \qquad x^*(y_{t+\tau},\epsilon) = \begin{cases} x(y_{t+\tau}) - \epsilon, & \text{for } \tau = 0, \\ x(y_{t+\tau}), & \text{for } \tau < \tau^*, \\ x(y_{t+\tau^*}) + (1-\delta)^{\tau^*-1}\epsilon, & \text{for } \tau = \tau^*, \\ x(y_{t+\tau}), & \text{for } \tau > \tau^*. \end{cases}$$

For fixed τ^* the ϵ–alternative policy in (6b) is feasible for all $|\epsilon|<\kappa$. Moreover any such policy generates the same values for the state variables at all $\tau>\tau^*$ as does the optimal policy (with probability one). Conditional on a value for τ^* then, the difference between the expected discounted values of net cash flows generated by the optimal and these ϵ–alternative policies just depends on cash flows between times t and $t+\tau^*$. τ^* is a random variable, but one whose realization is <u>independent</u> of ϵ.

Now construct the difference in discounted net cash flows between the optimal and the ϵ–alternative policies given in (6b), and note that an argument analogous to that proceeding lemma 5 leads to the conclusion in lemma 7.

7. Lemma

When investment is restricted to be nonegative, but the rest of the assumptions underlying Lemma 5 remain intact, then a necessary condition for a policy couple $\{\chi(y),x(y)\}$ to be optimal is that

$$-\partial c(x_t)/\partial x + E_{(t)}\Big\{\Sigma_{\tau=1}^{\tau^*-1}\beta^{\tau}[\partial\pi(\omega_{t+\tau},k_{t+\tau})/\partial k](1-\delta)^{\tau-1} + \beta^{\tau^*}\chi_{t+\tau^*}\{[\partial\pi(\omega_{t+\tau^*},k_{t+\tau^*})/\partial k](1-\delta)^{\tau^*-1} + \beta[\partial c(x_{t+\tau^*})/\partial x](1-\delta)^{\tau^*}\Big\} = 0,$$

where $E_{(t)}f()$ provides the expectation of f() conditional on F_t, and τ^* is defined as in (6).

Provided that τ^* is observable (in our example this requires observations on both investment and on whether or not the firm liquidated in a given year), and that τ^* is always less than T−t, where T is the final year of the panel, (7) can be used to generate Euler equation based estimators in the same way Lemma 5 does. Note however that the requirement that $\tau^*\leq T-t$ is a probability one requirement. If we simply select out those observations for which it ends up being true, we will be selecting the sample on the basis of behavior determined by information not available at date t, and any selection procedure based on such information will generate an inconsistency in the estimation procedure. The importance of this censoring problem is likely to vary from sample to sample, but one ought to be able to get an indication of just how important it is in any given sample by examining the empirical distribution of the realizations of τ^*, and adjusting the estimation algorithm accordingly (that is using only those years for which there is a sufficiently long subsequent period observed in the panel).

As an empirical example of the magnitude of the issues raised by the existence of discrete controls and of binding corners, we consider Olley's(1990) study of the telecommunications equipment industry (most of SIC 3661, and part of SIC 3663). The study constructs a thirteen year panel of enterprise level data (the basic data sources are the various censuses and surveys of manufacturing as combined in the LRD, see McGuckin,1989, for details on the LRD). The telecommunications equipment industry is an industry with large plants, but one which went through a major restructuring during the period covered by the data (this was caused by both a gradual deregulation process, and by technological change). Of the 376 plants that are in the data at some time during the sample period, 93, or approximately 25%, shut down by the end of the sample. So liquidation does occur quite frequently and should be accounted for in the analysis. Of the 2569 plant/year observations available (this includes the observations on plants who liquidated in the following year) 168, or about 6.5%, report exactly zero investment. Thus there is some evidence of there being a "boundary" at zero. These 168 zero observations, however, include only 16 sequences with two consecutive years of zero investment, and there is not one sequence with three consecutive zero investment years. Thus it seems that for this data set Euler equation techniques would be appropriate provided we used data on all plant–year observations at t (including those who liquidate in t+1) for all $t \leq T-2$,or possibly, $t \leq T-3$.

We now move on to consider boundaries whose values are determined, at least in part, by the actions of the agent in previous periods. There are really two differences from the last case. The first is due to the fact that once we have partial control over the value of the boundary in the coming period we can often, though not always, insure that we will be away from that boundary with probability one by appropriate choice of the control, and this simplifies the estimation problem considerably. The second difference is that models which have endogenous boundaries generally involve an extra (often quite difficult) set of measurement problems that arise from the need for a measure of that boundary. We come back to this measurement problem below.

We should note at the outset that not all models with endogenous boundaries will be able to generate Euler equations from simple compensating variations in adjacent periods. There is, therefore, room for developing necessary and sufficient conditions for Euler equations when boundaries are endogenous. Here we suffice with the simpler task of illustrating the logic of the argument that allows one to use Euler equations when they are available. To do so in as simple a fashion as possible we have to modify our example to allow for consumption, as well as production (this because the simple model with an added finance constraint does not admit Euler equations based on compensating perturbations in adjacent periods).[8]

Now the agent is allocating consumption and investment expenditures to maximize the expected discounted value of a time separable utility function. The boundary condition is obtained by constraining the agent to satisfy the "credit constraint" that the sum of investment and consumption expenditures cannot exceed the value of the agent's financial assets. These are denoted by A(t), so that if the two continuous controls (investment and consumption) are given by $x_{1,t}$ and $x_{2,t}$, the credit constraint is written as

$$A_t \geq x_{1,t} + x_{2,t}, \qquad (8a).$$

A(t) (which is the boundary of the correspondence determining the feasible choices for the continuous control) evolves according to

$$A_{t+1} = \begin{array}{ll} A_t(1+r) + \pi(\omega_t, k_t) - x_{1,t} - x_{2,t}, & \text{if } \chi_t=1, \\ A_t(1+r) + (\chi_{t-1} - \chi_t)\Phi - x_{2,t}, & \text{if } \chi_t=0, \end{array} \qquad (8b)$$

where for simplicity we have assumed a deterministic rate of return on financial assets (r). Note that the effect of a change in the discrete state here is to change the nature of the accumulation relationship.[9] If the agent shuts down the firm, he (or she) obtains a one period payoff of Φ, and

---

[8]If we did not allow for consumption expenditures we would have two state variables, A_t and k_t whose laws of motions are different linear functions of the same, single, control, x_t. Thus, it would in general be impossible to construct compensating perturbations for the control that would return <u>both</u> state variables to what their values would have been in the optimal program after two periods.

[9]This is analogous to the situation that would arise if we were to apply the Euler equation methodology to analyzing retirement decisions, or to choices among entering various

then must live off asset income. We shall assume that the primitives of this model have been chosen so that the optimal program is kept away from any lower bound on the x–choices.

With these assumptions consider the following family of alternative policies. Leave the discrete shutdown decision the same as in the optimal program in every period, and the continuous controls unchanged for all periods after t+1, but change the continuous controls in periods t and t+1 to

(9)
$$x_{1,t} = x^*_{1,t} - \epsilon_1 \text{, and } x_{2,t} = x^*_{2,t} - \epsilon_2$$
$$x_{1,t+1} = x^*_{1,t+1}(y_{t+1}) + (1-\delta)\epsilon_1 \text{, and } x_{2,t+1} = x^*_{2,t+1}(y_{t+1}) + g(y_{t+1}, \underline{\epsilon}),$$

where

$$g(y_{t+1}, \underline{\epsilon}) = \Delta\pi[y_{t+1}, \epsilon_1] - (1+r)\epsilon_2 - (r+\delta)\epsilon_1,$$

and $\Delta\pi[\cdot]$ is the difference in profits resulting from the difference in period t+1 capital stocks induced by ϵ_1, and $\underline{\epsilon} = (\epsilon_1, \epsilon_2)$. Without going into details we simply note that it is easy to show both that; the alternative program in (9) have been constructed so that they will be feasible with probability one for all sufficiently small $|\epsilon_1| + |\epsilon_2|$ provided <u>only</u> that the credit constraint is not binding in period t, and that the alternative and optimal programs generate the same value for the state variables after period t+2 (with probability one).

Given these facts, a simple extension to the logic of lemma 5 for the sub sample which is not at the boundary in the current period leads us to two Euler equations. These are obtained from the derivatives of the function defining the difference between the expected discounted value of utility from the alternative and from the optimal program with respect to ϵ_1 and ϵ_2 (evaluated at $\underline{\epsilon}=0$). The first Euler equation is familiar from the finance literature (see

---

welfare programs, see the references in Rust(1991).

Lucas, 1978). It states that the expected discounted value of the marginal utility weighted returns from the two forms of savings (investing in physical, and in financial, capital) must be equated at the margin. The second Euler equation is familiar from the literature on the life cycle hypothesis (see Hall, 1979). It states that if consumption expenditures are interior, the marginal utility of consumption should be a martingale (with respect to $\{F_t\}$).

The fact that this latter Euler equation is not destroyed by credit constraints, provided only that the credit constraint is not binding in the initial period, was exploited to do estimation and testing of a stochastic consumption model with credit constraints by Zeldes (1989) (see also the literature he cites; similar reasoning has been used to derive and analyze Euler equations for the investment decisions of firms in situations in which firms face credit constraints, see Gilchrist,1989, Himmelburg,1989, and Hubbard and Kashyap, 1989). Our discussion only generalizes to the extent that it allows for discrete controls (a fact which might allow one to integrate the closely related phenomena of shutdown and/or bankruptcy into the analysis).

In terms of our notation, those articles assume that all expenditures and assets are perfectly observed, and then proceed as follows. Calculate the right hand side of (8a). If it is greater than A(t) for a given agent, that agent could not have faced a binding finance constraint in the given period. Now separate the sample and use only the unconstrained agents in the estimation algorithm. Again, provided the selection is on variables which are measurable date t, the selected sample should abide by the constraints generated from the Euler equation at θ_0. Moreover a comparison of diagnostic tests done on the selected and the not selected parts of the sample can be illuminating (see Zeldes,1989).

Note that these procedures for generating Euler equations when there are boundaries which evolve endogenously assume that we can select out a sub sample that we <u>know</u> are not at a boundary in the current period. Empirically the question of whether this is possible depends on our ability to measure the variables determining the boundary condition. The reason we bring out measurement problems here, after ignoring them for the most part of this subsection, is that the variables determining the boundary conditions for the applications which have used these techniques to date are among the poorest measured of economic variables. For example, to determine whether the agent is at a boundary in the credit constraint example we need information on both total expenditures and total assets, two variables which are notoriously hard to measure. Most of the micro empirical literature to date has treated this measurement problem in an informal manner, using more or less stringent selection criteria, and focusing on the parameter estimates obtained from the more stringent criteria. Little has been done on more formal treatments of this problem (though a preliminary discussion of it is contained in Hajivassiliou and Ioannides,1989).

Once we allow for the case in which we do not know with certainty whether a given observation was at a boundary, the conditional expectation of the Euler equation is no longer necessarily zero at θ_0. Instead it becomes a sum of its expectation conditional on the constraint not being binding (which equals zero at θ_0) times the probability of the constraint not being binding, plus the expectation of the Euler equation conditional on the constraint being binding, times the probability of this latter event. A question that then arises is whether we can use the structure implied by the measurement problem, together with the constraints implied by the model at θ_0, to restrict the moments obtained from this "generalized" Euler equation at $\theta=\theta_0$ in a way that allows us to narrow the admissible range for (and, if possible, estimate) that parameter. We do know that at θ_0 the expectation of the Euler equation conditional on being at the boundary must be nonegative everywhere (this follows from the fact that a negative perturbation to the current choice is always feasible), and, we will often be able to show that both it, and the probability of being at a corner, is nonincreasing in the observed asset measure (as a case in point,in the credit constraint example the expectation of the Euler equation becomes the "lagrange multiplier" for the credit constraint). This gives us another set of constraints to use in obtaining information on θ. Just how far theory, together with some combination of parametric and nonparametric estimation procedures, can get us in this context, is an open question, but one that might well be worth pursuing, at least for problems to be analyzed on large data sets (see Matzkin,1990, for both an analysis of how qualitative characteristics derived from theory can be used to empirically analyze some static choice problems, and for a discussion of related computational issues).

There are, of course, more complicated constraints on the choices of continuous controls possible than those considered here. Our examples, however, do serve to illustrate the following points. First, if we have the ability to observe whether an agent is currently at a boundary of the feasible correspondence, the difficulty in establishing the conditions needed to generate stochastic Euler equations is in constructing a perturbation to the optimal policy which will be feasible in the following periods with probability one for all ϵ in an open neighborhood of zero. When the boundary is endogenous, this problem can often be circumvented by appropriate choice of controls in the alternative programs (this presumes sufficient markets to trade over time or between continuous controls). However in the case of endogenous boundaries measurement problems often make it difficult to determine whether an agent is at a boundary in the current period, and little work has been done to date on incorporating measurement error into the analysis. When the boundaries are set exogenously, modified Euler equation techniques can still be constructed, but we will generally require data that allows us to follow each agent over more than two periods of time to implement them (though the data itself ought to be able to tell us approximately how many consecutive periods are required). Finally, note that the presence (or absence) of discrete controls has little to do with our ability to generate Euler equations from perturbations to the continuous control (no matter whether there is the possibility of binding boundaries).

III.2 Unobserved State Variables

To date our ability to estimate the parameters of structural models with unobserved state variables, whether using Euler equations based estimators, or some other form of estimation technique, is severely limited. As a result we will pay special attention to it in this review. Section IV discusses the problems associated with integrating serially correlated unobserved state variables in estimation algorithms based on actually solving for the value function at different values of the parameter vector. This subsection explores the extent to which unobserved state variables can be incorporated into Euler equation based estimation algorithms.

We begin by setting out the problem generated by unobserved state variables, and then move on to discuss two special cases for which we can circumvent that problem. These cases are neither nested to, nor more general than, either the conditional independence assumption discussed extensively in Rust(1991), or the assumptions we develop in section IV for estimation in the presence of serially correlated unobserved state variables. More generally, the assumptions that are relevant for the problem at hand determine our ability to derive consistent estimators from alternative estimation techniques, and no single technique nests the others.

Our ability to obtain Euler equation based estimators when there are unobserved state variables depend on particular functional form assumptions. In order to be clear about precisely what assumptions are needed, we will have to move out of the confines of our example, and into a more general setting. Letting the state vector for a given individual, $y=(z,\nu)$, with z observed, and ν not observed, and recalling that $d(\cdot)$ is the vector of decision, or control, variables we write the Euler equation as

$$E_{(t)}h[d(z_{t+1},\nu_{t+1}),z_{t+1},\nu_{t+1},d(z_t,\nu_t),z_t,\nu_t;\theta_0]=0, \qquad (10)$$

where the expectation is conditional on all information available to the agent in period t (F_t). The estimation problem is that, since we do not observe ν, we cannot calculate the sample analogs to the moment conditions that can be generated from (10).

The first special case we consider mixes a partial conditional independence assumption with an exclusion restriction. Formally what we require is that we can partition the vector of unobservables, ν, into two components, ν_1 and ν_2, and then write the function $h(\cdot)$ in (10) as

$$h[d(z_{t+1},\nu_{2,t+1}),z_{t+1},\nu_{1,t+1},d(z_t,\nu_{2,t}),z_t,\nu_{1,t};\theta_0], \qquad (11a),$$

with ν_1 and ν_2 subvectors of ν, and

$$P(d\nu_1', d\nu_2', dz' \mid \nu_1, \nu_2, z) = P(d\nu_1' \mid z')P(d\nu_2', dz' \mid \nu_2, z), \qquad (11b),$$

with $P(d\nu_1' \mid z')$ a known family of distributions.

Equation (11a) assumes that one of the components of the partitioned vector of state variables (i.e.,ν_2) only impacts on the value of the Euler equation through its effects on the controls [i.e. through $d_t(\cdot)$], while the other (ν_1) can have an independent effect on the Euler equation but does not effect the control at all. (11b) states that ν_1, the unobserved component which has an independent effect on the value of the Euler equation, must be from a known family of conditionally independent distributions. However the conditional distribution of ν_2, the unobserved component that only affects the Euler equation through its impacts on the control, need not be either known, or restricted (ν_2 can, for example, be freely serially correlated).

To illustrate the content of the restrictions in (11) we go back to our example (assumption 2 and lemma 5), and augment it with unobserved state variables. Assume then that there is both an unobserved component in the ω process determining productivity (our ν_1), and randomness in the process determining the liquidation value of the firm (this value was previously the constant Φ). Call the unobserved component of the liquidation value ν_2, and assume ν_1 and ν_2 obey (11b). That is ν_1 is generated by a known family of conditionally independent distributions, but ν_2 may be freely serially correlated. z_t in this example contains the capital stock and any observed components of the ω process (for simplicity we ignore these in what follows).

There are two controls in this problem; the investment (x), and the shutdown (χ), decisions. For (11a) to be satisfied neither can depend on ν_1. If ν_1 is conditionally independent, investment cannot depend on it (investment only depends on the conditional distribution of future ν_1 values and that distribution is independent of ν_1). $\chi(\cdot)$ will not depend on ν_1 provided either that the exit decision must be made before the realization of ν_1, or that the exiting firm exits at the end of the period and obtains the profits from that period. Of the assumptions needed for (11a), then, the one that seems to be more problematic for our example is that the unobserved component of the productivity process is conditionally independent. Whether it is a reasonable assumption for the problem at hand depends on the nature of the factors which cause shifts in productivity in the industry being studied, and on the variables measured in the data being used.

The Euler equation for this example (the extension to Lemma 5) is

$$-\partial c(x)/\partial x + \beta\int_{\nu_2'}\int_{\nu_1'}\chi[\nu_2',k(1-\delta)+x;\theta]\{\partial\pi[k(1-\delta)+x,\nu_1']/\partial x +$$
$$(1-\delta)\partial c[x(\nu_2',k(1-\delta)+x)]/\partial x\}P(d\nu_1'\,|\,k(1-\delta)+x)P(d\nu_2'\,|\,k,\nu_2)=0,$$

at $\theta=\theta_0$. This is a special case of (11a) (one in which the initial value of ν_1 does not effect the Euler equation). We now note how combining this equation with a simulation technique (McFadden,1989, and Pakes and Pollard,1989) allows us to obtain consistent estimates of θ_0 (as will become clear, numerical integration would do equally well). Draw ν_1^* from $P[d\nu_1\,|\,k(1-\delta)+x]$, then construct

$$-\partial c(x_t)/\partial x + \beta\chi_{t+1}[\partial\pi(k_{t+1},\nu_1^*)/\partial k +(1-\delta)\partial c(x_{t+1})/\partial x],$$

and form the product of this function with alternative functions of observables known at date t. Now note that, since the expectation of the product function is 0 at $\theta=\theta_0$, sample moments obtained in this fashion can be used as a basis for a GMM estimation algorithm a la Hansen and Singleton(1982; for more detail see the discussion at the end of section III). Note, however, that since (11a) was not an assumption directly on the primitives of the model, we needed to verify it for our example before we could proceed with this estimation technique (there are alterative sets of assumptions on primitives that would lead one to 11a, and it was not clear that a listing of them would be any more useful than just stating 11a directly).

The more general case is only slightly more complicated. Note that (10) and (11) imply that

$$E_{(t)}\int h(d_{t+1},z_{t+1},\nu',d_t,z_t,\nu_{1,t})P(d\nu'|z_{t+1})=0$$

for almost every (d_t,ν_t,z_t). Consequently the integral of this expression with respect to $P(d\nu_{1,t}|z_t)$ must equal zero. Let $g(\cdot)$ be any square integrable function of (d_t,z_t). Then, from Fubini's theorem,

$$E\{[\int\int h(d_{t+1},z_{t+1},\nu'',d_t,z_t,\nu')g(d_t,z_t)P(d\nu''|z_{t+1})P(d\nu'|z_t)]|z_t,d_t\}=0.$$

Now simply substitute the observed values of the vector $(d_{t+1},z_{t+1},d_t,z_t)$ into h(), and simulate both ν' and ν'', for each individual. The sample average of these equations should converge to zero at $\theta=\theta_0$, and therefore can also be used as a basis for a method of moments estimation algorithm.

There is at least one more special case in which unobserved state variables do not hamper our ability to obtain consistent parameter estimates from Euler equations. This case also partitions the unobserved state vector into two components. The first only affects the Euler equation through its effect on the control (ν_2), while the second (ν_1) can also have a direct effect on the Euler equation. However the logarithm of the Euler equation must be additively separable into a function of ν_1 and a function which is independent of ν_1, and ν_1 must be constant overtime. More formally the requirements of this special case are that

$$\nu_{1,t+1} = \nu_{1,t} = \nu_1, \qquad \text{(12a), and}$$

$$h(\ ) = h_1(\nu_1;\theta_0)h_2(d_{t+1},z_{t+1},d_t,z_t;\theta_0), \qquad \text{(12b),}$$

with $h_1(\nu_1;\theta_0)\neq 0$ (a.s.). (12a) assumes that the value of ν_1 is constant over (at least one) two–period interval; an assumption which has been used extensively in panel data estimation problems (see Mundlak,1963, for an early discussion of its relevance, and the more recent discussion in Chamberlain, 1984, and the literature cited therein). (12b) is a strong assumption on the form of the Euler equation. It has been used in investigations of Euler equations for optimal intertemporal consumption choices, as in these cases differences in instantaneous utility functions across agents that appear multiplicatively in the marginal utility of consumption will, when combined with (12a), generate Euler equations of the form in (12b) (Zeldes,1989, uses one variant of this assumption).

Given (12) estimation and testing is straightforward. Together, (10) and (12) insure that the expectation of $h_2()$ conditional on all observed variables must be zero. Since $h_2()$ does not contain any unobservables, we can simply treat it as the Euler equation (ignoring ν_1) and base estimation on the restrictions implied by the fact that its conditional expectation is zero at θ_0.

We do not want to conclude this subsection with the wrong impression. All of the assumptions used to date to account for unobservables in structural estimation problems are quite restrictive. At this point all we can say is that (more or less simple) estimation algorithms have been developed under alternative sets of assumptions, and one or more of them may be appropriate to the problem at hand. The alternatives do have the saving grace that they usually generate overidentifying restrictions which can be used to formulate test statistics. The tests, however, are usually omnibus tests that do not distinguish very well between alternative possible sources of error, and the models we are dealing with are quite complicated. Good model selection criteria are likely to be more dependent on a detailed knowledge of the problem one is dealing with, and of the data that is available. The researcher will simply have to be familiar enough with his or her problem to have reasonably strong priors about what the major unobservable sources of differences in behavior across agents is, on how they relate to the alternative estimation procedures available, and, if there is a worry left over, on which diagnostic tests are likely to pick up problems in the more questionable assumptions.

III.3. Stochastic Accumulation.

We now sketch the basics of a model that allows for stochastic accumulation – a model that we will elaborate on when we introduce market interactions later in the paper.[10] Our purpose in this subsection is simply to illustrate the logic that allows one to generate Euler equations from models with stochastic accumulation when it is possible to do so. The deeper question of the conditions that allow one to generate Euler equations in models with stochastic accumulation is left for future research.

For simplicity assume that there is only one state variable, ω, or the efficiency of the firm, and that profits are increasing in it. The firm invests in research and exploration to improve its efficiency in future years but the outcomes of the research process are uncertain. That is the distribution of ω_{t+1} conditional on information at time t depends upon ω_t and x_t, so that $\{\omega_t\}$ is a controlled Markov process. Its primitives are given by the family of conditional distributions,

$$\mathbb{P}= \{ P(\cdot\,|x,\omega),\ (x,\omega) \in \mathbb{R}_+ \times \Omega \subset \mathbb{R}^2\}.$$

The family $\mathbb{P}$ is assumed to be stochastically increasing in x for each value of ω (increases in investment lead to better, in the stochastic dominance sense, distributions for future efficiency), stochastically increasing in ω for each given x (conditional on x the higher the current ω the better the distribution of tomorrow's ω), and continuous in the sense that when integrated against a continuous bounded function of ω', it produces a continuous bounded function of both x and ω.

The rest of the structure of this example is taken from the model with deterministic accumulation. In each period an active firm makes two decisions. One discrete decision (whether to remain active or to liquidate and receive the sell off value of Φ), and, if the discrete decision is positive ($\chi=1$), one continuous decision (the quantity to invest in research or x). The Bellman

[10]This model is taken from Ericson and Pakes, 1989, Part I.

equation which defines the value function for this problem is then

$$V(\omega)=\max\{\Phi,\sup_{x\in\mathbb{R}_+}[\pi(\omega)-c(x)+\beta\int V(\omega')P(d\omega'|\omega,x)]\}, \qquad (13).$$

Let $\{\chi(\omega),x(\omega)\}$ be the optimal policy. Then substituting that in (13) and rewriting in terms of the expected value of profits in the following period and the value of continuing thereafter we have,

$$V(\omega)=\max\{\Phi,\ \pi(\omega)-c[x(\omega)] + \beta\int\{\chi(\omega')[\pi(\omega')-c(x(\omega'))]\}P[d\omega'|\omega,x(\omega)]$$

$$+\beta\Phi\int[1-\chi(\omega')]P[d\omega'|\omega,x(\omega)] + \beta^2\int\chi(\omega')V(\omega'')P[d\omega''|\omega',x(\omega')]P[d\omega'|\omega,x(\omega)].$$

We want to find a set of alternative programs that leave the last term in this expression unchanged (then the difference between the value of the alternative and optimal programs just depends on actions and outcomes in periods t and t+1). The continuity properties of the family $\mathbb{P}$ insure that the following set of policies would suffice (were they feasible). Leave the shutdown policy the same as the optimal shutdown policy, and the investment policy the same as the optimal investment policy for all $t+\tau$ with $\tau\geq2$. However subtract ϵ from $x(\omega_t)$ and add $\Delta(\epsilon,\omega_{t+1})$ to $x(\omega_{t+1})$, where ϵ and $\Delta(\cdot)$ are chosen such that $\Delta(\epsilon,\cdot)=0$ at $\epsilon=0$, and the <u>distribution</u> of ω_{t+2} conditional on ω_t and each alternative policy, is the same as the distribution of ω_{t+2} conditional on ω_t and the optimal policy. More formally choose ϵ and $\Delta(\cdot)$ such that for every $\tilde{\omega}\in\Omega$

$$(14) \qquad \int_{\omega'}P[\omega''>\tilde{\omega}|\omega',x(\omega')]P[d\omega'|\omega,x(\omega)]$$

$$=\int_{\omega'}P[\omega''>\tilde{\omega}|\omega',x(\omega')+\Delta(\epsilon,\omega')]P[d\omega'|\omega,x(\omega)-\epsilon].$$

The optimal policy produces a distribution of ω_{t+2} conditional on ω_t as a convolution of $P[\cdot\,|\,\omega_t,x(\omega_t)]$ and $P[\cdot\,|\,\omega_{t+1},x(\omega_{t+1})]$. Equation (14) states that we can obtain the same convoluted distribution by perturbing x_t by $-\epsilon$ and x_{t+1} by $\Delta(\epsilon,\omega_{t+1})$. When this is so the difference in the expected discounted value of cash flows generated by the alternative and optimal programs just depends on the cash flows in periods t and t+1.

More formally, if the alternative in (14) is feasible and we let $V(\epsilon,\omega)$ be the value generated by the alternative program, then

$$V(\omega)-V(\epsilon,\omega)=-c[x(\omega)]+\beta\int\{\pi(\omega')-c[x(\omega')]\}\chi(\omega')P[d\omega'\,|\,\omega,x(\omega)]$$

(15)

$$+c[x(\omega)-\epsilon]-\beta\int\{\pi(\omega')-c[x(\omega')+\Delta(\epsilon,\omega')]\}\chi(\omega')P[d\omega'\,|\,\omega,x(\omega)-\epsilon].$$

By optimality (15) is nonegative for every feasible value of ϵ and equals zero at $\epsilon=0$. Thus provided we can show that there is a feasible ϵ–alternative policy that abides by (14) for all $|\epsilon|\leq\kappa$ (for some $\kappa>0$), and that (15) is differentiable in ϵ in the appropriate region, that derivative must be zero at $\epsilon=0$ — giving us an "Euler equation" for the problem with stochastic accumulation.

Whether or not these conditions are satisfied, and the form of the derivative if they are, depend on the properties of the family $\mathbb{P}$ [as well as on the differentiability of $c(\cdot)$ and the continuity of $\pi(\cdot)$]. Though it is beyond the scope of this paper to do a detailed investigation of the appropriate necessary and sufficient conditions, a simple example where the required conditions are satisfied will help illustrate how to proceed.

Let the family $\mathbb{P}$ have densities (with respect to some dominating measure) and assume those densities can be written as

$$p(\omega_{t+1}\,|\,\omega_t,x_t;\theta)=p_\xi[\omega_{t+1}-\mu(\omega_t,x_t;\theta)], \quad (16),$$

for every $(\omega_{t+1},\omega_t,x_t)\epsilon\Omega^2 X\mathbb{R}_+$, as would be the case if differences in ω and in x only caused a change in the mean of the distribution of future efficiencies [condition (16) is also satisfied in the computed version of the Ericson–Pakes model we discuss below; though for different reasons]. Assume also that; $p_\xi(\cdot)$ is differentiable in its argument for every possible value of ξ, that $\mu(\cdot)$ is everywhere differentiable in both its arguments, and that $\partial\mu(\cdot)/\partial x$ is both positive everywhere and goes to ∞ as x approaches zero (for all ω). This last condition insures that x is kept away from its lower bound of zero. The others insure (via the implicit function theorem) that there exists a differentiable function $\Delta(\epsilon,\omega')$ that satisfies (14).

Now $\omega_{t+1}=\xi_{t+1}+\mu(\omega_t,x_t)$, so for the family of alternative programs to satisfy (14) we need

$$\mu[\xi_{t+1}+\mu(\omega_t,x_t),x_{t+1}] = \mu[\xi_{t+1}+\mu(\omega_t,x_t-\epsilon),x_{t+1}+\Delta(\epsilon,\xi_{t+1})].$$

This, in turn implies that

$$\partial\Delta(\epsilon=0,\omega_{t+1})/\partial\epsilon = \{[\partial\mu(\omega_{t+1},x_{t+1})/\partial\omega][\partial\mu(\omega_t,x_t)/\partial x]\}/\{\partial\mu(\omega_{t+1},x_{t+1})\},$$

Now go back to (15), substitute in (16), and note that our assumptions together with the Lebesgue Dominated Convergence Theorem imply that the result is a differentiable function of ϵ. Taking that derivative and setting it equal to zero we obtain,

$$-\partial c[x_t]/\partial x$$
$$-\beta[\partial\mu(\omega_t,x_t)/\partial x]E_{(t)}\chi_{t+1}\{[\partial\pi(\omega_{t+1})/\partial\omega]+[\partial c(x_{t+1})/\partial x][\partial\mu(\omega_{t+1},x_{t+1})/\partial\omega]/[\partial\mu(\omega_{t+1},x_{t+1})/\partial x]$$
$$\} = 0.$$

This is not much more complicated than the Euler equation in Lemma 5 (the equation for the model with deterministic accumulation). However, we should emphasize that the procedure outlined here did rely heavily on the index restriction in (16).

III.4. Conditional Distributions Which Are Not Independent Across Agents.

Thus far we have assumed that the conditional distribution of the vector of the period t+1 state variables of the different agents (conditional on information known at date t) factors into the product of the distributions that those agents actually use in forming their own expectations. In addition to rational expectations, this requires independence of the conditional distributions of the $\{y_{t+1,i}\}_{i=1}^{N}$. It is this independence which insures that the average of the Euler equation disturbances, averaged over the individual's in the sample, converges to zero at the true θ_0, a property which lies at the heart of the proof of the convergence, as N grows large, of the Euler equation based estimator to the true θ_0.

Recall that y_{t+1} must include all variables which have either an independent effect on the value of the Euler equation, or are determinants of the value of control variables which, in turn, are arguments in the Euler equation. Frequently this leaves a lot of room for state variables whose realizations will either be common, or highly correlated, across agents (typical examples include prices, levels of technology, and governmental policy variables). At this point we should note that the fact that there is dependence in the conditional distribution of the $y_{t+1,i}$ across i does not necessarily rule out our consistency condition. That is, the same random variable may affect different agents in different ways, so that the dependence it induces in the realization of the Euler equation errors may not be strong enough to invalidate the convergence of the sample average of the true Euler equation disturbances to zero. On the other hand, when there are state variables that have important impacts on behavior that are likely to induce dependence across agents, then the arguments we have been implicitly relying on for the consistency of the Euler equation based estimators are at best incomplete (and may be seriously misleading).

There are at least two ways to investigate the possibility that dependence in the realizations of the state variables generates significant biases in Euler equation based estimators.

The first is empirical, using a combination of formal tests and less formal descriptive statistics to analyze the possible impact of dependence for the problem at hand. There are several computationally straightforward ways of proceeding here, and we come back to a more detailed discussion of them below. The second possibility is to provide theoretical conditions under which any dependence would impact on the Euler equation in a particular (and analyzable) way, and then check for, and possibly estimate, subject to them.

The latter possibility has recently been used in the literature on consumption choices, where the theory of complete markets has been used to structure the relationship between the increment in the marginal utility of consumption across households (see Altug and Miller,1990; and Altonji, Hayashi, and Kotlikoff, 1990). To date the empirical specifications used in this literature have focused on the opposite extreme to the one that is implicitly employed in the estimation procedures that do not allow for dependence – with complete markets the only source of uncertainty is one whose realizations are common to all agents. The truth, no doubt, lies somewhere in the middle (and much more difficult) case with only partial markets for future income streams. Note,however,that if any such markets exist they will, in and of themselves, induce conditional dependence in the realizations of the state variables of agents operating in the same submarket (for some eye–opening empirical evidence on such dependence, see Townsend,1990). In applications involving choices by firms, the implications of the dependence induced by the aggregate factors we often worry about (demand, factor price, technological, and policy changes) will depend on the nature of the dynamic equilibria established among the various actors (see section IV.3).

We now move on to consider what can be done to salvage Euler equation based estimation techniques that have desirable limit properties in dimension N when there is dependence across agents. Unfortunately, there has been little progress here. Empirical papers often attempt to account for the problem of dependence by adding time–specific dummy variables, say α_t, to the Euler equation, assuming that $E_{(t)}[h(\cdot;\theta)-\alpha_t]=0$, and then minimizing (by choice of both the vectors α and θ), a quadratic form in the sample analogs of the population restrictions implied by this equation. One way to see the implications of this procedure is to partition the vector of state variables, y_t, into a subvector whose conditional distribution is independent across agents, say z_t, and one that is not, say l_t, and then consider conditions which would imply the consistency (as N grows large) of the estimator of θ we obtain when we use the time–specific dummy variables. Using the more general notation introduced in section III.2 where $h(\cdot)$ is the Euler equation and $d(\cdot)$ is the vector of controls, one set of such conditions is

$$h(d_{t+1},y_{t+1},d_t,y_t) = h_1(d_{t+1},z_{t+1},d_t,y_t) + h_2(l_{t+1}), \quad (17a)$$

$$d(y_{t+1}) = d(z_{t+1}), \quad (17b)$$

and

$$P(dz_{t+1},dl_{t+1}|z_t,l_t) = P(dz_{t+1}|z_t)P(dl_{t+1}|l_t), \quad (17c).$$

(17a) states that the impact of l on the Euler equation is additively separable, while (17b) requires that l not effect the control at all. There are few, if any, empirical examples where the additive separability in (17a) arises naturally from the underlying primitives of the model, and, as we now show, it can be relaxed at the cost of a small change in the specification of the estimating equation. Replace assumption (17a) with

$$h(d_{t+1},y_{t+1},d_t,y_t) = h_1(d_{t+1},z_{t+1},d_t,z_t)h_2(l_{t+1}) + h_3(d_{t+1},z_{t+1},d_t,y_t) + h_4(l_{t+1}), \quad (17a').$$

(17a'), by itself, is a condition which, though clearly restrictive, arises quite frequently in applied work. It would, for example, apply to our leading investment example (assumption 2) if we were worried about common price (or demand) uncertainty, and those prices (or demand factors) had a multiplicatively separable effect on the current profit function (it also arises in the consumption example discussed by Hall,1979, once one allows for interest rate uncertainty). Now note that (17a'),(17b),(17c), and (10) imply that

$$h_{2,t+1}E_{(t)}h_1[d(z'),z',d,l,z] + E_{(t)}h_3[d(z'),z',d,z,l] + h_{4,t+1} = 0, \quad (18)$$

where $h_{2,t+1} = \int h_2(l_{t+1})P(dl_{t+1} \mid l_t)$, and $h_{4,t+1}$ is defined accordingly. It follows that the sequence of couples $\{[h_{2,t+1}, h_{4,t+1})]\}_{t=1}^{T}$ can be treated as parameters to be estimated along with θ, in an Euler equation based estimation procedure that will, given our assumptions and some standard regularity conditions, yield a consistent and asymptotically normal estimator as N, the number of agents, grows large (with T held fixed).

Note, however, that these results rely not only on (17a'), but also on (17b). The latter assumption requires that the controls in t+1 not be a function of the factor which induces the dependence in the state variables across agents (l_{t+1}). For most problems of interest, including our investment problem with common price uncertainty, this would be unlikely unless the common price process were serially independent (in which case its current realization would not impact on the perceived distribution of its future values, and therefore would not impact on investment and exit decisions). Serial independence of the process leading to the dependence in the realizations of the state variables of the various agents is often an unattractive assumption. On the other hand the assumptions underlying estimators based on equations such as (18) are testable, just as those based on an assumption of a lack of conditional dependence in the process generating the state variables are, and (18) is clearly less restrictive . Before moving away from the discussion of providing Euler equation based estimators with desirable asymptotic properties as N grows large holding T fixed when there is conditional dependence, it should be noted that neither we, nor other published work we are aware of, have attempted a very detailed investigation of the possibilities here, so that this is an area in which further research may well be warranted.[11]

---

[11]We have not attempted, for example, to use any additional restrictions that might result from the sampling process, for example the possible exchangeability of the vector of observations on different individuals; for a review of the implications of exchangeability see Aldous(1983). Similarly we have not attempted to make use of the fact that the factors generating the dependence across observations are often observed, which would allow us to compare different years with similar realizations in that factor.

We now move to a brief comparison of Euler equation based estimators that rely on limits in dimension T, holding N fixed, to those that rely on limits in dimension N, holding T fixed. We do this even though the vast majority of panel data problems have N much larger than T, for two reasons.[12] First, a comparison of the two limits will lead to simple sets of tests for conditional dependence in the state variable. Second, whether or not T is large for a given application is not only a function of the length of the panel, but also of the variance that the common factor induces in the sample average of the true Euler equation disturbances in the different years of the panel. If the average (over time) of these cross sectional average disturbances has a "small enough" variance, then the asymptotic approximations that rely on T growing large will be accurate (and recall that if the theory is correct the average cross–sectional Euler equation disturbances in the different years of the panel should be mutually uncorrelated so its variance should go down at rate T^{-1}). Moreover, if the approximations that rely on T growing large are accurate, <u>and</u> we choose an <u>appropriate</u> Euler equation based estimator (see below), we can often obtain estimators with desirable properties even in the presence of dependence.[13]

Once we allow for the possibility of conditionally dependent state variables, we have to be more careful about distinguishing differences in the properties of Euler equation based estimators obtained from different restrictions. To this end we introduce some additional notation. Let, $h[d(y_{i,t+1}),y_{i,t+1},d(y_{i,t}),y_{i,t};\theta] = h_{i,t}(\theta)$, $x_{i,t} \in \mathcal{F}_t$, and define

---

[12]The exceptions here are usually data sets that follow industries or countries over time.

[13]To illustrate I asked Stephen Zeldes to supply the values (and standard errors) of the coefficients of the time dummy variables he estimated in his analysis of the consumption Euler equation (1989, see the discussion in III.2). The estimates of the time dummies for the nine year panel on his preferred equation ranged from –.05 (.12) to + .03 (.06). It is reasonably clear from his estimates that one could accept the null that they are all zero, but this seems to be as much a function of the fact that the averages are estimated imprecisely as of any inherent smallness in the point estimates. On the other hand, the grand average of the Euler equation errors in Zeldes' study was –.01 with a standard error which was also,.01, and these are numbers that one might be willing to accept as close enough to zero with high enough probability.

$$N^{-1}\Sigma_{(i)}h_{i,t}(\theta)x_{i,t} = g^{x}_{*,t}(\theta), \qquad (19a),$$

$$T^{-1}\Sigma_{(t)}h_{i,t}(\theta)x_{i,t} = g^{x}_{i,*}(\theta), \qquad (19b),$$

and

$$T^{-1}N^{-1}\Sigma_{(t)}\Sigma_{(i)}h_{i,t}(\theta)x_{i,t} = g^{x}_{*,*}(\theta), \qquad (19c).$$

We consider estimators of θ obtained from minimizing a quadratic form in restrictions formed by; i)averaging over i for fixed t (as in 19a), ii) by averaging over t for fixed i (as in 19b), and iii) by averaging over i _and_ t (as in 19c).

Given appropriate regularity conditions, the first will yield consistent and asymptotically normal estimators as N grows large regardless of T provided that the evolution of the state variables are conditionally independent across agents. Note, however, that the first order conditions which define the estimator of θ in this case have terms which converge to the expectation of

$$g^{x}_{*,t}(\theta)\ \partial g^{x}_{*,t}(\theta)/\partial\theta,$$

and, if there is dependence in the conditional distributions of the state variables of the various agents, then the covariance between the cross–sectional average of the Euler equation disturbance, and its derivative with respect to θ, will not generally be zero at the true θ_0 . As a result, if there is dependence in the realizations of the state variables across agents, the estimator based on minimizing a quadratic form in restrictions of the form in (19a) will _not only_ be inconsistent when N grows large holding T fixed, but will _also be inconsistent as T grows large_.[14]

[14]There are, of course, special cases for which the required covariance is zero. This occurs when the derivative of g(·) with respect to θ depends only on variables that are measurable F_t. A case in point is the literature which tests for the rationality of price forecasts, see Keane and Runkle, 1990, and the literature cited there. The linear framework typcially used in this literature generates estimators which are unbiased regardless of the presence of dependence in the process generating the forecast error. As stressed by Keane and Runkle,

On the other hand, given appropriate regularity conditions, estimators based on minimizing a quadratic form in the time averaged Euler equation restrictions of the different individuals (as in 19b) will result in consistent and asymptotically normal estimators of θ as T grows large, for fixed N, even if the conditional distributions of the state variables of the various agents are dependent.[15] However, an analogous argument to the one used to show that conditional dependence destroys the consistency of Euler equation based estimators obtained from averaging over i for fixed t, shows that if T is not large enough the estimator obtained from restrictions such as those in (19b) will be inconsistent regardless of whether the observations are conditionally independent.

Finally the estimators obtained from averaging the restrictions over both i and t are consistent and asymptotically normal if either, N is sufficiently large and the state variables are conditional independent, or if there is conditional dependence but T is large enough.

One way of deciding between the various possibilities is to obtain estimators from restrictions that are averaged over both i and t (as in 19c), and then use a combination of formal testing and descriptive statistics to decide on whether any of the more restrictive alternatives seem relevant. An intuitive starting point would be to do a decomposition of the variance in the value of the restrictions (evaluated at the estimator obtained by averaging over both i and t) between time, individual, and idiosyncratic components. If the time component is small, we might not worry about conditional independence, and be willing to use both the efficiency gains

---

however, the standard errors of the coefficients obtained from the O.L.S. regression still need to be adjusted for the presence of dependence, and this adjustment can have very large impacts on the relevant test statistics.

[15]The required regularity conditions here are generally both more delicate, and harder to verify. We need the dependence in the $y_{i,t+1}$ to induce a dependence in the $h_{i,t+1}()$ that is weak enough to justify the use of a uniform law of large numbers in the consistency proof, and a stochastic equicontinuity condition in the proof of asymptotic normality. "Strong mixing" conditions will often suffice (see Billingsley, 1964), but these are not always satisfied for the problems of interest. For recent contributions to the literature on conditions that generate uniform laws of large numbers, and central limit theorems, in the presence of dependence, see Andrews (1990, and forthcoming) and the literature cited in those articles.

available from restrictions of the form in (19b) and the asymptotics that rely on N growing large. If the individual component is small we might be willing to use limits in dimension T and the efficiency gains available from restrictions of the form in (19a).

More formally let $\hat{\theta}$ be the estimator of θ obtained by minimizing a quadratic form in restrictions that are averaged over both i and t (as in 19c). Then under the null that the process generating the state variables of the alternative agents are conditionally independent, and some mild regularity conditions on the form of the Euler equation, we have

$$N^{-1/2}h_{*,t}(\hat{\theta}) = N^{-1/2}h_{*,t}(\theta_0) + o_p(1), \quad (20a)$$

while

$$N^{-1/2}h_{*,t}(\theta_0) \underset{d}{\to} \mathcal{N}(0,\text{diag}[\sigma_t^2]), \quad (20b)$$

where

$$\hat{\sigma}_t^2 \equiv N^{-1}\Sigma_{(i)}h_{i,t}^2 = \sigma_t^2 + o_p(1), \quad (20c)$$

and, consequently

$$\Sigma_{(t)}h_{*,t}^2/\hat{\sigma}_t^2 N \underset{d}{\to} \chi_T^2 \quad (20d),$$

where; $o_p(1)$ is notation for terms which converge in probability to zero (as N grows large), diag[x] denotes a diagonal matrix with x on the principle diagonal, χ_k^2 is a chi–square deviate with k degrees of freedom, and $\underset{d}{\to}$ denotes convergence in distribution (again as N grows large). (20d) provides our formal test statistic. Note that it only requires 20a, 20b, and 20c, so that the null is really broader than conditional independence and allows for forms of dependence that do not invalidate the consistency and asymptotic normality of the parameter estimates (see the discussion above) [all (20d) tests for is the assumption that $h_{*,t}(\theta_0) = 0$, the assumption that underlies the consistency of estimators whose desirable asymptotic properties are based on limits in dimension N]. Of course one can base similar tests on other moments that emanate from the Euler equation, but an intuitive place to start seems to be with the Euler equation itself.

It is almost as easy to build tests for the assumption that the time average of the Euler equation disturbances are small enough to be accounted for by the limiting approximations in dimension T, that is for the assumptions that underlie estimators that rely on restrictions such as those in (19b). For this we require regularity conditions which insure that

$$T^{-1/2}h_{i,*}(\hat{\theta}) = T^{-1/2}h_{i,*}(\theta_0) + o_p^T(1), \qquad (21a),$$

while

$$T^{-1/2}h_{i,*}(\theta_0) \xrightarrow[d]{T} N(0,V), \qquad (21b)$$

where $V = [v_{i,j}]$, and

$$\hat{v}_{i,j} \equiv T^{-1}\Sigma_{(t)}h_{i,t}h_{j,t} = v_{i,j} + o_p^T(1), \qquad (21c)$$

for each (i,j) so that

$$T^{-1}h_{*,t}\hat{V}^{-1}h_{*,t} \xrightarrow[d]{T} \chi^2_N, \qquad (21d)$$

where $\xrightarrow[d]{T}$ denotes convergence in distribution, and $o_p^T()$ denotes convergence in probability, both as T grows large. (21d) is fairly easy to calculate and provides a direct test of the assumption underlying the consistency of estimators that are based on restrictions of the form given in (19b).

One advantage of the set of tests in (20d) and (21d) is that it is possible for them to indicate that neither of the limiting approximations are relevant for the data at hand. In this case one would also be suspicious of the properties of the estimators obtained by averaging over both i and t. Note also that a similar set of test statistics could be built from comparisons of estimators of parameter vectors obtained by employing restrictions based on averaging over both i and t, to those based on averaging just over i, and then again to those based on averaging just over t (see Hansen,1982, or Chamberlain,1984, for details on the construction of the appropriate test statistics). Neither of these testing sequences are computationally burdensome, and some form of test for dependence should probably be applied as a matter of course in most applications of Euler equations based estimators that use panel data.

That concludes our discussion of the use of Euler equations in structural estimation. It is incomplete in several ways. In addition to leaving several open questions on the topics we did discuss, we left several important topics totally untouched. Perhaps foremost among the latter is the issue of the choice of estimator given only the restrictions that are embodied in the Euler equations (and, perhaps, some regularity conditions).

There is a large related econometric literature on the efficiency of estimators based on moment conditions that can guide us here. In their initial article Hansen and Singleton (1982) use the Euler equations to generate a finite number of moment restrictions, and then consider estimators based on minimizing a quadratic form in those restrictions. A result in Hansen(1982) shows that, given appropriate regularity conditions, the optimal weighting matrix is the inverse of the variance covariance matrix of the moment restrictions themselves (evaluated at the true value of the parameter vector). Chamberlain (1987) shows that, again subject to regularity conditions, the same result applies if we do not limit ourselves to estimators based on quadratic forms in the (finite number of) moment restrictions. Chamberlain then goes on to provide an efficiency bound (for regular estimators) when the restrictions we have to work with are specified directly as conditional moment restrictions (this generates an infinite number of moment restrictions, one for each possible value of the conditioning set; see also Hansen,1985, and Hansen, Heaton, and Ogaki,1988, for related work in the time series literature). Recall that the Euler equation restrictions are in fact conditional moments restrictions (conditional on the σ–algebra generated by variables known in period t); so, provided we are only using the Euler equations, we would ultimately like to obtain estimators which acheive the efficiency bound from the conditional moment restrictions they generate.[16]

If there are K parameters to be estimated, there will (again subject to regularity

[16]There remains the question of whether the conditional expectation of the Euler equation depends on the entire past history of all variables in the data set, or on just a subset of them. If one were willing to specify the entire structure of the underlying control problem, then the model itself would answeer this question. Alternatively, one could try to determine the relevance of different variables empirically by using an innitial consistent estimate of the paramter vector to construct estimates of the realized value of the Euler equation for the alternative sample points, and then examining its conditional expectation.

conditions) be a set of K "instruments" (measurable functions of the conditioning set) which generate moments (orthogonality conditions) whose sum of squares will be minimized at a value of the parameter vector whose limit distribution will achieve the efficiency bound. However, these "efficient" instruments involve the derivative of the conditional moments with respect to the parameter vector (they equal the derivative of the vector of conditional moments times the inverse of the conditional variance of those moment conditions; all evaluated at the true value of the parameter vector). In the Euler equation (and most other nonlinear) examples, computation of the derivative of the conditional moments requires knowledge of the conditional distribution of the endogenous variables (conditional on the state variables of the model), and then use of either numerical integration or simulation to calculate the appropriate integral.

Use of an instrument which requires a complete solution for all the endogenous variables destroys the raison d'etre for using Euler equation based estimation techniques in the first place. Chamberlain(1987) touches on the possibility of obtaining an estimator which achieves the efficiency bound by adding instruments sequentially from a sequence of functions which, in the limit, form a basis for a function space which is rich enough to include the efficient instruments. Newey (1990) considers the special case of homoskedastic conditional moments, and then provides conditions for achieving the efficiency bound using nonparametric (series and nearest neighbor) estimators of the efficient instruments (see also the related work on feasible GLS for heterosckedasticity by Carroll,1982, and Robinson,1987). Newey (1990) also provides a monte carlo example whose results show that use of the series estimator for the optimal instruments does amazingly well (though the nearest neighbor estimator did not). The interested reader should also consult the rapidly developing related literature on semiparametric efficiency bounds (see Chamberlain 1989, Newey 1990b, and the literature cited in those articles).

One final point. It is often worthwhile to examine the form of the "efficient" instruments even if one is unlikely (for whatever reason) to attempt to generate an estimator that attains the efficiency bound that result from them. This because an examination of the form of the efficient instruments frequently suggests sets of instruments which, though not strictly speaking efficient, should get one close to the efficiency bound, and be fairly easy to construct.

IV. Alternative Estimation Strategies, Serially Correlated Unobservables, and Invertibility Conditions.

It is natural to next ask what can be done in cases where Euler equation based estimation techniques cannot provide estimators with desirable properties for at least some of the parameters of interest. Generally, the alternative estimation strategies that are available depend on the model and data being investigated, but, at least in the context of models as simple as those used in our examples, it is probably most natural to look next at the possibility of specifying all the primitives of the model up to a vector of parameters, solving for the optimal choices implied by the different possible values for this vector, and then using either a maximum likelihood, or a minimum distance, estimator to fit the model to data.

Though estimation techniques which require computation of the value function are generally more computationally burdensome then Euler equation based techniques, their computational burden in models with a mixture of discrete and continuous choices is comparable to their burden in discrete choice models, and this is discussed extensively in the chapter by Rust(1991). Indeed, the computational issues only become significantly more difficult when we introduce interactions among agents, and are, therefore, required to solve for market equilibria. As a result we postpone further discussion of computation until we bring back in the market. For now we simply assume that we can compute the value function and the optimal policy, and look to see if this enables us to estimate the parameters of the model in situations where, because of the reasons noted above, Euler equation based estimation techniques are likely to fail.

To compute the value function we will generally also be required to make more detailed assumptions on the exogenous "forcing" processes then the assumptions we required for the Euler equation based estimators discussed in the last section. What we gain for these assumptions, and for the additional computational burden, is a set of predictions for the controls conditional on any given value of the parameter vector, all state variables, and the correspondence defining the feasible choices. Note that our ability to obtain these controls is independent of whether or not they are continuous or discrete (or, if continuous, are at a boundary of the choice set), of the form of the accumulation relationship (stochastic or deterministic), or of whether the distributions of the state variables are conditionally independent across agents.

However, the model's predictions for the controls are calculated as a function of all the state variables and the parameters of the problem. To match these predictions to data we need to express them in terms of only the observed state variables and these parameters. This requires an assumption on the distribution of the unobserved state variables. Consider first the case where these are serially independent so that their distribution at time t does not depend on their realizations at time t–1. Then, given any value for the parameter vector, we can integrate out over the current values of the unobservables that the model indicates would generate the observed controls, obtain the likelihood as a function of only observable magnitudes, and compute maximum likelihood estimators in the usual way (alternatively, we could compute, or simulate, the expected value of the controls for different values of the parameter vector, and obtain a method of moments estimator for θ). Though this may be a computationally difficult estimation procedure, it is always available, and requires no additional assumptions.[16]

On the other hand, when there are serially correlated unobservables the likelihood we calculate for the controls conditional on alternative values for θ is a function of the period t–1 value of the unobserved state variable. Thus, direct application of maximum likelihood is not possible. We could, of course, iterate backward, using the distribution of the period t–1 value of the unobserved state variable conditional on its value in period t–2 to form the distribution of the control in period t conditional on information in period t–2, and so on (see below), but we will still be left with the problem of an unknown value for the unobserved state variable in the initial period of the data. Alternatively, we could attempt to obtain the joint distribution of the values of the unobserved and observed state variables in some (preferably the initial) period, and use this to integrate out over the unobservable in forming the likelihood (again see below).

[16]Apart, perhaps, for those required to verify the regularity conditions needed to insure consistency and asymptotic normality of the maximum likelihood estimator. We should note that all we actually need for this procedure is the somewhat weaker assumption of conditional serial independence described in Rust,1988.

However, given serial correlation in the unobservable and even partial control of any of the observable state variables, the model itself will predict particular forms of dependence in the joint distribution of the observed and unobserved state variables. Thus before we can integrate out over the unobserved component we need to solve explicitly for the form of the conditional distribution of the unobserved state variable (conditional, that is, on the observed state variables), and this will require both additional assumptions, and an additional level of computational complexity. Analogous problems occur in developing method of moments estimators when there are serially correlated unobserved state variables.

As noted earlier the same issue arises in all stochastic dynamic programing models (discrete, continuous, or mixed). Indeed, the problem has a longer history than this; it has an almost identical structure to the problem labelled the "initial conditions" problem in Jim Heckman's discussion of discrete–time, discrete data, stochastic processes (see Heckman, 1981). Several solutions have been suggested in the literature. We begin by providing a brief review of some of them, focusing on their applicability to estimating structural models. Then we suggest an alternative which arises naturally for certain theoretical models with continuous (as well as discrete) controls, and provide proofs of its validity for the two leading examples used in this paper. Where applicable, the alternative is easy to adapt to the more complicated settings in which we allow for market interactions. We illustrate this below by looking at the problem of estimating the parameters of a Cobb–Douglas production function in the presence of a simultaneity (endogenous investment and labor choices) and a selection (attrition due to exit behavior) problem induced by a (serially correlated) unobserved productivity variable.

IV.1. The Problem.

We illustrate with the simple example introduced above; that of a firm choosing investment and exit strategies to maximize the expected discounted value of future net cash flow. Recall that $x(\cdot)$ provided the investment, and $\chi(\cdot)$ the exit, policies, of the firm. We shall assume here that k (capital) and ω (the state of productivity or demand) are, respectively, an observed and an unobserved state variable. Then

$$m_t=(k_t,x_t,\chi_t) \in M \subset \mathbb{R}^2 \times [0,1], \qquad (22a),$$

is the vector of observables for each firm in each period while

$$A(m_t;\theta)=\{\omega : x(\omega,k_t;\theta)=x_t \text{ and } \chi(\omega,k_t;\theta)=\chi_t\} \subset \Omega, \qquad (22b),$$

is the set of possible values for the unobservable, ω, that are consistent with the observable m_t vector, for different values of θ.

Capital accumulates deterministically, so conditional on $y_{t-1}=(\omega_{t-1},k_{t-1})$, the only source of randomness in $A(m_t;\theta)$ is the alternative possible realizations of ω_t. $\{\omega_t\}$ is a Markov process with transition probabilities given by the family $\mathbb{P}_\omega$ in (2). Assuming (for simplicity) that these have densities with respect to Lebesgue measure [these densities will be denoted by $p(\cdot\,|\,\omega,\theta), \omega \in \Omega,\ \theta \in \Theta$], the likelihood for the sequence of observables for a given firm <u>conditional</u> on the <u>initial</u> value of its <u>unobserved</u> state variable, is given by

$$\Pr(m_T,\ldots,m_1\,|\,\omega_1,\theta)=\Pi_{t=1,\ldots,T}\Pr[m_t\,|\,m_{t-1},\ldots,m_1,\omega_1,\theta], \qquad (23),$$

where

$$\Pr[m_t\,|\,m_{t-1},\ldots,m_1,\omega_1,\theta]=\int_{\omega_{t-1}}\Pr[A(m_t;\theta)\,|\,\omega_{t-1},\theta)p(\omega_{t-1}\,|\,m_{t-1},\ldots,m_1,\omega_1,\theta)d\omega_{t-1} =$$

$$\int_{\omega_{t-1}}\cdots\int_{\omega_2}\Pr[A(m_t;\theta)\,|\,\omega_{t-1},\theta]p(\omega_{t-1}\,|\,m_{t-1},\omega_{t-2},\theta)\cdots p(\omega_2\,|\,m_2,\omega_1,\theta)d\omega_{t-1}\cdots d\omega_2$$

and

$$p(\omega_j|m_j,\omega_{j-1},\theta)=\begin{cases}p(\omega_j|\,\omega_{j-1},\theta)/\Pr[A(m_j;\theta)\,|\,\omega_{j-1},\theta)], & \text{for } \omega_j \in A(m_j;\theta),\\ 0, & \text{elsewhere}\end{cases},$$

while

$$\Pr[A(m_j;\theta)\,|\,\omega_{j-1},\theta] = \int_{\omega_j \in A(m_j;\theta)} p(\omega_j\,|\,\omega_{j-1},\theta)d\omega_j.$$

The conditional likelihood for the sample, conditional on the initial values of all state variables, is formed as the product (across agents) of the likelihoods in (23). If either the initial value of the ω's of the different agents are known, of if there was no dependence in the process generating the $\{\omega_t\}$ so that $\Pr[A(m_t;\theta)\,|\,\omega_{t-1},\theta]=\Pr[A(m_t;\theta)\,|\,\theta]$, we could maximize the likelihood obtained in this fashion with respect to θ, and obtain a consistent and asymptotically normal estimator of that parameter (this presumes standard regularity conditions) .

When there is a serially correlated unobservable, several possibilities present themselves. The simplest is the conditional maximum likelihood estimator that treats the ω_1 of each individual in the sample as a parameter to be estimated [i.e. we maximize the likelihood in 23 with respect to both θ and the vector of ω_1 values]. As is well known, if limits of this estimator are taken as N grows large holding T fixed the estimator can, in general, be shown to be inconsistent (the number of parameters being estimated grows with the size of the sample and this generates a classic incidental parameter problem; see Neyman and Scott, 1948). On the other hand if limits are taken in dimension T, and the family $\mathbb{P}_\omega$ is sufficiently regular, then consistency and asymptotic normality are assured.

Panels are getting longer and it is reasonable to ask just how long they need to be before the conditional maximum likelihood estimator is reasonably well behaved. Surprisingly little research has been done on this point. Heckman (1981) reports simulation results for a problem with a single discrete alternative (no continuous control) and a disturbance process which is the sum of a (normally distributed) permanent effect and an i.i.d. (again normal) deviate, on an eight year panel. He concludes that when there are only exogenous determinants of the discrete choice, the conditional maximum likelihood estimator does well enough. However, when lagged values of the discrete choice also determine the current choice, the performance of the conditional maximum likelihood estimator is markedly worse. One's guess is that structural models that treat unobserved initial conditions as parameters to be estimated will perform more like the simulated models that allowed for lagged endogenous variables (the current choices in the structural models build up the values of the state variables that determine the choices in future periods). On the other hand there is enough of a difference between the models whose bias has been evaluated to date, and the current generation of structural models, that further Monte Carlo analysis seems to be warranted.

We next consider a class of solutions to the problem of serially correlated unobserved state variables that dates back, in a slightly different form, at least to the work of Kiefer and Wolfowitz (1956). In terms of the models considered here the Kiefer–Wolfowitz suggestion is to formulate the likelihood conditional on the initial value of the unobserved state variable [as in 23], obtain information on the joint distribution of the observed and unobserved initial values of the state variable, and then use the conditional distribution of the initial value of the unobserved state variable, say $p^*(\omega_1 | k_1, \theta_2)$, to integrate ω_1 out of (23), forming, thereby, a marginal likelihood [note that $p^*(\cdot | \cdot, \theta_2)$ will, in general, depend on a different set of parameters than those involved in the original choice problem]. The Keifer–Wolfowitz suggestion is to maximize this marginal likelihood with respect to the vector (θ, θ_2).

To do so we need the conditional density of the unobserved initial state, $p^*(\cdot | \cdot, \theta_2)$; a density which is typically unknown. There are at least two possible ways of proceeding. One is to look for a nonparametric estimator of $p^*(\cdot | \cdot, \theta_2)$ that allows us to find a semiparametric estimator for θ. Note that this requires a nonparametric estimator for a family of distributions for the unobserved initial state, one for each possible initial value of the observed state vector. We know of no research which has systematically explored this nonparametric alternative, so at this stage we simply relegate it to a topic for future research.

The second possibility is to use economic theory to derive the $p^*(\cdot)$ associated with the alternative possible values of θ. The way of proceeding here depends on the relevant model. We consider first models in which the joint distribution of the state variables of the agents active in a given market converges to some unique invariant measure (invariant to both the passage of time and to initial conditions). Models of markets with many agents, and exogenous forcing processes which are both independent across agents and ergodic, often have this characteristic (see Jovanovic and MacDonald,1990, and Hopenhayn and Rogerson,1990, for examples).

Assuming that the data is a random sample from this invariant measure, what we will need is the form of this invariant measure over (couples) of state variables. This generally requires knowledge of the parameters defining all the primitives of the model, and, in addition, an ability to compute the invariant measure associated with them. Since some of the parameters defining these primitive are the parameters we are trying to estimate, we would have to nest the problem of estimating the distribution of the initial conditions inside the estimation algorithm. That is, an evaluation of the marginal likelihood for a given value of the parameter vector would begin by calculating the invariant distribution associated with that value, and then use it as the $p^*(\cdot)$ needed to form the marginal likelihood.

Though in principle feasible, this is a very computationally intensive procedure. Moreover, in models with finite numbers of agents, and/or forcing processes which are not independent across them, both the analytic and computational problems get even more difficult. In these models the limiting characteristics of the market is often an ergodic distribution over the distribution of state vectors of the agents (as in the Ericson–Pakes model described below), and in order to integrate the unobservable initial condition out of the likelihood we would have to integrate also over the ergodic distribution of the distribution of state vectors. Though, as we show below, it might not be as difficult as once thought to calculate this ergodic distribution for one particular value of the parameter vector, calculating it repetitively for each different function evaluation required to find the maximum likelihood estimator is probably beyond our current computational capabilities.

Though in any given period there is an <u>endogenous</u> joint distribution of the observable and unobservable state variables, there may well be an initiation date for the process for each agent at which the required distribution of state variables is either a primitive to be estimated (at least up to a parameter vector), or easy to derive in terms of such primitives. Typical examples are models of firm behavior in which there is an entry date, or models of labor market behavior in which there is a date of first entry into the labor force. Given an initiation date for the process, a complete model will generate a joint distribution for the observed and unobserved state variables in the first period of the data conditional on the "age" of the agent at that time, any other presample information available, and the vector of parameters defining all the primitives of the model. This distribution is the p*() needed to obtain the marginal likelihood.

Note that the vector of parameters which define this marginal likelihood now contain also the parameters describing the entry process, and, perhaps, parameters describing changes in the environment that have occurred between the initiation of the process for the agent and the start of the panel. Thus this "solution" to the problem of serially correlated unobserved state variables does add an additional layer of computational complexity to the problem (deriving the conditional distribution of the unobserved state variables at the initiation date of the sample as a function of the parameters determining the distribution of the state variables at the initiation date of the process). However, it is an additional layer which has proven not to be too difficult in some applications, and, as a result, it is the only coherent treatment of the problem of serially correlated unobserved state variables that has been used in the estimation of structural discrete dynamic programming problems to date (see, for e.g., Miller, 1984, and Pakes, 1986).

The procedures discussed in the previous paragraphs are closely related. They both derive a form for the needed p*(·) distribution from economic theory. Indeed they only differ in that the latter makes use of presample information, and assumes that any relevant market outcomes that occur between the initiation date of the process and the beginning of the sample can be captured by a simple exogenous process. It will, therefore, be both easier to implement, and more realistic in its assumptions, when there is an exogenous "entry" date for the process we are trying to model which is close to the first sample year for each observation, and when the important sources of randomness are well described by a simple Markov process. [17]

[17]The Markov assumption is not innocuous, especially when we are trying to model a group of agents active in the same market. For example, though it may be reasonable to assume that agents take prices parametrically, it is much less reasonable to assume that agent's think the distribution of price tomorrow just depends on todays price (and no other characteristic of todays market) especially since current price is not a sufficient statistic for future price in most dynamic models; see the more detailed discussion in the next section.

IV.2. Invertibility Conditions.

Once we have a fully specified choice model, and the possibility of continuous controls, an additional procedure for dealing with the problem of serially correlated unobserved state variables presents itself. Where feasible, this solution is no more computationally burdensome then the inconsistent (in dimension N) conditional maximum likelihood estimator which treats the initial values of the unobserved state variables as parameters to be estimated. Additionally, it can often be combined with semiparametric estimation procedures to produce computationally simple estimators for models in which the value functions themselves are too difficult to compute (such as in models which allow for interactions among agents).

The technique does, however, require an invertibility condition. This condition states that there is a set of values for the observable vector each of which could only have been generated by a single value of ω — though the associated ω value can depend on θ. Below we formalize this condition and show that it is satisfied for the two leading examples used in this paper; the deterministic accumulation investment example used in this section, and the stochastic accumulation example used in the next. The proofs are, however, particular to these two classes of models. So both the feasibility of using the invertibility condition, and the form of the invertibility condition where feasible, must be investigated separately for each problem. This is the additional burden of the procedure we suggest. It is not computational, but it does require a detailed knowledge of the model one is using and the data at hand (and, as will become clear, whether or not one can obtain an invertibility condition depends on the observables available). On the other hand, provided the intuition underlying the invertibility condition is clear, one can sometimes circumvent the need for a formal proof of the condition by building a check for it into the estimation algorithm.

We first provide a formal statement of the invertibility condition (condition 24), and then show how it can be used to circumvent the initial condition problem generated by serially correlated unobservables. Recall that $m \in M$ is the vector of observables (controls and state variables), and $A(m;\theta)$ provides the set of ω values that could generate m given θ.

Condition 24. (the invertibility condition).

There is a subset of M, say M*, such that

$$\#A(m;\theta)=1, \quad \text{for all } \theta\in\Theta \text{ and } m\in M^*,$$

In (24), # provides the cardinality of a set, so the condition is that $A(m;\Theta)$ is a singleton for m in M*. Also, it is important to note that we do not require the condition to hold for all $m\in M$, but rather just for m in the subset M*, as M* will tend to be a proper subset of M in problems with discreteness in the choice set or the possibility of binding boundaries (see below).

Now assume (24). Then if m_τ is in M* we know the value of ω_τ for any value of θ; i.e. $\omega_\tau=\omega^*(m_\tau;\theta)$ for $m_\tau\in M^*$. So substitute $\omega^*(m_\tau;\theta)$ for the unobserved ω_τ in (23), and use this as the initial condition needed to compute a maximum likelihood (or a method of moments) estimator from the predictions of the model for periods $\tau+1$ to T. The product of this likelihood across agents will depend only on the period τ vector of observables for each agent, and the finite dimensional parameter vector to be estimated.

More formally, define the stopping time

$$\tau = \begin{cases} T & \text{if} \quad \cup_t m_t \cap M^* = \emptyset \\ \min\{t: m_t\in M^*\} & \text{otherwise.} \end{cases} \qquad (25a)$$

And form the truncated conditional likelihood

$$\Pi_{t=\tau+1,\ldots,T} \Pr[m_t \mid m_{t-1},\ldots,m_\tau,\omega^*(m_\tau;\theta)], \qquad (25b),$$

where

$$\Pr[m_t \mid m_{t-1},\ldots,m_\tau,\omega^*(m_\tau;\theta)]=$$

$$\int_{\omega_{t-1}} \cdots \int_{\omega_{\tau+1}} \Pr[A(m_t;\theta) \mid \omega_{t-1},\theta] p(\omega_{t-1} \mid m_{t-1},\omega_{t-2},\theta) \cdots$$

$$p(\omega_{\tau+1} \mid m_{\tau+1},\omega^*(m_\tau;\theta),\theta) d\omega_{t-1} \cdots d\omega_{\tau+1},$$

and the terms in the integral are defined as in (23).

Assume the set M^* has positive probability. Then maximization of the product (across agents) of the truncated conditional likelihood in (25b) with respect to θ, will, subject to "standard" regularity conditions (see Andersen, 1973 section 2.8, or Huber,1967) produce consistent and asymptotically normal estimators of that parameter.

We now go back to our example of a firm making investment and exit decisions to maximize the expected discounted value of future net cash flow. To prove that it satisfies the invertibility condition, at least for the subset of M for which $x>0$, we will have to restrict its primitives somewhat. The additional restriction that has empirical bite is that the derivative of the profit function with respect to capital must be increasing in the unobserved state variable, ω [more generally, $\pi(\omega,k)$ must be supermodular in the sense of Topkis,1978, see also Milgrom and Roberts,1990]. Though this condition is satisfied for most specifications used in empirical work (where ω generally represents either Hicks neutral efficiency differences in production, or quality differences among a group of differentiated products firms; see the examples below), it is easy to generate counter examples where it is not. In addition we will (partly to keep matters simple) impose additional regularity conditions on the family of probability distributions, $\mathbb{P}_\omega$. We begin with a lemma which insures that under our conditions the investment policy is nondecreasing in ω.

<u>26</u>.<u>Lemma</u> (monotonicity of the investment policy in the investment example).

Assume 2, that $\partial\pi(\omega,k)/\partial k$ is increasing in ω (everywhere), and that if $h(\cdot)$ is continuous (everywhere) and uniformly integrable with respect to a subset of $\mathbb{P}_{\omega'}$, say $\mathbb{P}^*$, then provided $P(\cdot|\omega_1)$ and $P(\cdot|\omega_2)$ are contained in $\mathbb{P}^*$, $|\int h(\omega')[P(d\omega'|\omega_1)-P(d\omega'|\omega_2)]| \leq \psi(h,\mathbb{P}^*)|\omega_1-\omega_2|$. Then

$$x(\omega,k) \text{ is nondecreasing in } \omega \text{ for each } k.$$

Proof. See appendix 1.

Theorem 27 provides the invertibility condition for the investment example.

Theorem 27.

Conditional on the assumptions underlying Lemma 26, condition 24 is satisfied for the subset of M on which $x>0$.

Proof.

From lemma 5 the investment choice must satisfy the Euler equation

$$F(x,k,\omega)\equiv-\partial c(x)/\partial x +$$
$$\beta\int \chi[\omega',k(1-\delta)+x]\{\partial\pi[\omega',k(1-\delta)+x]/\partial k+(1-\delta)\partial c[x(\omega',k(1-\delta)+x)]/\partial x\}P(d\omega'|\omega) = 0.$$

Assumption 2, together with the form of the optimal policy (see 3) insures that $F(\cdot)$ is a continuous function of ω for every (x,k). So it will suffice to show that for each (x,k), $F(\cdot)$ is strictly increasing in ω. (3) insures that $\chi(\cdot)$ is nondecreasing in ω', and the convexity of the adjustment cost function together with lemma (26) insure that $\partial c(\cdot)/\partial x$ is also, while $\partial\pi(\cdot)/\partial k$ is strictly increasing in ω' by assumption. Thus the integral is nondecreasing in ω' everywhere and strictly increasing for all ω' in the region where $\chi=1$. To complete the proof, then, one need only note that if $x>0$, $\chi[\omega',k(1-\delta)+x]$ must be 1 on a set of ω' with positive $P(\cdot|\omega)$ probability (else x=0 would generate a more profitable program).

Remark 1. The statement of lemma 26 given in the appendix allows the cost of adjustment function to depend also on k. The condition on c(x,k) that suffices for the lemma in this more general case is that it be nonincreasing in k for each fixed x. When the adjustment cost function depends on k the Euler equation for the investment choice becomes,

$$F(x,k,\omega)= -\partial c(x,k)/\partial x$$
$$+\beta\int \chi[\omega',k(1-\delta)+x]\{\partial\pi[\omega',k(1-\delta)+x]/\partial k+(1-\delta)\partial c[x(\omega',k(1-\delta)+x),k(1-\delta)+x]/\partial x$$
$$-\partial c[x(\omega',k(1-\delta)+x),k(1-\delta)+x]/\partial k\}P(d\omega'|\omega) = 0.$$

Now to obtain the strong monotonicity result in (27) we need also that $-\partial c(\cdot)/\partial k$ is nondecreasing in x. If $c(\cdot)$ is appropriately differentiable, then what we require is that $\partial^2 c(x,k)/\partial x\partial k < 0$, a condition which is satisfied for most cost of investment functions used in empirical work. ∘

Remark 2. Theorem 27 does not depend on the feasibility of negative investment. That is, if we constrained investment to be nonegative we could use the modified Euler equation in (7b) to prove the same condition (the proof would only be true for the subset of M for which the modified Euler equation is indeed satisfied, but that would include all observed vectors for which x>0). ∘

Theorem 27 implies the existence of a function $\omega^*(x,k;\theta)$ with the property that if $m \in M^*$, then $\omega=\omega^*(x,k;\theta)$. Following the discussion above, a truncated conditional maximum likelihood estimator for this problem can then be constructed by defining the $\tau(i)$ in (25a) to be the first observation on firm i for which we observe positive investment, substituting $\omega^*(x_{\tau(i)},k_{\tau(i)};\theta)$ for the needed initial condition into (25b), and then maximizing the sum (over firms) of the resulting log likelihoods with respect to θ.

Once we know the inverse function exists, alternative estimation strategies present themselves. Of particular interest are estimation techniques that use a nonparametric estimator of the inverse function, thereby circumventing the need to compute that function for different possible values of the parameter vector. Nonparametric alternatives are often feasible in situations where we want to control for ω in order to attenuate biases resulting from the presence of this serially correlated unobserved state variable.

Olley and Pakes (1990) study of productivity in the telecommunications equipment industry provides one example of the use of such a semiparametric estimation procedure. To construct their measure of productivity they required estimates of the industry's production function. This is specified in Cobb Douglas form as

$$(28) \quad q_{i,t} = \alpha_0 + \alpha_a a_{i,t} + \alpha_l l_{i,t} + \alpha_k k_{i,t} + \omega_{i,t} + \eta_{i,t},$$

where q,k, and l are the logarithms of output (value added), capital (constructed from a geometric decay assumption and data on the book value of the plant in the initial year the plant enters any of the census' files), and labor (manhours), while "a" is the plant's age (this allows for vintage effects, or for an initial sunk factor of production whose impacts decay from birth). The data are taken from the Longitudinal Research Data File at the U.S. Bureau of the Census (see McGuckin and Pascoe,1989). This is a thirteen year (1973–86) panel which follows information at the enterprise (plant) level of aggregation.

The model has two disturbances affecting observed productivity, ω and η. The distinction between them is that the firm is allowed to adjust its decisions to realizations of ω, but not to those of η (so that η is either measurement error, or a serially uncorrelated productivity shock that is realized after input decisions are made). Since ω effects the firm's decisions, it is the source of concern about the consistency of the O.L.S. estimates of the parameters in (29). There are two reasons for these concerns. The first, which dates back to Marschak and Andrews'(1944) classic article, is that if ω is serially correlated, and input decisions are at least partially subject to control, then inputs in place will be correlated with current ω, and this will generate a simultaneous equation bias in the O.L.S. coefficients. The second, which though discussed in the empirical literature for some time (see, for eg., Wedervang, 1965) had not been previously incorporated into the econometric analysis of production functions, is that we only observe firms that do not close down and, if more productive firms tend to be more profitable and survive longer, the selection on survival is, in part, a selection on ω, producing a selectivity bias in the O.L.S. coefficients (for a review of the literature on selectivity biases in the labor econometrics literature, see Heckman and Robb, 1986, and the literature cited therein).

To devise an estimation procedure which takes account of the simultaneity and selectivity biases we need a model for input and exit decisions. For this Olley and Pakes (1990) use the model in our example, augmented to allow for a labor choice and for an additional state variable (age). Labor is assumed to be variable so that it is contracted for at the beginning of the period and can be adjusted, perhaps at increasing cost, to realizations from the conditional distribution of ω_t. As in the model outlined above exit occurs whenever $\omega_t \leq \underline{\omega}_t(k_t, a_t)$, with $\underline{\omega}_t(\cdot)$ decreasing in k, while $x_t = x_t(\omega_t, k_t, a_t)$ with $x_t(\cdot)$ strictly increasing in ω for each (k,a) whenever $x>0$. The investment and exit function are indexed by t to allow for changes in market conditions over the period covered by the data (see Olley and Pakes,1990).[18]

Since we are only interested in the use of the invertibility condition we only reproduce the initial stage of their estimation algorithm, the stage that obtains the labor coefficient, α_l. The simultaneity problem here is a result of the correlation of l and k (through past choices) with ω,

[18]The framework used in Olley–Pakes allows for interactions among agents. They assume that the profits a firm earns in a given period depends not only on its own state variables, but also on the list of state variables of competing firms, and that the data are generated by a Markov–Perfect Nash Equilibrium in investment, exit, and entry strategies (see the next section). The model then becomes a modified version of the Hoppenhayn–Rogerson (1990) model of industry dynamics, and the firm's decisions depend on both the firm's own state variables, and the measure providing the list of state variables of the competing firms.

while the selection problem is a result of the fact that the survival process truncates the distribution of the ω observed in the data (and the truncation point is a function of the right hand side variables in the equation we want to estimate). We could account for both of these problems if we could condition also on ω. The invertibility condition tells us that for those observations with x>0 we can do just that by substituting $\omega_t{}^*(\cdot)$ for ω_t in 28 producing the equation

$$(29) \quad q_{i,t}=\alpha_0 + \alpha_l l_{i,t} + \phi_t(x_{i,t},k_{i,t},a_{i,t}) + \eta_{i,t}$$

where,

$$\phi_t(x_{i,t},k_{i,t},a_{i,t}) = \omega_t{}^*(x_{i,t},k_{i,t},a_{i,t}) + \alpha_a a_{i,t} + \alpha_k k_{i,t}.$$

The first stage of the estimation algorithm uses a polynomial approximation to the $\phi_t(\cdot)$ function to obtain a semiparametric estimator of $\alpha_l(\cdot)$. Since $\omega^*{}_t(\cdot)$ is a function of all the state variables of the problem (all the determinants of the investment choice), the semiparametric procedure does not allow us to separate out the effects of k and a on investment, from their direct effects on output. Olley and Pakes (1990) proceed to show how, by considerring also the restrictions from the expectation of $y_t - a_l l_t$ conditional on k_t, a_t and ω_{t-1}, one can also obtain consistent estimates of α_k and α_a (here b_l is the first stage root n consistent estimate of a_l).

Alternative estimates of α_l and α_k (standard errors in parenthesis) and the relevant sample sizes, are presented in table 1. The estimates in the first two columns are computed from a "balanced panel". The balanced panel is obtained by using only the information on the plants that were active for the entire 13 year period. Balanced panels are the traditional way of drawing samples for use in production analysis. Columns (3) and (4) use the "full" sample; this contains information on all plants ever active in all years that they are active (except those plant year observations that have zero investment, as these observations do not satisfy the invertibility condition).

The first point to note is that by using the balanced panel we discard about two–thirds of

Table 1

ALTERNATIVE ESTIMATES OF COBB-DOUGLAS PRODUCTION FUNCTION PARAMETERS*

Sample	Balanced Panel		Unbalanced or Full Panel	
Estimation Technique	O.L.S.	Within	O.L.S.	Olley/ Pakes
Coefficient of	(1)	(2)	(3)	(4)
Labor	.87 (.04)	.77 (.05)	.70 (.02)	.62 (.02)
Capital	.16 (.03)	.05 (.05)	.31 (.02)	/
Number of Observations	886	886	2,397	2,397

*Source, Olley and Pakes, 1991, Table 6. Estimated standard errors are in parenthesis. Other variables in all equations are plant age and a time trend. The balanced panel uses only the data on plants which were active in every year of the 13 years of the panel. The unbalanced panel uses data on all plants that were ever active in every year they were active.

the observations, so the potential for selection problems is large. Column 1 provides the O.L.S., while column 2 provides the within, estimators from the balanced panel (the within estimator uses deviations from firm–specific means for all variables; see Chamberlain, 1984, for a detailed discussion). The within estimator would be appropriate if the effects of selection and simultaneity differed by firm but were constant for a given firm over time (note how much at odds this is with the model; according to the model the firms who exit were firms who changed their perceptions of their future profitability over the period, and one would expect perceptions to be correlated with realizations).

The total and within coefficient estimators from the balanced panels are not unusual for production function estimates from balanced panels. The labor coefficient is higher than what seems plausible for the elasticity of output with respect to labor, and the capital coefficient is noticeably lower than what we think plausible for the capital elasticity. Theory suggests at least two explanations. First labor, being easier to adjust, is more highly correlated with the current value of ω (simultaneity). Second, the exit rule is decreasing in k, so low capital firms who survive need to be firms who drew exceptionally good productivity sequences, while firms with more capital will survive on much poorer productivity draws. This induces a negative correlation between capital and productivity among survivors. Note also that since labor and capital are positively correlated positive biases in one coefficient will tend to be associated with negative biases in the other.

Column 3 provides the O.L.S. estimates on the full sample. We expect moving to the full sample to alleviate much of the selection problem, though not necessarily the simultaneity problem. The results are quite striking. The capital coefficient more than doubles, and the labor coefficient moves down by over 20%. There should still be a bias in these coefficients that can be eliminated by substituting the polynomial approximation to the $\omega^*(\cdot)$ function for the unobserved value of ω_t Just as theory says it should, use of the polynomial approximation to the inverse function forces the labor coefficient down still further, by another 10%, so that the final estimate of the labor coefficient was close to labor's share in value added (the final estimate of the capital coefficient, which is not reported in the table, was .345 with a standard error of .05).

Where applicable, this combination of the use of theory (to prove the existence of an inverse function), and of semiparametric estimation techniques (that allow us to use that inverse function without ever computing its form), should be quite useful, as it ought to allow us to account for the problems induced by serially correlated unobservables with estimation algorithms which are computationally quite simple. There may, of course, be many cases in which difficulties arise in checking for the regularity conditions which insure the consistency and asymptotic normality of the semiparametric estimator, or in computing its variance covariance matrix and insuring there is a consistent estimator of it, or, perhaps, in finding an efficient semiparametric estimator for the structure at hand. However, the econometric literature on semiparmetric estimation has been advancing at an extremely rapid rate (see Ahn and Powell,1990; Andrews 1989a,1989b; Chamberlain,1989; Newey, 1989b and 1991; Powell,Stock and Stoker,1989; Robinson,1988; Sherman,1990; and the literature cited in those articles); and it may not be too long until many of the relevant issues are clarified. Then the major problem facing the applied researcher will be to formulate the invertibility condition for the problem at hand, and show how it can be used to identify the finite dimensional parameter vector of interest — a task that generally requires a deep understanding of both the appropriate model, and of the data at hand.

Having provided detail on one example where the invertibility condition takes on a relatively simple form, it is worth reemphasizing that the results on that form are model specific. Indeed, our ability to use investment (conditional on capital) as a proxy for the unobserved state variable in this model depends crucially on the fact that, in models with deterministic accumulation that satisfy the assumptions in 26, the expected increment to the value of the firm arising from a given increment in <u>capital</u> is increasing in the unobserved state variable. In models in which there is stochastic accumulation (such as the one outlined in section III.3) this monotonicity condition, and hence the associated invertibility condition, is unlikely to be satisfied.

In these models the role of investment is to improve the distribution of ω_{t+1} conditional on ω_t, and the increment to the value of the firm generated by given increments in ω are not characteristically monotonic in ω. In particular, boundedness of the value function implies that the increment in value per increment in ω must eventually be concave in ω, and there is often reason to expect the value function to be convex in ω at lower levels of development (see, for example, Ericson and Pakes,1989,part I). As a result we have to look for alternative ways of controlling for efficiency differences in situations in which stochastic accumulation model seems appropriate.

The version of the stochastic accumulation model presented in the next section is one in which firms are differentiated by the quality of the product they produce. Consumers have a distribution of tastes over the alternative products, and an increase in the quality of any one product (in its unobserved ω) will increase demand for that product conditional on any vector of prices and any vector of the ω's of a firm's competitors. Firm's are price setters, and the equilibrium in the spot market for current output is Nash in prices conditional on the ω's of all active firms. The ω's of the firms evolve over time according to the stochastic outcomes of the firm's investment decisions (and investment, entry and exit decisions are made to maximize the expected discounted value of future net cash flows).

Berry (1990, appendix 1) shows that under these conditions, and some mild restrictions on the distribution of preferences over consumers, there is a one to one map between the vector of market shares of the various competitors and the vector of unobserved efficiency differences. This map can therefore be inverted to obtain the unobserved efficiency differences as a function of the observed market shares and the parameters of the model. So there is an inverse function for this model, but it has a different form, and requires different data, then the inverse function for the model with deterministic accumulation.

The point to emphasize here is that the existence of the invertibility condition, and its form where available, depends on the details of the model that is appropriate for the problem one wants to analyze, and the data avilable. There is simply no substitute for a deeper understanding of the major sources of unobserved variation in the data, and on how these unobservables are likely to interact with the observed deviates.

Given the possible complexity of the invertibility condition, there may be cases where the intuition underlying it is clear enough, though the formal justification for its use is difficult to obtain. Our suggestion here is to begin by simply computing the value function (or the equilibrium condition) underlying the invertibility condition for different values of the parameter vector, and then inspecting the solution for the required properties. If one finds that the condition is satisfied, but the proof is still not available, one may be able to extend this numerical procedure one step further, and actually program a check of the invertibility condition into the computations at each iterative stage of the estimation algorithm.

This suggestion presumes that the estimation algorithm requires computation of the necessary relationships. As shown above, for more complicated models where computation of the needed functions can be very difficult, it is often easy to simplify the computational burden of the estimation procedure by combining an invertibility condition with a semiparametric estimator for the inverse function. When formal proofs of the invertibility condition for these more complicated cases are not available, but the intuition underlying it is still strong, the suggestion is to use it, together with the semiparametric estimator for the inverse function, to provide a simple "diagnostic" test for the presence of a serially correlated unobservables, and some indication of just how much of an effect it may have on the parameter estimates.

The idea behind using an invertibility condition is essentially the same as the idea of using a "proxy" to substitute for an unobserved variable; albeit a proxy whose values typically depend on the parameters being estimated as well as on the vector of observables, and a proxy which can typically only be justified for a subset of the possible realizations of the vector of observables. The connection to the use of proxy variables makes it clear that there is a long history of related research on accounting for unobservables in econometric models; most recently in the semiparametric selection models (see Ahn and Powell, 1990, Choi, 1990, Newey,1988, and the literature cited in these articles)[19].

Much of the prior literature on "proxy" variables focussed on linear models. Latent variable models (see the review by Aigner, Hsiao, Kapetyn, and Wansbeek, 1984), and dynamic factor analysis models (see Geweke,1977, and Sargent and Sims,1977) are two of the more successful examples. Also related is the analysis of the initial condition problem for dynamic linear models on panel data (see Anderson and Hsiao 1982, Pakes and Griliches,1984, and the literature cited in those articles). For the most part neither of these literatures worried about deriving the linear system analyzed from the primitives of a behavioral model, so the issue of the realtionship of the inverse function to those primitives did not arise (for notable exceptions in the context of dynamic representative agent models see Hansen and Sargent,1990; and in static equilibrium contexts with heterogeneity, see Heckman and Scheinkman,1987). Also, once one incorporates either discreteness in the choice set, or interactions among agents, nonlinearities typically appear in the relationship between the observable vector and the unobserved deviates we want to control for.

On the other hand most of the linear models did allow for disturbances in all the relationships of interest. In contrast our discussion has assumed that the nonlinear relationship between the observables and the unobserved state variable holds exactly. A logical next step would seem to be to allow for measurement error in some of the observables used in the model. The truncated conditional likelihoods, or the truncated conditional moment restrictions, would then be in terms of the "true" unobserved variables, and, since we would only observe the error prone deviates, estimation would require a solution to a nonlinear errors in variable problem.

---

[19]I would like to note that the idea of using invertibility conditions to account for serially correlated unobserved state variables in a model of asset pricing is clearly set out, though without the required proofs, in an unpublished manuscript Bent Christensen (1990), gave to me.

Research on nonlinear errors in variables models (see Fuller, 1987, chapter 3; Hausman, Ichimura, Newey, and Powell, 1991; Newey and Powell 1988, and Newey,1990) has been proceeding rather rapidly, so it may well be possible to incorporate errors of measurement into the analysis of invertibility conditions.

It is appropriate to conclude this section on a more general point. What is clear is that once we allow for serially correlated unobserved state variables, the properties of our estimators are going to have to depend on a set of very detailed assumptions on the way those variables effect the primitives of the model. A successful researcher, then, is likely to have to develop a fairly detailed understanding of what are the major sources of unobserved variation that effect behavior, and of how they interact with the other primitives of the problem. The alternative, however, is to assume that all unobserved deviates are serially uncorrelated. This is, of course, even more of a restriction than those needed for the models that allow for serially correlated state variables. Moreover, at least for many problems of current interest, it is an additional restriction which is simply untenable.

V. Market Interactions and the Computation of Equilibrium Responses.

We now consider one way of incorporating market interactions into our examples, make some brief comments on related estimation problems, and then focus in on computing equilibrium responses assuming that the parameters defining the primitives of the equilibrium problem have already been estimated.

To incorporate market interactions we allow the returns an agent earns in a given period to depend not only on the value of the agent's own state vector (y), but also on the vector of state variables of the other agents active in the market, $\hat{s}$. Recall that $s=(\hat{s},y)$, is the list of state variables of all active agents. It will be assumed that there is a finite upper bound to the number of agents active in a given period (a condition, which should, in general, be shown to be a consequence of the primitives of the model). So a particular value of s is a finite list of the state vectors of the firms currently active in the industry, and will be called an industry structure. In the deterministic accumulation example, then, the state vector for a given firm is a couple (ω,k), so s is a finite counting measure on $\Omega \times K$. Note also that the vectors (s,y), and $(\hat{s},y)$, carry precisely the same information so, for notational convenience, we will use (s,y) where possible.

The assumption that the current returns the agent earns depends on s, as well as the agent's y, implies that the likely profitability of a firm's investments depend on the investments of its (potential and actual) competitors. As before we shall assume that all decisions are made to maximize the expected discounted value of future net cash flow conditional on the current information set. That information set includes a distribution for the counting measure of possible industry structures in future years conditional on the current structure. The equilibrium notion we use to close the model insists that this distribution is in fact consistent with optimal behavior by all incumbents and potential entrants.

It is important to note that though we do allow the firm's profits to depend on the state variables of competing firms, we will, throughout, restrict those state variables to the set of variables which determine either current production costs or current demand conditions (to use the terminology of Tirole,1989, to "payoff relevant" state variables). Strategies, in turn, will be restricted to depend only on the vector (s,y) (in particular they cannot depend on previous actions). Our assumptions, then, require the equilibrium to be Markov–Perfect Nash in investment strategies(in the sense of Maskin and Tirole,1987,1988a,and 1988b).

The extent to which the focus on the Markov Perfect Nash assumption limits the nature of the equilibria we study depends on the dimensionality of y. Since the burden of the computational algorithm also goes up (and quite rapidly) in the dimensionality of y (see below), there will often be a trade off between the richness of the equilibria that the applied researcher allows for, and the computational burden of the subsequent analysis (and, as in other tradeoffs discussed above, our feeling is that it should be decided on a case by case basis according to the characteristics of the applied problem one wants to analyze). Note also that our discussion does not allow for nonpecuniary spillovers among firms (à la Roemer,1986), or for asymmetric information (for recent empirical work on structural models with asymmetric information in static contexts see Hendricks,Porter,and Wilson,1990, and Wolak,1990). These are reasonably glaring omissions which impose serious limits on the applicability of the results developed here. On the other hand, one has to start somewhere, and it is not analytically difficult to bring more detail into the current structure provided the basic behavioral assumptions used here are appropriate.

Our attitude towards structural estimation in applied problems where the interactions among heterogeneous agents are important, is that the strategy of estimating the model's parameters by solving for the complete set of dynamic equilibrium responses for different candidate values of the parameter vector, and then fitting these into an iterative maximum likelihood (or minimum distance) search procedure, has both computational and data requirements that are unlikely to be satisfied in the near future (at least for many problems of current interest). It therefore becomes essential to develop techniques that allow one to break the estimation problem down into smaller parts. Each part should allow the researcher to obtain an estimator for a subvector of the total vector of the model's parameters. This estimator should be obtainable without having to solve for the complete set of equilibrium responses, but should be consistent and asymptotically normal under the complete set of equilibrium assumptions. A typical breakdown will involve splitting off the static return function from the complete dynamic system and obtaining consistent estimates of its parameters in one part of the estimation algorithm, and then splitting off the problem of estimating the parameters defining the impact of investments on subsequent values of the state variables into another estimation subroutine. Part of the reason for our focus on Markov Perfect Nash equilibria is that they make it relatively easy to separate out the estimation of the parameters defining the one period return function from those defining the dynamic impacts of decisions.

On the other hand, once we have our estimated parameters, we will still want to use them to compute the equilibrium they imply, and then investigate how that equilibrium varies with policy and environmental changes. Thus we are still in need of an algorithm capable of computing equilibrium responses, but not one that needs to be fast enough to enable us to nest it into an iterative estimation algorithm. Section V.2 provides an algorithm for computing the equilibrium implications for the class of Markov Perfect Nah equilibria we focus on.

When we modify the deterministic accumulation example (assumption 2) to allow profits to depend also on the list of state variables of competing firms and then close it with an entry process, the example becomes a version of the Hoppenhayn–Rogerson (1990) model of industry dynamics (a model which is, perhaps, the most straightforward heterogeneous agent extension of the traditional production–investment model). This is the equilibrium model that underlies estimation of the Cobb–Douglas production function in the example in section IV.2. When we modify the stochastic accumulation example (section III.3) for the same factors the example becomes a version of the Ericson–Pakes(1990) model of industry dynamics. This will be the example used to illustrate the computational algorithm introduced below.

The last section mentioned estimation algorithms for subvectors of both these models parameter vectors that do not require iterative computation of equilibrium responses. We emphasized their that the availability and form of these estimation techniques depended on detailed characteristics of the model relevant for the data at hand. Though there are interesting and important general estimation issues here, we have chosen not to discuss them in this paper. Instead we focus on the problems involved in computing equilibrium responses conditional on having an estimate of the value of the model's parameter vector in hand. It would be inappropriate, however, to proceed directly to the computational issues without at least noting some of the problems that arise in generalizing the estimation techniques discussed in the last sections to models where there are market interactions.

Since much of our discussion of invertibility conditions already incorporated market interactions, we do not have much to add to our discussion of the use of invertibility conditions. So we go directly to the potential for integrating market interactions into Euler equation based estimation techniques. Here the prognosis is not as bright. At least in "small" markets, that is markets in which marginal changes in one firm's state variables in the current period generate nontrivial reactions by the firm's competitors in the following period, Euler equation based estimators will not generally be feasible. That this is so despite the fact that in some of these cases one can in fact use perturbations to the continuous control to derive Euler equations (eg. the alternating move games discussed in Maskin and Tirole; see in particular their 1987 article), is a result of the fact that the restrictions that result from these Euler equations will involve a term giving the reactions of the firm's competitors to the perturbation in the given firm's control. The needed "reaction function" is not a primitive of the model, but rather an endogenous construct, and to determine its form we generally have to know the form of the solution for the equilibrium responses. Without either more work, or more assumptions, then, we will be unable to use the restrictions embodied in the Euler equations to derive estimators for the model's parameters without solving first for the equilibrium responses generated by the different trial values of the parameter vector — a strategy which, as noted earlier, we want to avoid.

In some cases, of course, there will be ways around this problem. For example, in special cases where we can show that the reaction functions must have a simple form (eg. linear, as in the linear quadratic game literature; see, for example, Kydland,1975), we should be able to estimate the parameters of the reaction function along with the other parameters in the Euler equation from the restrictions that the Euler equation generates. A second possibility is to obtain a nonparametric estimate of the reaction function, substitute it into the Euler equation, and then derive a semiparametric estimator for the rest of the parameters that determine the restrictions emanating from the Euler equation (see the literature on semiparametric estimation referred to in the last section). Finally, we should note that even in cases where "dynamic first order" conditions cannot be used as a basis for estimation, one still may be able to use them to derive analytic characteristics of the optimal policy that insure that other form's of estimators are feasible. Of particular importance here is the use of first order conditions to show that invertibility conditions are indeed satisfied for some subset of the possible set of values of the observable vector.

We now leave the topic of estimation to consider computation of equilibrium responses.

V.1 Equilibrium Responses

An underlying purpose of structural modeling is to obtain a deeper understanding of the responses to policy and environmental changes. This will require, in addition to estimates of the appropriate parameters, an ability to compute the equilibrium implications of those estimates. This subsection assumes that we have estimated the model's parameters, and focuses on the problem of computing their implications.[20]

Once our models acknowledge the fact that agents do differ, and grant that their actions may impact on one another, then the computation of the responses we are typically after can become quite demanding. That this is so even if we are only after the aggregate impact of a given change, is a result of the simple fact that agents responses in realistic models are typically different nonlinear functions of the changing variable (the nonlinearity becomes most obvious once one allows for discrete alternatives), and the aggregate response we are after is usually a weighted average of the individual responses. Note that an analysis of aggregate impacts in such a world requires not only the distribution of responses of the agents currently active in the given market, but also, if agents can enter or exit that market, the equilibrium response of that distribution to the policy or environmental change.

Moreover, we are often specifically interested in the relationship of policy, or environmental, variables to the more detailed structure of the distribution of agents characteristics, and in how the equilibrium distribution of those characteristics is likely to react to the policy or environmental change. Obvious examples where more detailed knowledge of the determinants of the equilibrium distribution is of overriding importance are easy to come by in almost all aspects of economics. The analysis of the link between default probabilities and the market for finance capital, and of the effects of the various deregulatory changes on market structure, are examples that occur repeatedly in the

[20]This section draws heavily on Pakes and McGuire (1991).

finance and industrial organization literatures. More recently, the finding that almost all of the variance in gross job creation and gross job destruction is within time–period, within–industry, variance (see Davis and Haltwinger,1989), makes any analysis of the causes or the effects of job turnover in labor markets highly dependent on the detailed characteristics of the equilibrium from dynamic heterogeneous agent models (see Hoppenhayn and Rogerson,1990, for a start at such an analysis).[21] This section focuses on problems involved in computing the equilibria from dynamic stochastic heterogeneous agent models (assuming that all the parameters defining the primitives of that model are known).

As noted earlier we focus on Markov–Perfect Nash equilibria, and again we find it convenient to illustrate our points with a particular example. The theoretical structure of the example is taken from Ericson and Pakes (1990), and it generalizes the single agent stochastic accumulation example used in section III.3 to allow for market interactions. The differentiated product special case for the spot market for current output used in the computations is adapted from Berry(1990), and the structure of the computational algorithm is taken from Pakes and McGuire (1990).

The example provides us with a special case to use to test the computational algorithm. It has the additional, and at least for testing purposes desirable, characteristic that the Markov Process which defines the equilibrium in this example lives on a finite collection of points. We can, therefore, calculate the value functions and policies the equilibrium generates exactly (or at least to any desired degree of precision), and then

[21]The Hoppenhayn and Rogerson paper is also one of the few that worries about computation in heterogeneous agent models. They assume that all agents are zero measure and all sources of uncertainty are idiosyncratic, show that under their conditions the industry structure converges to a fixed s^* (and stays there) and then provide a simple way of computing s^*. Judd (1990) has computed Markov perfect equilibria for two agent models with no entry and exit, and Hansen and Sargent (1990) provide a computational algorithm for a class of heterogeneous agent models that allow for linear decision rules and equilibria (they assume quadratic preferences, linear technologies and information sets, no discrete choices, and that continuous choices are always interior).

compare the exact results to results based on various approximations.

We begin by calculating the exact solution. Then we illustrate the richness of the solutions one gets from structural heterogeneous agent models by simulating the ergodic distribution of market structures, and characterizing firm behavior, for a particular numerical example.

We also show, however, that the number of points at which we have to evaluate the value function to obtain the exact solution goes up as a polynomial in the number of agents ever active in the market. The number of computations per point evaluated also grows as a polynomial in this number. Exact computation will, therefore, become computationally impossible for a market with a large enough number of agents. For our example we cannot really go beyond a seven agent equilibrium on our sparc 1.

We therefore move on to show how one can use procedures based on polynomial approximations (and/or interpolation) to cut down the number of points at which we must calculate the value function in the computational algorithm. Our major result in this context is analytic. We prove that provided the value function of a given agent is symmetric (more precisely exchangeable) in the state vectors of its competitors, the number of polynomial coefficients one needs for a given order of polynomial approximation is independent of the number of agents active in the market.

To get some idea of how good the polynomial approximations could be, we fit polynomials directly to the exactly calculated value functions for our example. It is reasonably clear that the fit of a polynomial with a given number of coefficients does not depend on the number of agents active in the market (at least when we measure fits by a simple R^2 criteria). Two other points come out of these exercises. First, one can often do a lot better than using simple polynomial expansions , particularly if one knows something about the problem being analyzed. Second, and most encouraging, it seems that one can fit the polynomials, or the other approximating functions, to a small but reasonably diffuse subset of the total number of points, use the approximating functions obtained in this fashion to fit all the points, and do just about as well as one would have done by fitting the whole set of points directly.

V.1 The Example.

Recall that in the model with stochastic accumulation firms invest to explore, and if warranted, develop, profit opportunities (improved goods or techniques of production). The outcomes of the investment process are uncertain. Positive outcomes lead the firm to states where the good or service can be marketed more profitably. If the outcomes generate lesser increments than those of its competitors (both inside and outside the industry) the firm's profits deteriorate, and may lead to a situation in which it is optimal to abandon the whole undertaking (thus generating exit). A firm's supply to the spot market for current output, and its current profits, depend on its own level of development, a counting measure which provides the levels of development of its competitors in the industry, and on the level of development of an alternative outside of the industry. The level of development of the outside alternative evolves exogenously. Entry and investment decisions (which determine the levels of development of the actors in the industry) are made to maximize the expected discounted value of future net cash flow conditional on the current information set. The equilibrium notion is Markov Perfect Nash in investment strategies.

We begin by providing a brief description of each of the primitives of the model, starting with the profit function, then turning to the other primitives determining incumbent behavior, and finally to those determining the behavior of entrants. We then give a verbal characterization of the aspects of the model's equilibrium that we want to investigate (for more detail, see Ericson and Pakes, 1989).

The state variables determining the firm's perception of its opportunities are

$$(\omega,s) \in \Omega \times S \subseteq \mathcal{Z} \times \mathcal{Z}_+^{\mathcal{Z}},$$

where ω is an index of its own efficiency, s is a counting measure providing the number of firms at each possible efficiency level, and $\mathcal{Z}$ is notation for the integers. "s_t" defines the structure of the industry at each t.

Thus

$$\pi(\cdot,\cdot):\Omega X S \to \mathbb{R},$$

provides the "reduced" form of the current profit function. In the general case we need only that $\pi(\omega,s)$ is

i) increasing in ω for all s

ii) that there exists a complete preorder on S, say $\succeq$, s.t. if $s_1 \succeq s_2$, then $\pi[\omega,s_1] \geq \pi[\omega,s_2]$ for all ω, and that,

iii) $\sup \pi[\omega,s] \leq \pi^*$ and for each ω, $\pi[\omega,s] \leq (1-\beta)\Phi$ for sufficiently large s.

In iii) β is the discount rate, and Φ is the scrap or exit value (the value of the firm and its entrepreneur in its best alternative use) so $(1-\beta)\Phi$ is the per period return on the firm's transferable assets.

The special case we actually compute is a differentiated product model. Good "0" is the outside good, and goods 1,...,N are the goods produced by the firms competing in the industry. Each consumer purchases at most one good from the industry. The utility consumer "i" derives from purchasing and consuming good "j" is given by

$$U(i,j) = v_j - p^*_j + \epsilon(i,j),$$

where v_j is the quality or efficiency index, and p^*_j is the price, of the good, and i=1,...,M. Consumer "i" chooses good "j" if and only if it is preferred over all the alternatives, that is if

for q=0,1,...,N,

$$\epsilon(i,j) - \epsilon(i,q) \geq [v_q - v_j] + [p^*_j - p^*_q]$$
$$=[v_q - v_0] - [v_j - v_0] + [p^*_j - p^*_0] - [p^*_q - p^*_0]$$
$$= g[\omega_q] - g[\omega_0] + p_j - p_q,$$

where $g(\cdot)$ is increasing, $\omega_q = g^{-1}[v_q - v_0]$, and $p_q = p^*_q - p^*_0$. Let the set $c[\omega_j;p,s]$, where s is the counting measure providing the number of firms at each ω, be the set of ϵ's that satisfy the above set of inequalities, and hence induce the choice of good j.

Note that choices are determined entirely by the firm's efficiency relative to the efficiency of the outside good. So an increase in the firm's efficiency means that its efficiency has gone up relative to the outside alternative, and it's efficiency will decrease only if the improvements to the firm's own product are not as great as the improvements in the outside alternative. Also, movements in v_0 will cause synchronized movements in the relative efficiencies, in the ω's, of all firms in the industry, which in turn will generate a positive correlation in their profits (of course movements in the v_j will generate a negative correlation in the profits of firm "j" and its competitors). Finally, it is the "real" price of the good that matters.

Let $G(\cdot)$ provide the distribution of ϵ. Then the probability that a random consumer will choose good "j" is

$$\sigma[\omega_j;p,s]=\int_{\epsilon\subseteq c[\omega_j;p,s]} dG(\epsilon)=\exp[g(\omega_j)-p_j]/\{1+\Sigma\exp[g(\omega_q)-p_q]\},$$

where the last equality assumes that $G(\cdot)$ is multivariate extreme value. If there are N firms in the market, no fixed costs of production and constant marginal costs equal to mc, then it can be shown that if firms choose prices to maximize profits a unique Nash equilibrium exists (Caplin and Nalebuff,1991) and satisfies the vector of first order conditions

$$-[p_j-\mathrm{mc}]\sigma_j[1-\sigma_j] + \sigma_j = 0$$

for j=1,...,N. Profits are given by

$$\pi[\omega_j,s]=\{p[\omega_j,s]-\mathrm{mc}\}M\sigma[\omega_j;s],$$

where it is understood that the price and share vectors are calculated from the spot market equilibrium conditions.

The distribution of ω_{t+1} conditional on ω_t depends on the amount the firm is willing to invest in developing its product. We let the family of distributions for the increment in ω, i.e. for $\omega_{t+1} - \omega_t = \tau_t$, conditional on different values of x_t, be

$$\mathbb{P} = \{P(\cdot\,|x),\ x\in\mathbb{R}_+\},$$

which we assume to have finite support. This family is built as a convolution of two random variables. The first, say v_1 is the increment in efficiency the firm gets from its own research process, and is stochastically increasing in x. The second, say v, is an exogenous random variable which represents the force of the competition from outside of the industry (the efficiency of the outside alternative in the example above). Note that the possibility of advances by outside competitors imply both that; $P(\cdot)$ puts positive probability on negative values of τ, and that the realizations of τ are not independent across the firms that are active in a given period.

The example used in the computations puts $\tau = v_1 - v$, where,

$$v_1 = \begin{cases} 1 \text{ with probablity } ax/_{(1+ax)} \\ 0 \quad \text{otherwise,} \end{cases}$$

and,

$$v = \begin{cases} 1 \text{ with probability } \delta \\ 0 \text{ otherwise.} \end{cases}$$

By making the time period per decision small relative to the time period in the data we generate distributions of increments that make large changes in τ possible.

To choose optimal investment and exit policies incumbents need also a perceived distribution for

$$\hat{s}_{t+1} = s_{t+1} - e[\omega_{t+1}],$$

where $e[\omega_{t+1}]$ is a vector which puts one in the ω_{t+1} spot and zero elsewhere, conditional on s_t, and ω. This will be denoted by

$$q_\omega\{\hat{s}_{t+1}|s_t\} = \Sigma q_\omega\{\hat{s}_{t+1}|s_t,v_{t+1}\}p\{v_{t+1}\}.$$

Note that this distribution embodies the incumbent's beliefs about entry and exit.

We assume that the functions $q_\omega[\hat{s}|s]$ can be derived as the transition probability for $s_{t+1} - e[\omega_{t+1}]$ from some regular Markov transition kernel, say $Q[\cdot|\cdot]:SXS\to[0,1]$ and that S is compact. Ericson and Pakes show that the Markov–Perfect Nash equilibrium will generate transition kernels with these properties (i.e. that these conditions are indeed satisfied in equilibrium).

Given that $q_\omega(\cdot|s)$ provides the incumbents perceived distribution of future market structures, the Bellman equation for the firm's maximization problem is

$$V(\omega,s) = \max\{\Phi,$$
$$\sup_{(x\geq 0)} [\pi(\omega,s) - cx + \beta \Sigma V(\omega+\tau,\hat{s}+e[\omega+\tau])\, q_{\omega}[\hat{s}\,|\,s,v]\; p(\tau\,|\,x,v)\; p(v)]\}.$$

Ericson and Pakes (1989) provide a reasonably detailed exposition of the nature of optimal policies in this framework. What we require here is the fact that boundedness of the value function implies that if ω is high enough, the value of additional increments in ω can be made as small as we like. Since the return to investment in this model is determined by the increment in the value function generated by higher values of ω, the boundedness assumption insures that investment will be zero for all ω greater than some $\hat{\omega}$. Since firms cannot improve their quality index without some investment, states above $\hat{\omega}$ are "coasting states" from which the firm's ω can only deteriorate (and will stochastically). So there is an upper bound to the achievable ω states. Similarly, the possibility of exit generates a lower bound for the observed ω states. So we can, without loss of generality, take the set $\Omega = \{1,\ldots,K\}$.

To complete the description of the model we need also to specify the primitives which determine entry behavior. We have chosen a very simple model of entry where:

i) entry is sequential from an unlimited pool;

ii) entrants pay a (sunk) setup fee of $x_e(m)$, which is nondecreasing in the number of entrants (m), then obtain a draw from a distribution $P[\omega_0]$, and begin operation in the next period at the ω–location generated by the draw;

iii) each potential entrant enters if the EDV of net cash flow from entry exceeds $x_e(m)$.

Formally, if

$$V^e[s,m] = \beta\, \Sigma V[\omega,\hat{s}+e_{\omega}]q_{m-1}\,[\hat{s}\,|\,s,v]p[\omega_0\,|\,v]p(v)$$

then

$$m_s = \begin{cases} 0 \quad \text{if} \quad V^e[s,1] < x_1^e , & \text{else} \\ \min\{m \in I_+ : \quad x_m^e \le V^e_{s,m}, \; V^e_{s,m+1} < x^e_{m+1}\}, \end{cases}$$

with $\{q_{m-1}[\hat{s}|s,v]\}$ consistent with some Markov transition kernel on a compact set.

Note that the distribution of entering ω's is fixed over time. Thus the "ability" of entrants progresses at the same pace as the "ability" of the outside world (in terms of our example it advances with the ability of the outside alternative). If this were not the case entry would eventually go to zero and stay there. Also in the computational example we set $x_1^e = x^e$, and $x_2^e = \infty$, so the maximum number of entrants in any given period is one (the maximum number of entrants in any time interval depends on the number of decision–making periods in that time interval), and

$$p[\omega_0|v] = \begin{cases} 1 \text{ for } \omega_0 = \omega^*, & \text{if } v = 0 \\ 1 \text{ for } \omega_0 = \omega^* \text{-1} & \text{if } v = 1. \end{cases}$$

Ericson and Pakes show that under these conditions;

i) $\forall s, \; m_s \le m^*$.

ii) $\exists N^*$ s.t $\Sigma s_j \ge N^* \implies m_s = 0$.

Hence $\#S \le K^{N^*+m^*}$, i.e. there is only a finite number of industry structures possible. They also provide a formal proof of the existence of a rational expectations, Markov Perfect, Nash equilibrium under these assumptions.

The industry structures generated by this equilibrium will all be counting measures on Ω with a finite number of firms, i.e.

$$S = \{s=[s_1,...,s_K] : \Sigma s_j \leq N^*+m^*\}.$$

So the heart of the equilibrium is a stochastic process for $\{s_t\}$, defined on $(S^\infty,\underline{S},\underline{P})$. This process is a stationary Markov process, i.e. if $s^t=(s_1,...,s_t)$, then

$$Pr\{s_{t+1}=s' | s^t\} = Pr[s' | s_t] = Q[s'|s],$$

with transition kernel $Q[\cdot|\cdot]$, and initial condition s_0 (assumed in S).

The Ericson–Pakes paper also proves that the Markov kernel, $Q[\cdot|\cdot]$, implied by the model's assumptions generates an ergodic distribution of industry structures. In particular it is shown that;

i) the state space, S, contains a unique, positive recurrent communicating class, say $R\subseteq S$;

ii) $\exists$ a unique probability measure, say μ^*, whose support is R, and which satisfies, $\mu^* Q=\mu^*$.

iii) if $\mu_T[s_0,s]$, gives the fraction of time periods for which $s_t = s$, then $\mu_T[s_0,s]\longrightarrow \mu^*_s$ a.s. uniformly over $s\in S$.

Note that though $(1/T)\Sigma s_t \longrightarrow \mu^*$, s_t itself never settles down. Rather the structure of the industry is in perpetual flux. Depending on the nature of μ^* we can expect the industry to go through periods when output is concentrated in the hands of a small number of large firms, and then, perhaps as a reaction to a sequence of new inventions, to fracture into an industry composed of a large number of approximately equally sized firms. Of course even over periods when the industry structure remains relatively stable there will be heterogeneity in the outcomes of the active firms, with rank reversals, and simultaneous entry and exit as the normal course of affairs.

It is worth emphasizing, however, that the actual nature of the limit distribution, i.e. of μ^* (whether in fact it does include both relatively fractured and relatively concentrated structures), and the nature of the pattern of likely transitions between elements in that limit distribution (do we cycle over the divergent types of structures, or are there Poisson type events that take us more directly from one type of structure to another), depends on the precise values of the parameters that determine the primitives of our model. What the ergodic theorem tells us is that if we are willing to suffice with limit properties, we can analyze them, and how they react to different values of the parameter vector, without specifying initial conditions.

V.2 Computation of Equilibria.

We now come back to the task of tracing out the characteristics of $Q[\cdot|\cdot]$ implied by different values of the parameters defining the primitives of the model, and of describing how sample paths are likely to change in response to policy and environmental changes. This brings us back to the need for a computational algorithm that allows us to solve for (or simulate) the stochastic process generating $\{s_t\}$ for different values of the parameters of the model.

$Q[\cdot|\cdot]$ is calculated from the optimal policy and the primitives, as together, these generate the transition probabilities for all encumbents and potential entrants from any initial state. Though the actual computation may well be complicated (see Ericson and Pakes,1989), given the optimal policies, i.e. $\{\chi(\omega,s),x(\omega,s),V^e(s,m)\}$, and an initial s, it is easy to simulate different sample paths and then derive an empirical distribution which will converge (uniformly) to the true $Q[\cdot|\cdot]$. So the whole computational problem is in finding the optimal policies. We turn now to a description of a computational algorithm [taken from Pakes and McGuire(1990)], designed to find these polices, and a discussion of its properties. The algorithm, together with the functional forms given in the example provided above, is then used to generate and characterize the evolution of that industry.

<u>A Computational Algorithm</u>

We work off value functions for problems with a limited number of active firms, and then push that limit up. Start with the value function for 1 active firm. This is a straightforward contraction which sets the support of (the upper and lower limits for) ω, i.e. it sets Ω.

For N=2 we need to calculate $V[\omega_1,\omega_2]$ for $\{\omega_1,\omega_2\} \in \Omega^2$. Start with $V_0[\ ,\ ] = \pi[\omega_1,\omega_2]$. Then get $V^n[\omega_1,\omega_2]$ as the solution to:

$$\text{(30a)} \qquad V^n[\omega_1,\omega_2]=\pi[\omega_1,\omega_2]-cx+ \beta\Sigma_{\tau_i,\tau_j,v}V^{n-1}[\omega_1+\tau_1+v,\omega_2+\tau_2+v]p[\tau_1|x_1,v]p[\tau_2|x_{2,}v]p(v),$$

where the couple $[x_1,x_2]$ satisfy the Kuhn–Tucker conditions

$$\text{(30b)} \qquad x_i\{-c + \beta\Sigma_{\tau_i,\tau_j,v}V^{n-1}[\cdot,\cdot]p_j(\partial p_i/\partial x_i)p(v)\}=0,$$
$$x_i \geq 0,$$

for $i \neq j$, $i,j = 1,2,\ldots,p_i = p[\tau_i|x_i,v]$, and so on. One procedure for calculating the fixed point defining the equilibrium would be to repeat this step iteratively until $\|V^n-V^{n-1}\|$ was below an acceptable tolerance for an appropriate norm $\|\cdot\|$.

Note that this procedure differs from a straightforward "doubly nested" fixed point calculation. The latter would begin with a candidate function for the process generating the competitors states [for $p(\tau_2|s,\nu)$ in our two firm example], use it to solve the implied contraction for $V[\cdot|p(\tau_2|\cdot)]$ and the associated investment policy, then use that investment policy to update the process generating the competitors states [to update $p(\tau_2|s,\nu)$], and then iterate on this double nest until convergence. By solving for all the x's simultaneously at each iteration we have done away with one of the "nests" in this fixed point algorithm. However, the procedure in (30) requires the solution to an implicit nonlinear system of equations at each iteration of the fixed point, and the fact that the system of equations does not have a explicit solution increases computational time dramatically.

An alternative is to use x_j^{n-1} in the Kuhn–Tucker condition that solves for x_i (for $j \neq i$). If we ignore the constraint that $x \geq 0$, this gives an explicit solution for each x_i. If $x_i \leq 0$, set it equal to zero. This does away with the requirement of a nonlinear search at each iteration, and decreases computational time accordingly. Of course we could iterate on this "policy step" until convergence and (provided convergence is achieved) obtain an exact solution without ever having to simultaneously solve the nonlinear system. Though it did not prove helpful to iterate on this step in the examples presented here, it is more likely to prove helpful once we substitute polynomial approximations into the computational algorithm (see below).

For expositional simplicity we ignored entry in the discussion above. To account for entry, consider states where $V[\omega_2,\omega_1] = \Phi$. In such cases we use $\pi[\omega_1]$ instead of $\pi[\omega_1,\omega_2]$, and calculate $V^e[\omega_1]$. If this term is greater than x^e, we calculate $V[\omega_1,\omega_0]$ for $V[\omega_1,\omega_2]$, where the transition probabilities for ω_0 are given by $P[\omega_0]$.

We still continue iterating until $\|V^n - V^{n-1}\|$ is within a given tolerance, but now we also check $\|x^n - x^{n-1}\|$. A fixed point to this problem can be shown to satisfy all the requirements for our equilibrium if the maximum number of firms, set either endogenously by the model's parameters, or by an artificial barrier to entry, call it N, is 2. Now push N up to 3 and do the fixed point calculation again starting at

$$V^{n=0,3}[\omega_1,\omega_2,\omega_3] = V^{n=\infty,2}\{\omega_1,\max[\omega_2,\omega_3]\}.$$

This procedure should be repeated until we reach an n where, for all industry structures at which $\Sigma s(i)=n$, $V^e(s,1)<x^e$. The n which satisfies this condition is an upper

bound to the number of firms ever in the industry (i.e. it is N; of course this is only true if the initial s has no more than N firms).

Some caveats are in order before proceeding. We have not proved a contraction property for our algorithm yet, so we have no way of knowing for sure whether it converges. On the other hand it has converged for every set of parameter vectors we have tried. We have not proven uniqueness either. We did calculate all results we describe here thrice, starting each at different initial conditions [once at A(·), once at zero, and once at V(·) from the smaller N, see above]. In each case we got to precisely same answer.

Computational Burden

Roughly, the computational burden of this algorithm is the product of: 1) time per grid point evaluation, 2) number of grid points evaluated at each iteration, 3) number of iterations until convergence.

An explicit calculation can be provided for the number of grid points. First note that one does not have to evaluate all of them since symmetry implies that $V[\omega,1,k] = V[\omega,k,1]$, for all $(\omega,k,1)$. Indeed the number of points one needs to evaluate are the number of distinct N–element vectors with $\omega_1 \geq \omega_2,\ldots, \geq \omega_N$, (and, $1 \leq \omega_i \leq K$, for $i=1,\ldots,N$). Lemma 31 provides an exact calculation for the number of distinct N element vectors that satisfy this condition. Note that it grows as a polynomial in N (it is bounded from below by K^N/N, and from above by $(K+N-1)^N/N$).

31.Lemma

The number of distinct sequences $[\omega_1,\ldots,\omega_n]$ with $\omega_i \geq \omega_{i-1}$ and $\omega_i \in [1,\ldots,K]$ $(i=1,\ldots,N)$, say $E(K,N]$, is given by

$$E[K,N] = \binom{K+N-1}{N} = \frac{(K+N-1)!}{(K-1)!N!}$$

Proof

First note that for $N \geq 2$

$$\mathcal{E}[K,n] = \sum_{j=1}^{K} \mathcal{E}[j,N-1]$$

since when we put the number "j" in the N^{th} slot we have $\mathcal{E}[j,N-1]$ sequences of $[\omega_1,\ldots,\omega_{N-1}]$ with $\omega_i \geq \omega_{i-1}$ for i = 1,....,N−1 and $\omega_i \in [1,\ldots,K]$. Given this fact we can proceed with an inductive proof for the theorem (in N).. The initial condition of the inductive argument [N=2] is true by enumeration, so what we need to show is that if $\mathcal{E}[K,N]$ satisfies the equation in the statement of the lemma so does $\mathcal{E}[K,N+1]$. From above

$$\mathcal{E}[K,N+1] = \sum_{j=1}^{K} \mathcal{E}[j,N] = \sum_{j=1}^{K} \binom{j+N-1}{N} = \binom{K+N}{N+1}$$

as required, where the last equality can be shown by induction on K. o

The approximation methods we discuss below are designed to overcome the problem that the number of grid points evaluated grows as a polynomial in the least upper bound to the number of active firms. There will still be the issue, however, that the time per grid point evaluation grows as a similar (though not quite so large) polynomial in N (it would be exactly the same if all industry structures were connected in the sense that it was possible to pass from any one to any other in a single period; though this is definitely not the case in our example). That is at each grid point we do evaluate, we need to evaluate the value function at every achievable industry structure in the following period.

On the other hand, we expect the number of iterations needed for convergence to go down as N goes up. As we increase N, the effect of an additional active firm on the value of being at a particular point ought to diminish, so the final iteration for the value function calculated at the N–1 firm equilibrium should be closer to the N–firm value function we are looking for. However, we have no formula for the rate at which this will occur.

We calculated 2 to 6 firm equilibria (i.e. $N^*=2,...,6$) for different values of the parameter vector for our example, and found that the computational time for the 6 firm equilibria was about 5.5 hours on our sparc station. The no–entry barrier equilibria value of N^* for most of these runs was 6 firms. However the time required to calculate the equilibria went up by a factor of about 5 every time we went from an N to an N+1 firm equilibria. This in spite of the fact that the number of iterations required before our convergence criteria was met typically got multiplied by fractions between .5 and .7 when we moved up N in units of one (though this varied between runs).

Thus, though the computational techniques presented here may suffice for computing equilibria for markets with a small number of agents, we will need to improve on them in order to analyze many of the markets of interest. We come back to this point below. First we glance at some of the summary statistics from one run of our example.

<u>Descriptive Output From One Set of Computations</u>

To illustrate the type of dynamic stochastic equilibrium that results from this class of dynamic heterogeneous agent models, we briefly go over some summary statistics from one set of parameter values. Those values are: δ (the probability that the outside alternative moves up) = .7; $\beta = .925$, x^e (sunk entry cost) = .2, Φ (scrap value) = .1, m (size of market) = 5, spread = 3 and a = 3 (parameters determining the efficacy of own investment in increasing the probability of quality improvements), mc (marginal cost) = 5.

Table 2 provides some statistics which help describe the ergodic distribution for this industry. Part A of the table indicates that the ergodic process characteristically has either three or four firms active in a given period. There is, however, lots of entry and exit, so the firms active in equilibrium are not always the same three or four firms. Note also that entry and exit are positively correlated; in most years when there is entry there is also

Table 2

Characteristics of Ergodic Distribution

$\delta = .7$ $\alpha = 3$ $\beta = .925$ xe = .2 phi = .1
spread = 3 m = 5 c = 5

No. of Time Periods	10,000
% with 6 firms active	.3
% with 5 firms active	1.9
% with 4 firms active	27.9
% with 3 firms active	69.9
% with 2 firms active	.0
% with entry and exit	10.10
% with entry only	4.44
% with exit only	2.18
% with entry or exit	16.75

exit (Part B of the table: this is in a stark contrast to models of industry dynamics that do not allow for idiosyncratic sources of change).

One thousand four hundred and fifty one firms participated in the industry during the 10,000 periods simulated, however, most were active only a short period of time (Table 3; part A). Almost half of them dropped out after their first year of operation. Both mortality and hazard rates decline markedly over the first seven or eight years, giving the indication that this initial period looks very much like a "learning" period. About 11% of new entrants survive eight years, and after that the hazard has no particular shape (one should be aware that these are estimated mortality rates; their standard errors are on the order of .005).

Part B of Table 3 provides characteristics of the realized values of the firms which participated. The first point to note is that over 90% of the firms which participated in this industry had a net loss from their endeavor (generated negative realized values). Most lose about .1 (the difference between the entry and exit fees), but there are those who invest for awhile, never move up the "quality ladder", and eventually drop out, losing also their investments in the interim. Among the 10% whose realized values were positive, the mean realized value was very high (9.3 giving a benefit/cost ratio of 46.5), and the distribution was very skewed. The industry is most often reasonably fractured (the one firm concentration ratio averaged .37 in an industry in which there are almost always either three or four active firms), but periodically a firm will surge ahead of its competitors and stay there for reasonable lengths of time (the standard deviation of the one firm concentration ratio was .11).

These parameter values generate an industry in which it is relatively cheap to start up and explore some new idea. Most start ups are not successful. The few that are grow to become major actors in the industry, and earn phenomenal rates of profit. Of course, eventually, even the most profitable firms are passed over by the developments of its competitors and find it optimal to exit.

Table 3A

Lifetime Distribution

(Based on 1,451 "lives" in 10,000 time periods)

δ = .7 α = 3 β = .925 xe = .2 phi = .1
spread = 3 m = 5 c = 5

Mean = 22.7, Median = 2, Standard Deviation = 101.8

Lifetime	Frequency	Percent	Implied Hazard	Cumulative Percent
1	617	42.5	42.5	44.5
2	401	27.6	48	70.2
3	126	8.7	29.2	78.8
4	55	3.8	17.9	82.6
5	36	2.5	14.3	85.1
6	20	1.4	9.4	86.5
7	16	1.1	8.2	87.6
8	14	1.0		88.6
9	7	.5		89.0
10	5	.3		89.4
≥ 10	146	10.06		
≥ 50	96	6.6		
≥ 100	77	5.32		

Table 3B

Realized Value Distribution

δ = .7 α = 3 β = .925 xe = .2 phi = .1
spread = 3 m = 5 c = 5

Mean = .58, Median = −.1, Standard Deviation = 3.60

128 positive entries, mean is 9.3
1223 negative entries, mean is .28

Obs/Num	RV	Lifetime	Sum/Rv
1	72.8	79	72.8
2	24.0	70	96.9
3	22.2	182	119.1
4	21.9	501	141.0
5	20.2	530	161.2
10	17.8	54	253.4
50	10.44	78	779.4
100	4.39	334	1135.9
128	.21	9	1192.2
129	−.06	4	1191.9
1000	−.1		
1451	−4.03		846.7

At this point it would be useful to perturb the model in ways that correspond to possible policy or environmental changes, do additional computations, and compare the results. One of the great advantages of structural modelling is that it generates an ability to do such comparisons, and Pakes and McGuire (1990) illustrate by considering the effects of alternative possible regulatory changes on market structure and welfare (see also Judd's, 1990, numerical analysis of alternative duopolies). However, these comparisons are topics for whole different papers, so we now return to computational issues.

V.3 Computational Approximation

We now consider computational techniques that attempt to reduce the computational burden of obtaining the equilibrium by fitting the value function at only a small fraction of the points in S, and then using the information obtained from those values to predict the value function at other points as needed. More generally, all we require is an approximation to a function which determines policies at any point in S, and there are many different ways of doing this. The symposium in the JEBS (1990) reviews and compares several different approximating techniques in the context of computing equilibria for a representative agent stochastic growth model. Judd (1990), sketches a general framework and computes equilibria from models with two agents (no entry or exit), and the article by Marcet (1990) in this volume reviews progress in this field to date.

Many of these techniques fit polynomials in a set of functions that span, or form a basis for, a "rich enough" collection of approximating functions (the Chebyshev or Legendre polynomials for example) to a small set of points, and then use the fitted polynomial to predict the other points as needed. An alternative is to fit the function directly at a small number of points, and then interpolate, either linearly, or using a spline, to other points. We begin by showing how to embed such approximations into the computational algorithm described above. Note that the heterogeneous agent problems we are interested in are by nature multidimensional, the dimensionality of the state vector for any given agent going up with the number of other agents active in the market.[22]

Recall that a function $f:\Omega^N \to \mathbb{R}$ is a polynomial of order λ if for all $\omega\in\Omega^N$

$$f(\omega_1,\dots,\omega_N) = \sum_{p=0}^{\lambda} \sum_{h_N=0}^{p-\Sigma h_i} \cdots \sum_{h_1=0}^{p} \alpha(h_1,\dots,h_N)\omega_1^{h_1}\dots\omega_N^{h_N}$$
$$\equiv \sum_{h\in H^N} \alpha(h)\omega(h)$$

with $\alpha(h)\in\mathbb{R}$ for all $h=\{h_1,\dots,h_N\}\in H^N$, where $H^N=\{h\in \mathbb{Z}^N{}_+ : \Sigma h_i \leq \lambda\}$. The collection of all such polynomials (obtained by varying α), together with the usual operations of addition and scalar multiplication, is a vector space (over the real numbers), say $\mathscr{V}_\lambda$. A basis for this vector space is the set of tensor products of the $\omega_i^{h(i)}$ with h varied over H^N (see Hoffman and Kunze,1972,section 5.6). These are just the functions implicit in the $\omega(h)$ in the equation above. Though we do not pursue it here, the following discussion could be generalized by looking for an approximation in a vector space spanned by the tensor products of $g(\omega_i)$ for suitably chosen $g(\cdot)$.

The iterative procedure used to calculate the fixed point defining the value function for our problem can be modified to find an approximating polynomial, a $\hat{V}\in \mathscr{V}_\lambda$ as follows. Define a set of basis points, say $\omega(j)\in\Omega^N$, for j=1,2,.... If there are J basis functions, the basis points must generate at least J linearly independent values for those functions. Starting at some initial guess for the vector α, let the estimate of the coefficients at the $n-1^{th}$– iteration of the recursive calculation be α^{n-1}. Now calculate the value function at the basis points by substituting

[22]Throughout we will consider the case where N, the least upper bound to the number of agents ever active in a given period, is less than or equal to $K = \#\Omega$. In this case the dimensionality of the state vector is smaller when we calculate value functions as a function of the vector of ω values of all active agents. When $N \geq K$, use of a counting measure on Ω as the state vector minimizes the dimensionality of the state space.

$$\hat{V}^{n-1}[\omega(j)] = \omega(j)'\alpha^{n-1},$$

into the n^{th} iteration of the recursive calculation in (30). (30b) then produces an x^n, which when substituted back into the (30a) that used $\hat{V}^{n-1}$ produces a new value function, say $V^{*n}()$, at each of the basis points. We choose α^n to minimize the Euclidean distance between $\omega'\alpha^n$ and $V^{*n}()$ at the basis points. That is, if W is the matrix formed from the rows $\omega(j)$,

$$\alpha^n = [W'W]^{-1}W'V^{*n\prime}.$$

This procedure can be generalized slightly by approximating a monotone function of $V(\cdot)$ by a polynomial in the basis functions, instead of approximating $V(\cdot)$ itself.

Without further restrictions the number of functions needed to form a basis for $\mathcal{V}_\lambda$, and hence the minimum number of points at which we need to fit the value function for this approximation, still grows as a polynomial in N. However, we have not yet used the fact that the value function is symmetric, more precisely exchangeable, in the vector $(\omega_2,\ldots,\omega_N)$. If we restrict our search to the subspace of $\mathcal{V}_\lambda$ that satisfy the restriction that, for all $\omega^N \epsilon \Omega^N$,

$$\hat{V}(\omega_1,\ldots,\omega_N) = \hat{V}(\omega_1,\pi_2,\ldots,\pi_N), \qquad (32),$$

for any N–1 dimensional vector $\pi = (\pi_2,\ldots,\pi_N)$ which is a permutation of $(\omega_2,\ldots,\omega_N)$, we reduce the number of required basis functions dramatically. Indeed, provided $N \geq \lambda$, the number of required basis functions becomes <u>independent of N</u>. That is the content of the following theorem.

<u>33. Theorem</u>

The space of polynomials of order λ satisfying equation 32, together with the usual operations of addition and scalar multiplication, is a vector space, say $\mathcal{V}^{\mathcal{E}}_{\lambda} \subseteq \mathcal{V}_{\lambda}$, with dimension

$$\dim \mathcal{V}^{\mathcal{E}}_{\lambda} \leq \Sigma_{p=0}^{N} \Sigma_{i=0}^{p} \Delta(i) = \varphi(\lambda), \qquad (33.1)$$

where $\Delta(i)$ is the number of partitions of the number i (see below). Further, 33.1 holds with equality if $N \geq \lambda$. Note that $\varphi(\lambda)$ is independent of N.

Proof

The fact that addition and scalar multiplication preserves partial exchangeability proves that the subspace of functions satisfying (32) is a vector space. The proof of 33.1 is a result of the following lemma.

33.2 Lemma (proved in Appendix 2).

An $f \in \mathcal{V}_{\lambda}$ is also a member of $\mathcal{V}_{\lambda}^{\mathcal{E}}$, if and only if for all $h \in H^N$,

$$\alpha_f(h_1, h_2, \ldots, h_N) = \alpha_f(h_1, \pi_2, \ldots, \pi_N),$$

for any $(\pi_2, \ldots, \pi_N)$ which is a permutation of $(h_2, \ldots, h_N)$.

Define $m_j(h)$ to be the j^{th} largest element in the vector $(h_2, \ldots, h_N)$ for $j=1,\ldots,N-1$ (using any tiebreaking rule that preserves the natural order of pairs that are ordered). Then lemma 4 implies that we can form a basis for $\mathcal{V}_{\lambda}^{\mathcal{E}}$ by simply adding together the basis functions from $\mathcal{V}_{\lambda}$ that have

$$\alpha_f(h_1, \ldots, h_N) = \alpha_f(h_1, m_2, \ldots, m_N),$$

for each distinct value of the vector $(h_1, m_2, \ldots, m_N)$. What remains is to determine the number of distinct α coefficients this generates. Let p(h) be the order of the basis function corresponding to $\alpha(h)$, that is $p(h) = \Sigma h(i)$. Then the number of distinct α coefficients generated by h vectors with $p(h) = p$, and a particular value of h_1, is the number of ways the number $p-h_1$ can be allocated among N−1 locations (without regard to order). If $N \geq \lambda \geq p-h$, this is simply the number of partitions of $p-h_1$, or $\Delta(p-h_1)$ (see below). Consequently, the number of distinct α coefficients required to generate all distinct coefficients for the p^{th} order basis functions is $\Psi(p)$, where

$$\Psi(p) = \sum_{i=0}^{p} \Delta(i)$$

$\varphi(\lambda)$ is derived by summing this equation over $p=0,1,\ldots,\lambda$.

The theorem implies that there are only two distinct first order coefficients

$$\alpha(1,0,\ldots,0), \quad \text{and} \quad \alpha(0,1,0,\ldots,0)$$

with associated basis functions

$$\omega_1, \quad \text{and} \quad \Sigma_{i=2}\ \omega_i.$$

Similarly, there are four distinct second order coefficients

$$\alpha(2,0,\ldots),\ \alpha(1,1,0,\ldots),\ \alpha(0,1,1,\ldots), \text{ and } \alpha(0,2,0,\ldots)$$

with basis functions

$$\omega_1^2,\quad \omega_1\Sigma_{i=2}\,\omega_i,\quad \Sigma_{i_1=2}\Sigma_{i_2=2}\omega_{i_1}\omega_{i_2},\ \text{ and }\ \Sigma_{i=2}\omega_i^2.$$

More generally, there are $\Delta(p-j)$ p^{th} order coefficients with $h_1 = j$,

$$\alpha(j,p-j,0,\ldots),\ \alpha(j,p-j-1,1,0,\ldots),\ \alpha(j,p-j-2,2,0,0,\ldots),$$

$$\alpha(j,p-j-2,1,1,0,\ldots),\ldots,\ \alpha(j,1,1,\ldots,1,0,\ldots),$$

with associated basis functions

$$\omega_1^j \sum_{i=2}^{N} \omega_i^{h-j},\quad \omega_1^j \sum_{i_1=2}^{N} \sum_{i_2=2}^{N} \omega_{i_1}^{h-j-1}\omega_{i_2},\quad \omega_1^j \sum_{i_1=2} \sum_{i_2=2} \omega_{i_1}^{h-j-2}\,\omega_{i_2}^2,$$

$$\omega_1^j \sum_{i_1=2} \sum_{i_2=2} \sum_{i_3=2} \omega_{i_1}^{p-j-2}\omega_{i_2}\omega_{i_3},\ldots,\quad \omega_1^j\Sigma_{i_1}\ldots\Sigma_{i_{p-j}}\omega_{i_1},\ldots,\omega_{i_{p-j}}$$

The general formula for $\Delta(q)$ requires fairly detailed notation (see, for eg., Abramowitz and Stegum, 1972, p. 825; it is a sum of Stirling numbers of the second kind). For convenience, we provide a listing of $\Delta(q)$ and $\varphi(q)$ for q=1,...,12, in Table 4.

Table 4

q	0	1	2	3	4	5	6	7	8	9	10	11	12
$\Delta(q)$	1	2	3	4	5	7	11	15	21	30	41	55	75
$\varphi(q)$	1	3	7	14	26	45	75	120	186	276	407	593	854

Recall that if a λ–order polynomial is a good approximation to the value function, then we need only calculate the value function at $\varphi(\lambda)$ points. For comparison, the pointwise technique used to calculate the results reported earlier required calculating the value function at 639,000 points; and this for a vector of parameters that generated an ergodic distribution of industry structures with an upper bound of six active firms. Thus, at least for industries with a moderate number of firms, polynomial approximations restricted to the subspace of exchangeable polynomials should allow us to cut the number of points at which we evaluate the value function by several orders of magnitude.

The other point to remember is that the CPU time required to compute the value function is a product of: the number of points evaluated at each iteration, the time per point evaluated, and the number of iterations required before convergence. Though the number of points evaluated will fall dramatically as a result of imposing the restriction that $\hat{V} \in \mathcal{V}_{\lambda}^{\mathcal{E}}$, the complexity of the calculations at each point evaluated will increase. The reason is that at each point we require the integral of the value function over the states achievable from that point in the next period, and the values of the value function required for the integrand in this computation must now be computed as a product of basis functions and polynomial coefficients (instead of just calling them up from memory, which is what is done when we calculate the value function pointwise). On this count alone, then, we would not expect substitution of the approximation technique to cut computational time by the same factor as it cuts the number of points at which we need to evaluate the value function. In addition, substitution of the approximating technique is likely to change the number of iterations needed before convergence is achieved (though it is not clear in which direction this change will go). Thus, the crucial question of just how much of a saving in CPU time we will generate by approximating the value function by a $\hat{\mathcal{V}} \in \mathcal{V}_{\lambda}^{\mathcal{E}}$ is still unresolved, and all we can say at this stage is that this form of approximation <u>may</u>

enable us to calculate the equilibria in problems for which the number of active agents is quite large.

To begin our examination of the use of a $\hat{V} \in \mathcal{V}_\lambda^{\mathcal{E}}$ to approximate the value function, we fit the approximating basis to the actual value function for our example (recall, that we obtained the value function from an "exact" pointwise calculation). We started here for two related reasons. First, we thought that if the polynomial coefficients obtained by fitting the approximating basis to the true numbers did not provide an adequate approximation to the value function, then we could not expect that the polynomial coefficients obtained by fitting the approximating technique into our computational algorithm to provide an adequate approximation. Second, there are several ways of modifying the procedure used to obtain the polynomial approximation to the value function, and one simple way of comparing the alternatives is to compare how well they do in approximating the true numbers.

In this latter context we mention four points. First, since the sum of (partially) exchangeable functions is an exchangeable function, one can add any exchangeable function to an exchangeable basis and still maintain the exchangeability of the approximating function. This is one way of embodying exogenous information into the approximation algorithm, and we illustrate below by adding the profit function to the basis used for approximating the value function with quite dramatic results. Second, we have proceeded throughout as though the basis were being fit directly to the value function. Instead, we could fit the basis to any monotone transformation of the value function, and modify the computational algorithm accordingly. The Ericson–Pakes paper proves that for some simple cases of their model the value function is "S–shaped" in the firm's own ω, and it is presumed that this general shape characteristic persists for a larger class of primitives, including the primitives used in our calculations. So we present results from fitting the logit transform of the value function, as well as the value function per se. Third, it is possible to use different degrees of polynomial approximation for the ω_1 dimension, then for the $(\omega_2,\ldots,\omega_N)$ dimension, and, finally, we need not restrict ourselves to fitting $\varphi(\lambda)$ points [any number greater than $\varphi(\lambda)$ will do].

Table 5 present some results from the fitting exercise. The entries in the table are the R^2's obtained from OLS fits of the value function to alternative approximations. For most of the approximations we present the fits from the value function when the number of active firms is restricted to be no more than 4, and 5 as well as for the unrestricted case (where the least upper bound on the number of active firms in the ergodic distribution is 6). We also present most results as the order of the polynomial being used in the basis functions varies from 2 to 6 (this gives us the alternative rows of the table).

The columns labeled V provide the R^2's from fitting the value function to the partially exchangeable basis of polynomials. The columns labelled LOV fit the logit transform of the value function, but then transform back to the actual numbers to calculate the fit. The columns labelled A fit the actual value functions, but add the profit function to the set of basis functions. The columns labeled A2 add both the profit function and an interaction of the profit function with the first order polynomials to the basis functions. The numbers above all these columns refers to the number of points at which we obtain values for the value function, or the cardinality of S (this is the number of observations for the OLS regressions).

The entries for the columns labeled mod3 and mod3A are found in a slightly different way. Here we took only the value function at those ω points that were mod3 in the vector sense (i.e. each element of the ω vector was divisible by 3) and projected these on to the basis functions to obtain the polynomial coefficients. We then use the polynomial coefficients obtained in this way, to predict the value function at all points, and calculate the R^2 obtained from fitting the true values to these predicted values. The "number of points" headings above these columns refer to the number of points used in the first stage of this procedure (the number of mod3 points). The column labeled mod3A adds the profit function to the basis used in the first stage. Finally, in subpanel 6 we also present results

Table 5 - R^2 's for Alternative Approximations*

Firms	4						5				
# of points	25270 mod3 = 588						138958 mod3 = 1470				
Order	V	LOV	A	A2	mod3	mod3a	V	A	A2	mod3	mod3a
2	.804	.905	.977	.986	.802	.976	.807	.971	.982	.804	.970
3	.877	.911	.987	.992	.873	.985	.880	.985	.991	.877	.982
4	.927	.947	.993	.995	.919	.992	.928	.991	.994	.922	.990
5	.951	.974	.995	.996	.939	.994	.953	.994	.996	.941	.993
6	.968	.977	.996	.997	.942	.995	.968	.996	.997	.943	.994

Firms	6							Interpolation-6 firms		
# of points	639331 mod3 = 3,235 ergodic = 2,485							3234	8778	1064
Order	V	erg	ergA	ergA2	erg-mod3	erg-mod3A	erg-mod3A2	mod3	mod3 (2-6)	mod6 (2-6)
2	.815	.834	.971	.988	-	-	-	.989	.993	.902
3	.887	.908	.986	.994	-	-	-			
4	.932	.957	.993	.996	.865	.979	.990			
5	.954	.978	.996	.998	.917	.987	.993			
6	.969	.988	.998	.999	.942	.993	.993			

* For the approximating technique relevant for the alternative columns, see the explanation in the text.

from fitting only the 2485 points in the ergodic distribution of industry structures.

Several points stand out from the table. First, as expected, the same order of polynomials (and hence approximately the same number of basis functions) produce about the same fit regardless of N, or the number of firms ever active in equilibrium (at least if fit is measured by R^2). Second, in comparing the alternative ways of approximating the value function, it seems that using the logit transform only improves the fit marginally (at least when the fit is already quite good), but adding the profit function to the set of basis functions improves the fit rather dramatically. When one has exogenous information on either the form of the value function, or on an alternative function which is expected to "mimic" the properties of the value function, one should probably use it directly.

Third, and probably most importantly, when we fit the exchangeable basis to a small number of (reasonably diffuse) points (the mod3 points), and then use the coefficients obtained from that fit to predict the value function at all possible points, we seem to do just about as well as we do when we fit the basis to the entire set of points directly — at least if the polynomial basis is rich enough to give a good direct fit.

We now move on to examine how well we fit the 2485 points in the ergodic distribution. Since they are less than 1% of the total points being fit in the six firm equilibria, we were worried that the fit of the points in the ergodic distribution (weighted by there probability in the invariant measure) might not be similar to the overall fit of the points in the six firm equilibria (and it is the ergodic points that we want accurate estimates of for most subsequent analysis). The columns labelled erg,ergA, and ergA2,fit the ergodic points directly. The columns labelled ergmod3, ergmod3A, and ergmod3A2, take the polynomial coefficients obtained from fitting the set of mod3 points from the entire six firm equilibria, and use those to predict the points in the ergodic distribution. If anything we seem to fit the points with positive probability in the ergodic distribution better than we fit the entire space of points (even in cases where we use the coefficients predicted from the entire set of points). This gives some reason to believe that the points at which our approximation is not fitting well are points which would not be used intensively in policy and descriptive simulations.

The last subpanel of the table presents some results from fitting interpolated values of the value function. The points from which we interpolate are, respectively in the three columns: all mod3 points, mod1 for the ω of the firm in question and mod3 for the other firms' ω's, and mod1 for the ω of the firm in question and mod6 for the other firm's ω's . It seems that in order to obtain the same fit as obtained from the polynomial approximations, the interpolation procedure requires a larger number of interpolation points than either the number of basis functions required to achieve this fit in the polynomial approximations, or the number of points we used to obtain the polynomial coefficients.

Since the value function per se is not what we are interested in, we also did some limited experiments on whether the investment strategies implied by these approximations were sufficiently close to the investment strategies calculated from the pointwise solution. To do this we simply substituted the approximations into (30b) and calculated the implied investment strategies. We used three measures of fit. The first was the R^2 obtained from comparing the two investment strategies. The second separately substituted the alternative investment strategies into the simulation program used to compute ω_{t+1} and then computed the R^2 from comparing the two ω_{t+1} series, and the third did the same but computed the R^2 from the two series for $\omega_{t+1}-\omega_t$. The three R^2's were, respectively, .98, .99, and .91. These were obtained using polynomial approximations made directly to the ergodic points, and if we use instead the polynomial coefficients obtained by fitting polynomials to the mod3 points from the entire 6 firm equilibrium, and then fit to the ergodic distribution, the results are somewhat worse; .91, .98, and .78, respectively. Still, an R^2 of .8 for first differences, and of .98 for levels seems reasonable, and we could do better by increasing the order of the polynomial we fit.

For a "first cut" we view these results as encouraging. Still they do leave two unanswered questions. First, will we get as good an approximation if we obtain the approximating functions directly from the recursive algorithm described above? Second, are fits as good as those shown in the table "good enough" for either estimation, or for descriptive and policy simulations. At this stage all we can say is that there is work in progress which should help to clarify these points, at least for models similar to those discussed in this paper.

VI. Conclusion

This paper has attempted to clarify some of the modelling, econometric, and computational issues that arise in bringing dynamic structural models into empirical use. The discussion focussed on selected technical issues that have been of concern in applied work; the uses and limitations of Euler equations, incorporating serially correlated unobservables into our models, and computing equilibrium responses to dynamic heterogeneous agent models. Throughout we used examples to illustrate the main points. The exposition of the examples also carried with it an implicit view of structural modelling — so much so that it did not seem necessary to add a section with a more general discussion of when, why, and how, one might engage in it.

It might, however, be useful to conclude with some practical points that often get lost in the more abstract debates on the methods and merits of structural modelling. Our discussion of these points will be premissed on the following "fact". We, as applied researchers, attach a "structural" interpretation to the numbers we eke out of our data every time we use those numbers to analyze the interactions between economic agents, or between an agent and his or her environment. This is just as true when the numbers we use in the analysis are simple "reduced form" correlations, as it is when the numbers used in the analysis are parameter estimates from a complex structural model. So there is really no room for debate on the issue of whether structural models are "useful". The debate must, therefore, be about whether the cost of formalizing the structural models being implicitly used in the analysis, and then possibly parameterizing them with the data, is worth the benefits from this (sometimes quite costly and time consuming) endeavor.

The answer to this question is undoubtedly that sometimes it is worthwhile, and sometimes it is not; and when it is, it is to varying degrees. The cost benefit calculation depends on a myriad of factors including; the complexity of the problem being analyzed and the possibilities for drawing misleading conclusions from simpler forms of analysis, the quality of the data, computational difficulties, prior knowledge on the likely appropriateness of the assumptions that need to be fed into the structural model, and the comparative advantages of the researcher. General rules are hard to come by when so many of the important dimensions of the decision are problem specific. There are, however, a few considerations that one might keep in mind in formulating one's own strategy.

First, it is often useful to begin an empirical project with a reasonably detailed "reduced form" analysis of the data. This for several reasons. First it is likely to suggest just why a more detailed structural model might be useful. Second, the reduced form analysis should indicate the aspects of reality that will need to be built into a structural model for that model to be able to account for the data, and, finaly, thoughtful reduced form analysis often allows one to get some feel for the likely benefits from a structural modelling effort.[23]

Second, given a set of reduced form results, it is often helpful to write down a simple structural model that captures the essence of what one thinks might lie behind them regardless of whether one intends to take that model to data (note that by this we mean writing down a model all of whose assumptions, including the assumptions on its disturbances, are formulated entirely in terms of the primitives affecting economic behavior). The understanding that comes from this modelling exercise is typically useful in several ways. First, it clarifies the problems associated with placing any given interpretation on the reduced form estimates. Second, it crystalizes the trade off between assumptions, computational problems, and data requirements that will be faced in attempting to build a structural estimation algorithm. Finally the modelling exercise will frequently lead to simple diagnostic tests and/or correction procedures for problems that

[23]Unless one has large amounts of data and is extremely careful about how it is used, the preliminary reduced from analysis will often also call into question the interpretation of subsequent standard errors and test statistics. However, without a more detailed theory of learning that allows for the mix of inductive and deductive reasoning actually used in empirical work, this seems like a cost we will just have to bear.

seem likely to be important in interpreting a particular relationship; corrections that might not require either all the assumptions, or the computational burden, that would be needed in order to estimate a complete structural model. The developments in semiparametric estimation (see the references in the text) are particularly exciting in this context. Semiparametric techniques often allow one to circumvent many of the computational issues and some of the more detailed assumptions that would need to be adressed had we to solve the complete model for alternative possible values of its parameter vector before we engage in any estimation.

Third, given the complexity of the issues we typically want to analyze, and the limited data and computational resources available, any successful effort at structural modelling is going to have to abstract from certain aspects of reality. The choices of what to abstract from, and the issue of how that abstraction impacts on what we can learn from our estimates, are both legitimate topics for discussion. In engaging in such discussion, however, a few general points should be kept in mind.

First, given the limitations of our data sets and computational procedures, it seems reasonably clear that one should lean heavily on any prior knowledge available about the applied problem at hand (and there is frequently quite a bit of it available). This, in turn, is going to make the modelling problem more complex; there will be no single framework that is likely to abstract from just the "right" features of reality for a multitude of problems, so that modelling flexibility is going to be required.

Second, even given diligent prior work we are unlikely to come down to exactly the "right" model. The reason we engage in structural modelling in spite of this fact is the belief that there is continuity in the map between the assumptions and the implications of interest, so that the more we know about our problem and the better we are able to incorporate that knowledge into our model, the closer our model will be to mimicking reality.

The fact that structural models cannot be rich enough to encompass all aspects of reality does, however, make it easy to pick them apart. All of the assumptions used are laid out in front of the reader, so it is easy to find transgressions from reality, and these transgressions are frequently large enough to be picked up in formal test statistics. Again, what has to be kept in mind when evaluating a structural modelling exercise is that some model is going to be implicitly used in the subsequent descriptive and policy analysis whether we like it or not. As a result the relevant question is often not whether the model is exactly correct, or whether it satisfies some formal statistical test. Rather it is whether we believe the implications of the internally consistent structural model whose parameters have been obtained from the data more than we believe the implications of the alternative lines of reasoning available. Of course we also have to be careful not to fall into the habit of accepting the implications of the structural estimates as gospel (forgetting that corners had to be cut to obtain them). There is always room (indeed a need) for doubt (especially if it is constructive), and it will always be possible to make further improvements (though sometimes it might not be worthwhile).

It is easy to close this paper on an optimistic note. The one fact that seems clear from the developments over the last decade of modelling is that advances in theory, computation, statistical methods, and data sources have generated dramatic increases in our ability to take economic models to data and come back with useful interpretations of reality. Moreover, if anything, the rate of increase in our abilities to engage in such endeavors has been accelerating, making it an exciting time to be engaged in empirical research.

Bibliography

Abramowitz, M. and I. Stegun (1972); Handbook of Mathematical Functions, Dover Publications, New York.

Ahn, H. and J.L. Powell (1990); "Semiparametric Estimation of Censored Selection models with a Nonparametric Selection Mechanism," mimeo, University of Wisconsin–Madison, April 1990.

Aigner, D., C. Hsiao, A. Kapetyn, and T. Wansbeek, (1984); "Latent Variable Models in Econometrics", pp. 1321–1393 in Z. Griliches and M. Intrilligator (ed.s), The Handbook of Econometrics.

Aldous, D.J. (1983); "Exchangeability and Related Topics", Springer Lecture Notes in Mathematics, No. 117, Ecole D'ete de Probabilities de Saint–Flour XIII,pp 2–199.

Altonji, J., F. Hayashi, and L. Kotlikoff, (1990); "Risk Sharing, Altruism, and Factor Structure of Consumption", mimeo, University of Pennsylvania.

Altug, S., and Miller, R. (1990); "Household Choices in Equilibrium", Econometrica, 58, pp.543–570.

Andersen, T. and C. Hsiao (1982); "Formulation and Estimation of Dynamic Models Using Panel Data", Journal of Econometrics Vol. 18 pp. 47–82.

Andersen E.B., (1973); Conditional Inference and Models for Measuring, Mentalhygiehnisk Forlag, Kopeernhavn.

Andrews, D.W.K. (1988); "Laws of Large Numbers For Dependent Non–Identically Distributed Random Variables", Econometric Theory, 4, pp.458–467.

Andrews, D.W.K. (1989); "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," Discussion Paper #874R, Cowles Foundation, Yale University, June 1989.

Andrews, D.W.K. (1989a); "Asymptotics for Semiparametric Econometric Models: I Estimation," mimeo, Cowles Foundation, Yale University

Andrews D.W.K. (forthcoming); "An Empirical Process Central Limit Theorem For

Dependent Non–Identically Distributed Random Variables", Journal of Multivariate Analysis.

Benveniste, Lawrence M., and Jose A. Scheinkman, 1979; "On the Differentiability of the Value Function in Dynamic Models of Economics." Econometrica 47:727–732.

Berry, S. (1991); "Discrete Choice Models of Oligopoly Product Differentiation," mimeo, Yale University.

Billingsley, P. (1968); Convergence of Probability Measures, Wiley Series in Probability and Statistics, John Wiley and Sons, New York.

Blume, L., Easley D., and O'Hara, M. (1982); "Characterization of Optimal Stochastic Dynamic Programs", Journal of Economic Theory, 28, pp. 221–234.

Caplin, A. and B. Nalebuff (1991); "Aggregation and Imperfect Competition: the Existence of Equilibrium" Econometrica, Vol. 59, pp. 25–59.

Carroll, R. (1982); "Adapting for Heteroskedasticity in Linear Models" Annals of Statistics, vol.10, pp1224–1233.

Chamberlain, G. (1982); "Multivariate Regression Models For Panel Data", Journal of Econometrics, 18, pp 5–46.

Chamberlain, G. (1986); "Panel Data", pp. 1247–1318 Handbook of Econometrics, edited by Z. Griliches, and M. Intrilligator.

Chamberlain, G. (1987); Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", Journal of Econometrics, pp.305–334.

Chamberlain, G. (1989); "Efficiency Bounds for Semiparametric Regression," mimeo Department of Economics, Harvard University

Choi, K. (1990); "The Semiparametric Estimation of the Sample Selection Model Using Series Expansion and The Propensity Score", mimeo the Department of Economics, the University of Chicago.

Christiansen, Bent (1990); "Dynamic Programming, Maximum Likelihood, Stock Price Volitability and the ICAPM", mimeo, NYU, Business School.

Cosslett, S. (1988); "Probabilistic Choice Over a Continuous Range: An Econometric Model Based on Extreme Value Stochastic Processes", mimeo, Department of Economics, the Ohio State University.

Dagsvick, J.K.(1988); "The Generalized Extreme Value Utility Model for Continuous Choice", mimeo, Central Bureau of Statistics, Oslo.

Das, S. (forthcoming); "A Micro Econometric Model of Capital Utilization and Retirement", Review of Economic Studies.

Davis, S. and J. Haltwinger, (1990); "Gross Job Creation, Gross Job Destruction: Microeconomic Evidence and Macroeconomic Implications," NBER Macroeconomics Annual, 5, 123–168.

Ericson, R. and A. Pakes (1989); "An Alternative Theory of Firm and Industry Dynamics," Discussion Paper No. 445, Columbia University.

Fuller, W. (1987); Measurement Error Models, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York.

Geweke, J. (1977); "The Dynamic Factor Analysis of Economic Time–Series Models" in D. J. Aigner & A.S. Goldberger, Latent Variables in Socio–Economic Models, Amsterdam: North Holland.

Gilchrist, S. (1989); "An Empirical Analysis of Corporate Investment and Financing Hierarchies Using Firm Level Panel Data," unpublished Ph.D. dissertation, University of Wisconsin.

Hajivassiliou, V., and Ioannides Y., (1990); "Dynamic Switching Regression Models of the Euler Equation: Consumption and Liquidity Constraints", mimeo, Department of Economics, Yale University.

Hall, R. (1979); "Stochastic Implications of the Life–Cycle Permanent Income Hypothesis: Theory and Evidence," Journal of Political Economy, Vol. 86, pp. 971–87.

Hansen, L (1982); "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica, 50, pp.1029–1054.

Hansen, L., and Singleton, K. (1982); "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models", Econometrica,50, pp.1269–1286.

Hansen, L. (1985); "A Method for Calculating Bounds on the Asymptotic Covariance of Generalized Method of Moments Estimators", Journal of Econometrics, vol.30, pp205–238.

Hansen, L. and T. Sargent, (1990); "Recursive Linear Models of Dynamic Economies", Mimeo, the Hoover Institution.

Hansen, L. and T. Sargent (1982); "Two Difficulties in Interpreting Vector Autoregressions", in L. Hansen and T. Sargent (ed.s) Rational Expectations Econometrics, Westview Press, forthcoming.

Hansen,L., J.Heaton, and M.Ogaki (1988); "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions", Journal of the American Statistical Association, pp.863–71.

Hausman,J., H. Ichimura, W. Newey, and J. Powell (forthcoming); "Semiparametric Identification and Estimation of Polynomial Errors–in–Variables Models," Journal of Econometrics.

Hayashi, 1987; "Tests for Liquidity Constraints: A Critical Survey and Some New Observations" in T. Bewley (ed) Advances in Econometrics, pp. 91–120 (Econometric Society 5th World Congress), Cambridge University Press, New York.

Heckman, J. and R. Robb (1985); "Alternative Methods for Evaluating the Impact of Interventions" in J. Heckman and B. Singer (ed.s) ; Longitudinal Analysis of Labor Market Data, Cambridge University Press.

Heckman, J. and J. Scheinkman (1987); "The Importance of Bundling in a Gorman–Lancaster Model of Earnings", Review of Economic Studies, LIV, pp.243–255.

Heckman, J. (1981); "The Incidental Parameter Problem and the Problem of Initial Conditions in Estimating a Discrete Time–Discrete Data Stochastic Process", in

Structural Analysis of Discrete Data with Econometric Applications, C. Manski and D. McFadden (eds.) pp. 179–197.

Hendricks, K., R.H. Porter and C.A. Wilson (1990); "Auctions for Oil and Gas Leases with an Informed Bidder and a Random Reservation Price," mimeo, Northwestern University.

Himmelburg, C. (1990); "Essays in the Relationship Between Investment and Internal Finance," unpublished Ph.D. dissertation, Northwestern University.

Hoffman K. and R. Kunze (1972); Linear Algebra, Second Edition, Prentice Hall, Englewood Cliffs, New Jersey.

Hoppenhayn H. and R. Rogerson (1990); "Labor Turnover and Policy Evaluation In A Model Of Industry Equilibrium", mimeo, Graduate School of Business Stanford University.

Hubbard G. and A. Kashyap (1989); "Internal Net Worth in the Investment Process," NBER Working Paper.

Journal of Business and Economic Statistics (1990); "Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods", symposium edited by J. Taylor and H. Uhlig, in the Journal of Business and Economic Statistics, pp. 1–55.

Jovanovic, B. (1989); "Observable Implications of Models with Multiple Equilibria", Econometrica, 57, pp. 1431–39.

Jovanovic, B. and G. MacDonald (1990); "Competitive Diffusion", mimeo. University of Western Ontario.

Judd, K. (1989); "Minimum Weighted Residual Methods in Dynamic Economic Models", mimeo, the Hoover Institution.

Judd, K. (1990); "Cournot Versus Bertrand: A Dynamic Resolution", mimeo, the Hoover Institution.

Keane, M. and D. Runkle, (1990); "Testing the Rationality of Price Forecasts: New Evidence From Panel Data", the American Economic Review, 80, pp. 714–735.

Kiefer, J. and J. Wolfowitz, (1956); "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters" Annals of Mathematical Statistics: 27: 887–906.

Kushner, H. (1965); "On the Stochastic Maximum Principle: Fixed Time of Control", Journal of Mathematical Analysis of Application, 11, pp. 78–92.

Kushner, H. (1965); "On Stochastic Extremum Problems: Calculus", Journal of Mathematical Analysis and Applications, 10, pp. 354–367.

Kydland, F. (1975); "Noncooperative and Dominant Player Solutions In Discrete Dynamic Games", International Economic Review, 16, pp. 321–335.

Lucas, R.(1978); "Asset Prices In An Exchange Economy", Econometrica, 46, pp.1429–1445.

Marcet 1990

Marschask, J and W.H. Andrews (1944); "Random Simultaneous Equations and the Theory of Production," Econometrica, Vol. 12, pp. 143–205.

Maskin, E. and J. Tirole (1988a); "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs," Econometrica, Vol. 56, pp. 549–569.

Maskin, E. and J. Tirole (1988b); "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles" Econometrica, Vol. 56, pp. 571–599.

Maskin, E. and J. Tirole (1987); " A Theory of Dynamic Oligopoly, III: Cournot Competition," European Economic Review, Vol. 31, pp. 947–968.

Matzkin, R. (forthcoming); "Nonparametric and Distribution–Free Estimation of the Binary Threshold Crossing and the Binary Choice Models", Econometrica.

McFadden, D. (1989); "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," Econometrica, Vol. 57, pp. 995–1026.

McGuckin, R. and G. Pascoe (1988); "The Longitudinal Research Data Base (LRD): Status and Research Possibilities", U.S. Bureau of the Census, Center for Economic Studies DP#88–2.

Mundlak, Y. (1963); "Estimation of Production and Behavioral Functions From A Combination of Time Series and Cross Section Data", in Measurement in Economics; Essays in Honor of Yehuda Grunfeld, ed. C. Christ.

Miller, R. (1984); "Job Matching and Occupational Choice," the Journal of Political Economy, Vol. 92, pp. 390–409.

Newey, W.(1988); "Two–step Series Estimation of Sample Selection Models", mimeo, Department of Economics, Princeton University.

Newey, W. (1989); "Series Estimation of Regression Functionals," Econometric Research Program Memorandum #348, Princeton University, June 1989.

Newey,W. (1989b); "Distribution Free Simulated Moments Estimation of Nonlinear Errors–in–Variables Models", mimeo, Princeton University.

Newey,W. (1989c); "The Asymptotic Variance of Semiparametric Estimators", mimeo, Department of Economics, M.I.T.

Newey, W.K. and J.L. Powell (1988); "Nonparametric Instrumental Variables Estimation," manuscript, Department of Economics, University of Wisconsin, December (revised October, 1989).

Newey, W.(1990); "Efficient Instrumental Variables Estimation of Nonlinear Models" Econometrica, Vol. 58, pp 809–39.

Newey, W.(1990b); "Semiparametric Efficiency Bounds", Journal of Applied Econometrics, Vol.5, pp.99–135.

Neyman, J. and E. Scott (1948); "Consistent Estimates Based on Partially Consistent Observations", Econometrica, Vol. 16.

Olley, S. (1991); "An Econometric Study of the Telecommunications Equipment Industry", unpublished Ph.D. Dissertation, University of Wisconsin.

Olley, S. and A. Pakes (1991); "The Dynamics of Productivity in the Telecommunications Equipment Industry", mimeo, Yale University, January 1991.

Pakes, A. and D. Pollard (1989); "Simulation and the Asymptotics of Optimization Estimators," Econometrica, Vol. 57, pp. 995–1026.

Pakes, A. (1986); "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," Econometrica, Vol 54, pp. 755–784.

Pakes, A. and Z. Griliches(1984); "Estimating Distributed Lags in Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs", Review of Economic Studies, pp.243–262.

Pakes, A. and P. McGuire(1990); "Computation of Markov Perfect Equilibria: Numerical Implications of a Dynamic Differentiated Product Model", mimeo, Yale University.

Resnick, S. and Roy, R.(1989); "Random USC Functions, Max–Stable Processes, and Continuous Choice", mimeo, Cornell University.

Roberts, K., and Weitzman, M. (1981); "Funding Criteria for Research, Development, and Exploration Processes", Econometrica, 49, pp.1261–1288.

Robinson, P. (1987); "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form", Econometrica, Vol.55, pp.875–891.

Robinson, P. (1988); "Root–N–Consistent Semiparametric Regression", Econometrica, 56, pp. 931–954.

Romer, P. (1986); "Increasing Returns and Long–Run Growth," Journal of Political Economy, Vol. 94, pp. 1002–37.

Runkle, D. (1983); "Essay in Empirical Macroeconomics", Unpublished Ph.D. Dissertation, M.I.T.

Rust, J.(1987); "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher" Econometrica, 55, 999–1034.

Rust, J.(1991); "Dynamic Structural Models; Problems and Prospects: Discrete Decision Processes", forthcoming in J.J.Laffont and C.Sims (ed.s): Advances in

Econometrics, proceedings of the sixth world congress of the Econometric society.

Sargent, T.J. and C.A. Sims (1977); "Business Cycle Modeling without Pretending to Have Too Much A Priori Economic Theory." in C.A. Sims (ed.) New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank Minneapolis, Minneapolis.

Sherman, R. (1991); "U–Processes and Semi–Parametric Estimation", Unpublished Ph.D. Dissertation, Department of Statistics, Yale University.

Stokey, N., R.E. Lucas and E.C. Prescott (1989); "Recursive Methods in Economic Dynamics, Harvard University Press, Cambridge, Massachusetts.

Tirole, J. (1989); The Theory of Industrial Organization, MIT Press, Cambridge, Massachusetts.

Topkis, D.M (1978); "Minimizing a Submodular Function on a Lattice", Operations Research, 26, pp.305–320.

Townsend, R. (1990); "Risk and Insurance in Village India", Mimeo, University of Chicago.

Wedervang, F.(1965); Development of A Population of Industrial Firms, Oslo, Norway; Scandinavian University Books.

Wolak, F. (1990); "Estimating Regulated Firm Production Functions with Private Information: An Application to California Water Utilities", mimeo, Stanford University.

Zeldes, S. (1989); "Consumption and Liquidity Constraints: An Empirical Investigation," Journal of Political Economy, Vol. 97, pp. 305–346.

Appendix I

Proof of Lemma 26

(monotonicity of the investment policy with deterministic accumulation)

26. Lemma.

Assume 2, that $\partial\pi(\omega,k)/\partial k$ is increasing in ω, that $\partial c(x,k_1)/\partial x \leq \partial c(x,k_2)/\partial x$ is nonincreasing in k for all x, and that if $h(\cdot)$ is continuous (a.e) and integrable with respect to a subset of $\mathbb{P}_{\omega}$, say $\mathbb{P}^*$, then provided $P(\cdot|\omega_1)$,and $P(\cdot|\omega_2) \in \mathbb{P}^*$,
$|\int h(\omega')[P(d\omega'|\omega_1)-P(d\omega'|\omega_2)]| \leq \psi(h,\mathbb{P}^*)|\omega_1-\omega_2|$.

Proof.

The proof assumes that the optimal policy is unique. Given this we show that the solution to the finite horizon problem, say $x^T(\omega,k)$, is, for all T, weakly increasing in ω for each k. It is straightforward to show that this implies that the limit function, $x(\omega,k)$ must also be weakly increasing.

For every T define

$$V^T(\omega,x,k) = -c(x,k)+\pi(\omega,k)+\beta\int V^{T-1}[\omega',k(1-\delta)+x]P(d\omega'|\omega)$$

where $V^{T-1}(\omega,k)$ provides the value of a T−1 horizon problem. Two properties of this function will be used below. First we use the fact that maximization implies that $V^T(\omega,x,k) \leq V^T(\omega,k)$ for all $x \in \Gamma(k)$,and all $(\omega,k) \in \Omega \times K$. Second we need the fact that if $V^T(\omega,x,k)$ has isotone differences in (ω,x), i.e.if

$$[V^T(\omega_1,x_1,k)-V^T(\omega_1,x_2,k)]-[V^T(\omega_2,x_1,k)-V^T(\omega_2,x_2,k)]\geq 0$$

whenever $\omega_1 \geq \omega_2$ and $x_1 \geq x_2$, then $x^T(\omega,k)$ is nondecreasing in ω [Briefly, since both Ω and K are totally ordered, the fact that $V^T(\cdot)$ has isotone differences implies that it is supermodular; see Topkis,1978, Theorem 3.2. This, together with the fact that $\Gamma(\cdot)$ is independent of ω, implies the result; see theorem 6.1 of Topkis, 1978].

This latter result implies that to prove the theorem we need only show that for any T, $V^T(\cdot)$, has isotone differences. We now use induction to prove this fact. For the initial condition of the inductive argument we need only note that

$$[V^1(\omega_1,x_1,k)-V^1(\omega_1,x_2,k)] - [V^1(\omega_2,x_1,k)-V^1(\omega_2,x_2,k)]$$
$$= \beta\int \{\pi[\omega',k(1-\delta)+x_1]-\pi[\omega',k(1-\delta)+x_2]\}[P(d\omega'|\omega_1)-P(d\omega'|\omega_2)] \geq 0,$$

where the inequality follows from the supermodularity of $\pi()$ and the fact that the family $\mathbb{P}_\omega$ is stochastically increasing in ω (see assumption 2).

For the inductive step assume that $\{V^t(\omega,x,k)\}_{t=1}^{T-1}$ is supermodular in (ω,x) for each k. Lemma * below shows that this implies that $V^{T-1}(\omega,k)$ is supermodular in (ω,k). Consequently

$$[V^T(\omega_1,x_1,k)-V^T(\omega_1,x_2,k)] - [V^T(\omega_2,x_1,k)-V^T(\omega_2,x_2,k)] =$$
$$\beta\int \{V^{T-1}[\omega',k(1-\delta)+x_1]-V^{T-1}[\omega',k(1-\delta)+x_2]\}[P(d\omega'|\omega_1)-P(d\omega'|\omega_2)] \geq 0,$$

where the inequality follows from Lemma *, and the fact that $\mathbb{P}_\omega$ is stochastically increasing in ω. ○

<u>Lemma *</u>.

$V^T(\omega,k)$ is supermodular in (ω,k) on $\Omega\times K$, if $\{V^t(\omega,x,k)\}_{t=1}^{T}$ are supermodular in (ω,x) on $\Omega\times\Gamma(k)$, for each $k \in K$.

<u>Proof.</u>

Again by induction on T. The initial condition of the inductive argument is analogous to the initial condition of the inductive argument in Lemma 26 and need not be repeated. For the inductive step assume that $\{V^t(\omega,k)\}_{t=1}^{T-1}$ are supermodular on $\Omega\times K$. Then, using the shorthand that for any $f:\Omega\times K\to\mathbb{R}$, $f(i,j)=f(\omega_i,k_j)$,

$$(1) \qquad [V(1,1)-V(1,2)] - [V(2,1)-V(2,2)] \geq$$
$$\{V^T[\omega_1,x(\omega_2,k_1),k_1]-V^T(\omega_1,k_2)\} - \{V^T(\omega_2,k_1)-V^T[\omega_2,x(\omega_1,k_2),k_2]\}$$
$$\equiv\Delta+\beta\int\{V^{T-1}[\omega',k_1(1-\delta)+x(\omega_2,k_1)]-V^{T-1}[\omega',k_2(1-\delta)+x(\omega_1,k_2)]\}[P(d\omega'|\omega_1)-P(d\omega'|\omega_2)],$$

where,

$$\Delta \equiv [\pi(1,1)-\pi(1,2)] - [\pi(2,1)-\pi(2,2)] > 0.$$

Consequently if

$$k_1(1-\delta)+x(\omega_2,k_1) \geq k_2(1-\delta)+x(\omega_1,k_2)$$

the proof is complete. So assume to the contrary that

$$x(1,2)-x(2,1)-(k_1-k_2)(1-\delta)\equiv\kappa>0.$$

The assumption that $\partial c(x,k)/\partial x$ is nonincreasing in k, together with the Euler equation in Lemma 2, and the convexity of the investment cost function implies that $x(1,1)+k_1(1-\delta)>x(1,2)+k_2(1-\delta)$ whenever $k_1>k_2$. Hence by continuity of the optimal policy, and the hypothesis of the inductive argument (which insures that $x^T(\omega,k)$ is weakly increasing in ω) there exists an $\omega^*\in[\omega_1,\omega_2]$ such that

$$k_1(1-\delta)+x(\omega^*,k_1) = k_2(1-\delta)+x(\omega_1,k_2) \equiv \bar{k}.$$

Substituting ω^* for ω_1 in (1) we have

$$[V^T(\omega^*,k_1)-V^T(\omega^*,k_2)] - [V^T(\omega_2,k_1)-V^T(\omega_2,k_2)] \geq$$
$$[\pi(\omega^*,k_1)-\pi(\omega^*,k_2)] - [\pi(\omega_2,k_1)-\pi(\omega_2,k_2)] \geq 0,$$

where the last inequality follows from the supermodularity of $\pi()$.

Next we will show that $\omega_1-\omega^* \geq J(k_1,k_2)$. That will complete the proof because it will imply that we can break the move from ω_1 to ω_2 into a finite number of steps (each of which preserves isotone differences). To this end note that Lemma 2 implies that

$$\partial c[x(\omega_1,k_2),k_2]/\partial x - \partial c[x(\omega^*,k_1),k_1]/\partial x =$$
$$\beta\int \chi(\omega',\bar{k})\{\partial\pi(\omega',\bar{k})/\partial k + \partial c[x(\omega',\bar{k}),\bar{k}]/\partial x\}[P(d\omega'|\omega_1)-P(d\omega'|\omega^*)]$$
$$\equiv\int h(\omega')[P(d\omega'|\omega_1)-P(d\omega'|\omega^*)].$$

The convexity of $c(\cdot)$ and the assumptions on $\mathbb{P}_\omega$ then imply that

$$\beta^{-1}[\partial c(x=0,k_1)/\partial x][k_1-k_2](1-\delta)\leq\psi(h,\mathbb{P}^*)(\omega_1-\omega^*),$$

or

$$\omega_1-\omega^* \geq\beta^{-1}[\partial c(x=0,k_1)/\partial x][k_1-k_2](1-\delta)/\psi(h,\mathbb{P}^*)\equiv J(k_1,k_2),$$

where $\mathbb{P}^*$ includes the interval $[\omega_1,\omega_2]$, as required.∘

Appendix 2: Proof of Lemma 33.2

Lemma 33.2.

An $f \in V_\lambda$ is also a member of $V_\lambda^{\mathscr{E}}$ if and only if for all $h \in H^N$,

$$\alpha_f(h_1, h_2, \ldots h_N) = \alpha_f(h_1, \pi_2, \ldots \pi_N),$$

for any $(\pi_2, \ldots, \pi_N)$ which is a permutation of $(h_2, \ldots, h_N)$.

Proof.

We prove that partial exchangeability of the value functions implies partial exchangeability of the coefficients. The other direction of causation is immediate since the sum of partially exchangeable functions is partially exchangeable.

The proof is by induction on $d = d(h_2, \ldots, h_N)$, the number of non–zero elements in $(h_2, \ldots, h_N)$. To prove the initial condition of the inductive argument (that the lemma is true for $d = 1$) consider ω–vectors of the form.

$$\omega_1 = \varphi_1, \quad \omega_j = \varphi \text{ for some } j \neq 1, \text{ and } \omega_{j'} = 0 \text{ for } j' \neq j,1. \quad (A.1),$$

For ω–vectors satisfying A1, equation (2) implies

$$\hat{V}(\cdot) = \sum_{p=0}^{\lambda} \sum_{h_1=0}^{p} \alpha(h_1, h_j = p-h_1)\, \varphi_1^{h_1} \varphi^{p-h_1} \equiv \underset{\sim}{\varphi}' \underset{\sim}{\alpha} = c(\varphi_1, \varphi), \quad (A.2)$$

where $\alpha^p(h_1, h_j = p-h_1) = \alpha(h_1, 0, \ldots, h_{j-1} = 0, h_j = p-h_1, 0, \ldots, 0).$

Note that the dimensions of the α–vector is

$$\sum_{p=0}^{\lambda} (p+1) = (\lambda+1)(\lambda+2)/2,$$

and that both $\underset{\sim}{\varphi}$ and $c(\varphi_1,\varphi)$ are independent of j. Construct $(\lambda+1)(\lambda+2)/2$ independent $\underset{\sim}{\varphi}$ vectors by setting $\varphi_1 = i$ and $\varphi_2 = 1,\ldots,K$, where $K = \# \Omega$, for $i = 1,2,\ldots,K$. This will be possible provided $(\lambda+1)(\lambda+2)/2 < K^2$, that is provided:

$$\lambda+1 \leq K,$$

a condition we assume in what follows. Stack the analog of equation (A.2) for the alternative values of $\underset{\sim}{\varphi}$ and solve the resulting system for

$$\underset{\sim}{\alpha} = \Phi^{-1}\underset{-}{C}, \qquad \text{(A.3)}$$

where $\underset{-}{C} = [C(1,1), C(1,2)\ldots]$ and the rows of Φ are the values of the vector $\underset{-}{\varphi}$ used to form the basis above.

Note that the $\underset{-}{\alpha}$ defined by (A3) is independent of "j". Consequently,

$$\alpha(h_1,0,\ldots,0,h_j{=}m,0,\ldots,0) = \alpha(h_1,m,0,\ldots,0)$$

for all h with $d(h) = 1$, $h_j{=}m$, and $h_{j'}{=}0$ for $j' \neq j,1$, as required.

Now assume the lemma for $d(h) \leq d-1$. To prove this implies the lemma for $d(h) =$

d consider ω–vectors of the form:

$$\omega_1 = \varphi_1, \omega_{j_2} = \varphi_2, \dots, \omega_{j_{d+1}} = \varphi_{d+1}, \qquad \text{(A.4a)}$$

with $j_q \neq j_{q'} \neq 1$ for $q, q' = 2,\dots,d+1$, while

$$\omega_j = 0 \quad \text{for } j \notin \mathcal{J} = \{j_1 = 1, j_2, \dots j_{d+1}\} \qquad \text{(A.4b)}$$

To analyze this case it will be useful to let $\underline{1}$ be the N– dimensional vector which puts a 1 at every location in $\mathcal{J}$ and a zero elsewhere, and e_i to be the N–vector that puts a one at the i^{th} location and a zero elsewhere. Then equation (2), together with the initial condition of the inductive argument [note that this sets $\alpha(h)$ for all h such that $p(h) = \Sigma h_i \leq d$] implies that for any ω–vector satisfying (A.4).

$$c(\varphi_1,\dots,\varphi_N)\cdot[\varphi_1,\dots,\varphi_{d+1}]^{-1} = \alpha(\underline{1}) + \sum_{i=1}^{d+1} \alpha(\underline{1} + e_{j_i})\varphi_i$$

$$+ \sum_{i_2 \geq i_1}^{d+1} \sum_{i_1=1}^{d+1} \alpha(\underline{1} + e_{j_{i_1}} + e_{j_{i_2}})\varphi_{i_1}\varphi_{i_2} + \dots +$$

$$\sum_{i_{\lambda-(d+1)} \geq i_{\lambda-(d+2)}}^{d+1} \dots \sum_{i_1=1}^{d+1} \alpha(\underline{1} + \sum_{k=1}^{\lambda-(d+1)} e_{j_{i_k}}) \Pi_{k=1}^{\lambda-(d+1)} \varphi_{i_k}$$

$$\equiv \underline{\alpha}' \underline{\varphi} = c(\varphi_1,\dots,\varphi_{d+1}), \qquad \text{(A.5).}$$

The number of coefficients in the $h+1^{st}$ term in (A.5) is the number of sequences $[i_1,\dots,i_h]$

with $i_j \geq i_{j-1}$ and $i_j \epsilon [1,2,\ldots,d+1]$, for $j=1,\ldots,h$. By lemma 32 this is

$$\mathcal{J}[d+1,h] = \binom{d+h}{h}$$

Now use the fact that

$$\sum_{h=0}^{x} \binom{d+h}{h} = \binom{d+x+1}{x}$$

(see for e.g. Abronowitz and Stegum, 1972, p. 822) to show that the total number of coefficients in A.5 is

$$\binom{\lambda}{\lambda - (d+1)}.$$

Note, that $\underset{\sim}{\varphi}$ and $c(\varphi_1,\ldots,\varphi_{d+1})$ are independent of $\mathcal{J}$ and construct, $\binom{\lambda}{\lambda-(d+1)}$, linearly independent $\underset{\sim}{\varphi}$ vectors by letting φ_i range over $1,\ldots,K$, for $i=1,\ldots,d+1$. That this is possible follows from the fact that $(\lambda+1) \leq K$. Use these vectors to determine the $\underset{\sim}{\alpha}$ coefficients in (A.5) as in the argument leading to (A.3) Since this $\underset{\sim}{\alpha}$ vector is independent of $\mathcal{J}$ we have for all h such that $d(h) = d$

$$\alpha(h_1,\ldots,h_N) = \alpha(h_1,\pi_2,\ldots,\pi_N)$$

for any $(N-1)$ dimensional vector $(\pi_2,\ldots,\pi_N)$ which is a permutation of $(h_2,\ldots,h_N)$. □