Estimation from the Optimality Conditions for Dynamic Controls.

Steve Berry, and Ariel Pakes.*

December 2000

This note introduces an estimation technique based on the optimality conditions from dynamic models. The technique has at least two virtues; i)it is easy to implement, and ii) it is logically consistent with the underlying model (and hence can be used to analyze the model's implications). It is driven by the assumption that agents perceptions of the distribution of the discounted value of future net cash flows are rational in the sense that they are, at least on average, consistent with the realizations of those net cash flows.

Since this assumption is used extensively in applied work, our estimator should be useful in a wide variety of settings. We are, however, particularly interested in using it to estimate parameters from dynamic models of oligopolistic industries. Consequently after a brief introduction to the estimator, we will focus on how our estimator can help analyze behavior in oligopolistic industries.

In this context there are three issues that motivated the development of our estimator. First, the first order conditions currently used in empirical models of markets are typically the equations from a *static* Nash equilibrium (in either prices or in quantities). For these equations to be an accurate description of pricing (or quantity setting) behavior the conditional distribution of the firm's future states, conditional on whatever investment decisions are made in the interim, does not depend on the price (or quantity) decision being studied. There are many important examples in which current prices or quantities will have independent impacts on future states, as well as on current profits, and in these cases the logic behind the static price or quantity setting equation falls apart. Examples include models with: learning by doing; either durable, experience, addictive, or network goods; and collusion. Our estimation technique can allow for any (or all) of these extensions, and nests the standard static models as testable special cases.

*Yale University and the NBER, and Harvard University and the NBER, respectively. We are greatful to Liran Einav for several thoughtful suggestions, as well as for programming assistance

Second, the technique allows us to use the information in investment choices to help in estimation (investment here is defined broadly enough to include R&D and advertising choices, as well as investment in traditional capital stock). In this respect our technique has computationally properties similar to those of Euler equation estimation techniques (these were introduced into economics for the estimation of dynamic models by Hall,1979, and Hansen and Singleton,1982). That is, neither the Euler equation nor our technique requires the researcher to solve explicitly for the value function and/or the investment policy of the firm in order to use the implications of the investment choices in estimation. However, as discussed in Pakes (1993), Euler equations cannot (in general) be derived for problems that involve interactions among agents (as in the dynamic oligopoly models that are our main interest), while our technique can.

Finally our technique will allow the data to provide the researcher with information on the value of sunk cost parameters. Sunk cost parameters (usually costs of entry and selloff values at exit) are typically the hardest parameters to find information on in industrial organization models. Most of our information on the primitive parameters that underlie IO models comes from estimates of demand or production functions, or from the response of a continuous state variable to investment decisions. The only decisions that are directly related to sunk costs are the entry and exit decisions themselves, and the form of that relationship is typically too complex to use in an estimation algorithm. Still the values of the sunk cost parameters are crucial to many of the policy implications that motivate the empirical analysis.

Our technique works off the equilibrium conditions for the choice of controls. Standard first order conditions for continuous controls depend upon the (derivative of the) expected discounted value of future net cash flows. This expectation is not directly observed. However, under our rational expectations assumption the actual discounted value of future net cash flows will be an error prone measure of this unobserved expectation, and, in equilibrium, the error in this proxy for the continuation value will be mean independent of all variables known at the time the expectation is made. This fact leads to a particularly simple estimating equation if we use the stochastic accumulation framework described in Pakes (1993) for the evolution of the state variables of the model, so we focus on that case in this paper. Then the first order condition can be expressed as a conditional expectation of known primitive functions, and a standard method of moments estimation algorithm can be used. A more complicated condition is available for the choice of discrete controls (like entry and exit), and we discuss the use of these conditions later in the paper.

The major difficulty that is likely to arise in the use of our technique is that the entire realized discounted value of future net cash flow is not observed for firms that are still in operation at the end of the sample period. This generates a "truncation remainder" problem similar to the "initial conditions" problem highlighted in Heckman $(\cdot)$. We suggest several ways of treating this truncation remainder problem, and use Monte Carlo procedures to examine both its importance, and the performance of our suggested solutions.

We begin by outlining our model; first for the firm's problem in a general setting, and then in the context of Markov Perfect equilibria. Next we go to a Monte Carlo analysis of two examples. The first involves a monopolist selling an experience good, and the second the pricing and investment choices of a collusive industry.

1 The Firm's Problem.

The firm's problem is choose a sequence of actions, say $\{a_{t+\tau}\}$, to maximize the expected discounted value of net cash flows (of $\{n_{t+\tau}\}$) conditional on the information sets (on $\{\sigma_{t+\tau}\}$) that will be available when those actions are to be taken. Thus if we let $A(\sigma_t)$ define the feasible choices of a_t (typically this will be the positive orthant), the firm's problem is to choose a_t to

$$sup_{a_t \in A_t} E[\sum_{\tau=0}^{\infty} \beta^\tau n_{t+\tau} | \sigma_t, a_t] \equiv E[\sum_{\tau=0}^{\infty} \beta^\tau n_{t+\tau} | \sigma_t] \equiv V(\sigma_t), \tag{1}$$

where the expectation is taken with the understanding that optimal actions will be taken in each future period, and β is the discount rate. We will refer to $V(\sigma)$ as the value of the firm.

We make the standard assumptions that $0 \leq \beta < 1$, $\{\sigma_t\}$ evolves as a Markov process, and net cash flows are bounded from above. Blackwell's theorem then implies that the solution to our problem can be obtained from the unique value function that solves the Bellman equation

$$V(\sigma) = sup_{a \in A(\sigma)} \{n(a, \sigma) + \beta \int_{\sigma'} V(\sigma') dP(\sigma' | \sigma, a)\}, \tag{2}$$

where $P(\cdot | \cdot, a)$ is the Markov transition kernel for $\{\sigma_t\}$ conditional on the action a.

Throughout we will make the following two assumptions.

Assumption 1 (Rational Expectations)

$$\sum_{\tau=0}^{\infty} \beta^\tau n_{t+\tau} = V(\sigma_t) + \nu_t$$

with,

$$E[\nu_t | \sigma_t] = 0.$$

Assumption 2 (Smoothness) *$P(\sigma' | \sigma, a)$ has support which is independent of a and a density, $p(\sigma' | \sigma, a)$, which is differentiable in a for almost every σ' and every σ. Moreover $n_t = n(\sigma_t, a_t)$ is a differentiable funciton of a for almost every σ.*

Assumption 1 is a rational expectations assumption that has become standard in the literature. Assumption 2 does rule out some familiar models, and it can be relaxed. For example it rules out a standard capital accumulation model where capital in the next period is a deterministic function of capital this period and investment. On the other hand if we were to be a bit more realistic and allow for "noise" in the relationship between investment expenditures and the "constant quality" capital aggregate, it is easy to get back to a model which satisfies our Assumption 2. Moreover, as discussed below, once we relax that assumption the estimation algorithm itself gets considerably more complicated (it requires semiparametric procedures).

Equation 2 and our assumptions imply that if a_t is in the interior of $A(\sigma_t)$ it must satisfy the first order condition

$$0 = \frac{\partial n(a_t, \sigma_t)}{\partial a_t} + \beta \int V(\sigma_{t+1}) \frac{\partial p(\sigma_{t+1}|\sigma_t, a_t)}{\partial a_t} d\sigma_{t+1}$$

$$= \frac{\partial n(a_t, \sigma_t)}{\partial a_t} + \beta E[V(\sigma_{t+1}) \frac{\partial lnp(\sigma_{t+1}|\sigma_t, a_t)}{\partial a_t} |\sigma_t].$$

Substitute $\sum_{\tau=1}^{\infty} \beta^\tau n_{t+\tau} - \nu_{t+1}$ for $V(\sigma_{t+1})$ in this equation, note that $\frac{\partial lnp(\sigma_{t+1}|\sigma_t, a_t)}{\partial a_t}$ is a function of σ_{t+1} and so is uncorrelated with ν_{t+1}, and use the fact that for any x, $E(E[x|\sigma_{t+1}]|\sigma_t) = E[x|\sigma_t]$ to rewrite the above equation as

$$0 = \frac{\partial n(a_t, \sigma_t)}{\partial a_t} + \sum_{\tau=1}^{\infty} \beta^\tau n_{t+\tau} \frac{\partial lnp(\sigma_{t+1}|\sigma_t, a_t)}{\partial a_t} + \epsilon_t,$$

where

$$\epsilon_t = \beta E[V(\sigma_{t+1}) \frac{\partial lnp(\sigma_{t+1}|\sigma_t, a_t)}{\partial a_t} |\sigma_t] - \sum_{\tau=1}^{\infty} \beta^\tau n_{t+\tau} \frac{\partial lnp(\sigma_{t+1}|\sigma_t, a_t)}{\partial a_t},$$

which, together with our assumptions, implies

$$E[\epsilon_t|\sigma_t] = 0.$$

Indexing all functions with the parameter vector of interest (θ), we rewrite this latter equation as

$$E\left[\frac{\partial n(a_t, \sigma_t; \theta)}{\partial a_t} + \sum_{\tau=1}^{\infty} \beta^\tau n(a_{t+\tau}, \sigma_{t+\tau}; \theta) \frac{\partial lnp(\sigma_{t+1}|\sigma_t, a_t; \theta)}{\partial a_t} |\sigma_t\right] = 0 \tag{3}$$

at $\theta = \theta_0$. This is the equation which underlies our estimates of θ_0.

Letting i index different firms, assume temporarily that all firms eventually shut down (with probability one), that we oberve them until shut down, and that though the firm might recieve a selloff value for shutting down, once it does shut down it receives no further net cash flows (the selloff value itself is counted in the net cash flow of the terminal period). In this case the potentially infinite sum in equation (3) is a finite sum and we can construct the integrand in the expectation in (3) for different values of θ, as

$$q(z_{i,t},\theta) \equiv \frac{\partial n(a_{i,t},\sigma_{i,t};\theta)}{\partial a_{i,t}} + \sum_{\tau=1}^{\infty} \beta^{\tau} n(a_{i,t+\tau},\sigma_{i,t+\tau};\theta) \frac{\partial lnp(\sigma_{i,t+1}|\sigma_{i,t},a_{i,t};\theta)}{\partial a_{i,t}},$$

for $t = 1,\ldots,T_i$, where T_i is the (random) shutdown time of firm i, and $i = 1,\ldots,N$. Then if $h(\sigma,\theta)$ is a sufficiently rich function of (σ,θ) and traditional regularity conditons are assumed to hold, a standard method of moment argument implies that an estimator of θ_0 which minimizes a norm of

$$G_N(\theta) = N^{-1}\sum_{i,t} m(z_{i,t},\theta) \equiv N^{-1}\sum_{i,t} q(z_{i,t},\theta)h(\sigma_{i,t},\theta).$$

will be consistent and asymptotically normal.

The required regularity conditions, the choice of instruments (of $h(\cdot)$), the norm we minimize, and the exact form of the limit distribution will depend on the details of the problem at hand. We therefore discuss these issues in the context of our examples.

Finaly keep in mind that we have limited the discussion to using the information contained in the choices of continuous controls for the subset of periods in which those controls were not chosen to be points on the boundary of $A(\sigma_{i,t})$. Since whether or not the control is on the boundary is known as a function of $\sigma_{i,t}$, this selection is done on the basis of information known at t. As a result it is not related to the rational expectations disturbance $\nu_{i,t}$, and does not cause a selection bias in our estimator. On the other hand there is information on the parameter of interest in the choices that are on the boundary that we are not using. Similarly there is information in the discrete choices (like the choice of whether to enter or to exit) that we are not using. We come back to how one could use these sources of information, and how they are likely to help, below.

2 An Application to Markov Perfect Equilibrium.

The result in equation (3) applies both to models of single agents choosing policies in games against "nature", and to dynamic games. In the context of single agent models our suggestion has both advantages and disadvantages relative to the more standard Euler equation estimation techniques introduced into economics by Hall(1979) and Hansen and Singleton(198?)(see the discussion below). We are, however, primarily interested in developing techniques which can be used to analyze dynamic games, particularly those that appear in applied Industrial Organization settings. The perturbation conditions that underlie Euler equations are simply not satisfied in game theoretic models. Consequently Euler equations techniques cannot be used as a basis for estimating (or analyzing) the parameters of the models we are intersted in. We now show how easy it is to apply our estimator to environments that involve dynamic games.

To illustrate, we take a symmetric information, Markov Perfect, model of industry dynamics that extends the Ericson and Pakes (1995) framework slightly (see Pakes, 1999 for a discussion)[1]. In those models, $\sigma_{i,t} = (\omega_{i,t}, \omega_{-i,t})$, where $\omega_{i,t}$ is a vector of the firm's own state variables (typically variables which determine its cost function and the characteristics of the product it markets) and $\omega_{-i,t}$ are the same variables for the firm's competitors. We will allow for two controls, price and an investment, denote them by $a = (\mathbf{p}, x)$, and assume only that $a \in A = \mathcal{R}_+^2$ for all σ. x could be traditional investment, advertising, research and development expenditures (or with obvious extensions, a vector valued list of the three).

If we let $\pi(\cdot)$ denote profits, χ be the indicator function which is zero when the firm exits and one elsewhere, $c(\cdot)$ be the costs of investment function, and ϕ be the selloff value of the firm, then net cash flows are given by

$$n_{i,t} = \chi_{i,t}[\pi(\omega_{i,t}, \mathbf{p}_{i,t}, \omega_{-i,t}, \mathbf{p}_{-i,t}; \theta) - c(\omega_{i,t}, x_{i,t})] + [\chi_{i,t-1} - \chi_{i,t}]\phi. \tag{4}$$

That is net cash flow equals profits minus the cost of investment if the firm continues in operation, and it equals the selloff value if the firm is sold off.

The transition probabilities for $\omega_{i,t}$ given $a_{i,t}$ are primitives of the problem, prespecified up to a parameter vector. They are written as $p(\omega_{i,t+1}|\omega_{i,t}, a_{i,t}; \theta)$ and are assumed to satisfy our assumption 2. Since we deal with a simultaneous move game, the transition kernel for the state vector of the firm's competitors does not depend directly on the firm's choice of actions. Consequently, we can factor the transition kernel for $\sigma_{i,t+1}$ conditional on $a_{i,t}$ and $\sigma_{i,t}$ as

$$p(\sigma_{i,t+1}|\sigma_{i,t}, a_{i,t}; \theta) = p(\omega_{i,t+1}|\omega_{i,t}, a_{i,t}; \theta)p(\omega_{-i,t+1}|\sigma_{i,t}) \tag{5}$$

In Markov Perfect equilibrium $p(\omega_{-i,t+1}|\sigma_{i,t})$ is also a function of θ, as it must be consistent with the behavior of the other competitors each of whom (including the potential entrants) face problems similar to that of the firm we focus on. As a result if we were to add assumptions to our model which would insure a unique equilibrium, we could, at least in principal, solve for the transition kernel for the states of the firm's competitors in terms

[1] We stress, however, that the estimation algorithm proposed here can accomodate situations in which assymetric information is important as well.

of θ, and use that information in the estimation algorithm also. However this would force us to solve a very complicated fixed point for every θ evaluated in the estimation algorithm (for a discussion of the computational burden of such a calculation, see Pakes, 1999). As a result we will ignore this information in what follows.

We begin with the estimating equation that is derived from the first order condition for price. Substitute (5) and (4) into (3), use the fact that the conditional distribution of $\omega_{-i,t+1}$ does not depend on $a_{i,t}$ to factor out the logarithmic derivative of $p(\omega_{i,t+1}|\omega_{i,t}, a_{i,t})$, and rearrange terms to produce

$$E\left[\left(\frac{\partial\pi(\omega_{i,t},\mathbf{p}_{i,t},\omega_{-i,t},\mathbf{p}_{-i,t};\theta)}{\partial\mathbf{p}_{i,t}}+\frac{\partial lnp(\omega_{i,t+1}|\omega_{i,t}a_{i,t};\theta)}{\partial\mathbf{p}_{i,t}}\sum_{\tau=1}^{\infty}\beta^{\tau}\pi(\omega_{i,t},\mathbf{p}_{i,t},\omega_{-i,t},\mathbf{p}_{-i,t};\theta)\right)|\sigma_t\right] \tag{6}$$
$$=0 \quad \text{at} \quad \theta=\theta_0.$$

There are two things to note about this equation. First (6) gives us a first order conditon for price in terms of primitives that are specified at the outset. That is we can compute this first order condition for different values of θ without ever computing an equilbrium. As a result we can find our method of moments estimator of θ_0 without ever computing an equilibrium. Second we can compute the estimator without ever specifying which of the many possible equilibrium that satisfy the set of first order conditions (analogous to 6) that arise for the different firms at different tuples of states. The estimation procedure lets the data choose which of these equilibria are relevant[2].

That is we are able to compute a consistent and asymptotically normal estimate of θ_0 without either;

- computing the fixed point which defines equilibrium behavior, or
- making the assumptions required for that equilibrium to be unique.

We now come back to the substantive issues that generated our interest in equation (6). Estimating equations derived from the first order conditions to pricing problems are used intensively in I.O. Partly this is because cost data is usually proprietary, so that one of the few ways we have of getting either an idea of what either costs or the markup over cost must be is to assume an equilibrium for prices and then estimate costs and markups from the first order conditions for equilibrium pricing behavior[3]. A related reason that pricing

[2]Of course once we go to use the estimates of the parameters to analyze potential policy or environmental changes both the computational and the uniqueness issues will come back to haunt us. Still at least the computational problem at that stage will be much smaller; we will not need to compute the equilibria repeatedly every time we evaluate a different parameter vector in the estimation algorithm.

[3]Of course the markup terms in the pricing equation also provide information on demand parameters, and this is frequently also very important.

equations are used so intensively is that we are often interested in the nature of the pricing equilibrium per se, as in, for example, the study of collusive behavior. Then the empirical work is typically directed at examining which of the possible equilibrium pricing models seems most consistent with the data.

Usually the pricing equations taken to data are the equilibrium conditions from *static* pricing models; i.e. models in which prices only affect current profits and do not have an independent effect on the evolution of the state variables of the system (independent of investment expenditures). When this is the case the second term in (6) is zero. It is this special case which underlies the usual way of teaching introductory I.O. classes. That is we usually break the analysis of market equlibrium into two parts; a static pricing or quantity setting game which can be solved independently of the solution to the dynamics of the problem, and a dynamic investment game which takes the solution to the static game as input and sets the investment policies that will determine the evolution of the state of the system (see Pakes, 1999, and the discussion there).

More advanced courses point out that we would expect "dynamic" pricing behavior in many instances of importance in economics, in which case there is no convenient breakdown into static and dynamic controls. One case of interest is when costs tomorrow respond to quantities (or prices) today, as when learning by doing leads us to believe that current quantity has significant effects on the cost function in future periods(see Benkard,?, and the literature he cites). Another is when demand tomorrow responds in a significant way to demand today, as when; goods are durable, consumer learning effects are important (see Ackerberg,?, the literature he cites, and the example in the next section), or network effects are important (see Marcovitch,? , and the literature she cites). Finally we also expect dynamic pricing when collusion is an important part of the environment, since then future pricing decisions can depend on current strategies (thus in collusive models with punishment schemes we typically would add a state variable which keeps track of previous policies to our state vector, see our second example which is taken from Fershtman and Pakes, 2000).

One of the advantages of our framework is that we can allow for dynamic pricing equations with little in the way of added computational costs. Moreover, the traditional static model will typically be nested to the dynamic model of interest, and so can be easily tested. The tests simply asks whether prices are lower when lower prices (higher demand) are expected to lead to large increments in future net cash flow.

We now move to the first order condition for investment for our problem. It can be derived analogously and is

$$E\left[\left(\frac{\partial - c(\omega_{i,t}, x_{i,t};\theta)}{\partial x_{i,t}} + \frac{\partial lnp(\omega_{i,t+1}|\omega_{i,t}, a_{i,t};\theta)}{\partial x_{i,t}} \sum_{\tau=1}^{\infty} \beta^{\tau} \pi(\omega_{i,t}, \mathbf{p}_{i,t}, \omega_{-i,t}, \mathbf{p}_{-i,t};\theta)\right) |\sigma_t\right] \quad (7)$$
$$= 0 \quad \text{at} \quad \theta = \theta_0.$$

Investemnt equations for single agent models are typically analyzed using Euler equation techniques, so a brief comparison is in order. There are several aspects of our problem which rule out applying Euler equation techniques to estimate its parmeters; (i)we are analyzing a dynamic game, (ii) we are using a stochastic accumulation model, (iii) we allow future actions to be at the corner of the choice set, and so on (for a fuller, though still limited, discussion of the range of application of Euler equation techniques see Pakes, 1992). On the other hand to use our method one does have to specify the transition probabilities of the states conditional on the controls, at least up to a parameter vector – a specification not required in using Euler equation techniques[4].

There is one final point to be made before leaving the Markov Perfect example, and it is quite important. Unlike other estimation techniques based on first order conditions, our technique does contain some information on the sunk cost parameters. Since we have not treated entry explicitly the only sunk cost parameter in our problem is the selloff value or ϕ. Were we to introduce entry we would also specify a sunk cost the entrant would have to pay in order to enter and our entry decision would be a function of it. Sunk costs are typically the parameters that are hardest to obtain information on in I.O. problems, and their magnitude is a crucial determinant of the nature of an industry's reaction to policy and environmental changes.

Of course it may be difficult to extract information on sunk costs from pricing and investment decisions. This is a good reason for making a more serious attempt to integrate the information from the entry and exit decisions per se (see below).

3 Truncation Remainders.

For firms that are still alive at the end of the sampling period $\sum_{\tau=1}^{\infty} \beta^{\tau} n(a_{t+\tau}, \sigma_{t+\tau}; \theta)$ is not observed. That is we only have complete data on firms who have exited. However if we were to select data only from those firms who do selloff before the end of the sample, we would be selecting firms on the basis of (a different function of) the same random draws that determine the residuals from our estimating equations, and hence incur the possiblity of a selection bias.

Let T be the final year of the sample, and note that for the firms still alive at T we can

[4]It is likely, however, that our method will be less sensitive to slight misspecifications in timing assumptions then is Euler equations. This is because our method is based on the returns for investment over the entire future, while Euler equation techniques are only concerned with returns over the period until the compensating perturbations is made. Since this if often a single period, a slight misspecification in the timing of decisions and outcomes could cause a significant "noise to signal" ratio, and make it hard to obtain reliable estiamtes from Euler equation techniques.

write

$$E\left[\sum_{\tau=1}^{\infty}\beta^{\tau}n(a_{i,t+\tau},\sigma_{i,t+\tau};\theta)|\sigma_{i,t+1}\right]=$$

$$E\left[\left(\sum_{\tau=1}^{T}\beta^{\tau}n(a_{i,t+\tau},\sigma_{i,t+\tau};\theta)|\sigma_{i,t+1}]+\beta^{T+1-t}V(\sigma_{i,T+1})\right)|\sigma_{i,t+1}\right].$$

Substituting into our moment condition (equation (3)), rearranging, and passing through an expectations conditional on $\sigma_{i,t}$, we have

$$E\left[\frac{\partial n(a_{i,t},\sigma_{i,t};\theta)}{\partial a_{i,t}}+\frac{\partial lnp(\sigma_{i,t+1}|\sigma_{i,t},a_{i,t};\theta)}{\partial a_{i,t}}\left(\sum_{\tau=1}^{T-t}\beta^{\tau}n(a_{i,t+\tau},\sigma_{i,t+\tau};\theta)+\beta^{T+1-t}V(\sigma_{i,T+1})\right)|\sigma_{i,t}\right]=0 \tag{8}$$

at $\theta=\theta_0$.

The truncation remainder problem is that $V(\sigma_{T+1})$ is not observed for those firms who survive until T, and the firms who survive until T are selected (in part) on the basis of the realizations of the disturbance terms in the estimation equations (the ϵ_t above). As a result the estimation procedure outlined in section 1 does not produce consistent instrumental variance estimates of θ (even when we restrict the analysis to the data on firms who exit before T)[5].

There are different ways of trying to account for this problem and the relative performance of these "solutions" is likely to depend on the type of data at hand. If our sample runs from $t=1,\ldots,T$, and T is large then we should be able to obtain consistent estimates of θ from the data on firms who were active in the early sample years by estimating $V(\sigma_{T+1})$ pointwise where needed, obtaining a different value for each firm which survives to the end of the sample period. If on, the other hand, T is short and our original consistency arguments were based on having data on a large number of firms (say N), then we might think of using a nonparametric estimate of $V(\sigma_{T+1})$ to obtain a semiparametric estimates of θ (especially if the number of variables in, or the dimension of, σ_T were small)[6].

Of course we do not know which of the possible procedures for accounting for the truncation remainder problem is likely to produce "better" estimates of θ , or indeed if there is a

[5]The truncation remainder problem is similar, though not identical, to the initial conditions problem discussed in Heckman($\cdot$) and some of the "solutions" to that problem discussed in the literature (see Pakes 1993) apply also here. On the other hand the fact that we can analyze the sources of the truncation remainder with the information in our data makes our problem somewhat easier to handle than the initial condition problem, see the discussion below.

[6]This is similar to the problem we would incur if we were to drop the smoothness assumption, i.e. assumption 2 above. In that case, provided the value function was differentiable, we could use the realized discounted future net cash flow to form a nonparametric estimate of the derivative of the value function, and then relate that estimate to the choice of the control.

need for an estimation procedure that worries about truncation at all, for samples with the size and characteristics of those we actually work with. What we will do is provide monte carlo evidence on both the severity of the biases the truncation remainder might generate and on the performance of alternative corrections for examples that ought to be of some interest to applied I.O. We turn to those examples now.

4 Example: A Monopolist Selling an Experience Good.

Our first example is a single agent monopoly problem. What differentiates our problem from standard monopoly problems is that we allow the distribution of future demand to depend on current demand, as is the case with "experience" goods. As noted above this will imply that the firm's optimal pricing strategy is "dynamic"; prices will be set below the prices that would maximize current profit because low prices induce larger future demand.

4.1 Primitives of the Problem.

Demand wil be given by $Q = MK(\omega)exp[-\alpha\mathbf{p}]$, where M denotes the number of consumers in the market, $K(\omega) = exp[\omega\kappa]/(1 + exp[\omega\kappa])$ is the fraction of consumers who are aware of the product, and $\mathbf{p}$ is its price. $K(\omega)$ is a montone increasing S-shaped function of ω. It sets the fraction of M who would purchase the good if price were zero, and it's evolution will depend on past demand. It is the only state variable of the problem which exhibits dependence over time.

From above

$$\omega = \kappa^{-1}log\left[\frac{(QM^{-1}exp[\alpha\mathbf{p}])}{(1 - QM^{-1}exp[\alpha\mathbf{p}])}\right],$$

so given price, quantity, and the parameter vector, we can compute ω.

We assume that

$$\omega_{t+1} = \omega_t + \tau_t - \xi_t,$$

with $\tau_t = 1$ with probability $aQ/[1 + aQ]$ and 0 elsewhere, while $\xi = 1$ with probability δ and 0 elsewhere. That is in any decision period (and there may be many decision periods per data period), ω_t can either go up one, stay the same, or go down one. The probability of it increasing is an increasing function of demand during the period.

Marginal cost will be independent of output, but repond to an independently and identically distributed factor cost $\{r_t\}$. That is

$$mc(r_t) = \lambda r_t$$

where λ is an unknown parameter of interest.

The scrap value of the firm is given by ϕ.

4.2 The Control Problem.

Since we are considerring a monopoly problem, it does not matter whether we have the firm choosing prices or quantities. For simplicity we choose quantities and write the Bellman equation for our problem as

$$V(r,\omega) = max\{\phi, sup_{Q\geq 0}(Q[\mathbf{p}(Q,\omega) - mc(r_t)] + \beta \sum_{r',\omega'} V(r',\omega')p(r')p(\omega'|\omega,Q))\}.$$

Assuming the firm produces positive output, the first order condition for optimal quantity $(Q^*(\omega,r))$ is

$$\mathbf{p}(r,\omega,Q^*) - \lambda r_t - \alpha + \beta \sum_{r',\omega'} V(r',\omega')p(r')\frac{\partial p(\omega'|\omega,Q)}{\partial Q}|_{Q=Q^*} = 0.$$

Second order conditions are also satisfied, so there is a unique $\mathbf{p}$ associated with each (r,ω) couple.

Note that with our demand equation $Q/[\partial Q/\partial \mathbf{p}]$, or the markup from a static model (in which quantity demanded today was not a determinant of demand tomorrow), is just the constant α. Since $p(\cdot|\cdot,Q)$ is stochastically increasing in Q, and $V(\cdot,\omega)$ is increasing in ω, our first order condition implies that price is always less than $\alpha+\lambda r$. That is the monopolist pricing an experience good always prices lower than what it would price if there was no "experience" component to demand.

In particular for our specification

$$\mathbf{p}(r,\omega) = \lambda r + \alpha - \beta\Delta(\omega)\frac{a}{(1+aQ)^2}$$

where

$$\Delta(\omega) \equiv \sum_{r'}\{[V(r',\omega+1) - V(r',\omega)](1-\delta) + [V(r',\omega) - V(r',\omega-1)]\delta\}p(r').$$

Since $V(\cdot)$ is increasing in ω, $\Delta(\omega)$ is always positive. The price is an decreasing function of the "slope" of the value function, and prices will increase or decrease in ω according as the value function is a convace or convex function of ω.

For every r there is an $\underline{\omega}(r)$ below which a firm exits. $V(\omega,r) = \phi$ for $\omega < \underline{\omega}(r)$ but is increasing in ω thereafter. So the value function starts out being a convex function of ω, and it remains so until reaching it single point of inflection, after which the value function turns concave and remains so thereafter (that is the value function "inherits" the S-shape of $K(\omega)$). Consequently price at low ω is quite high (near $\alpha + \lambda r$) and falls as we increase ω until the inflection point. In our specification the price at the inflection point is actually less than cost, implying a *negative markup*. The price then gradually rises as the value function approaches its asymptote, eventually approaching $\alpha + \lambda r$. Quantity will act in exactly the opposite fashion; it will be initially small, increase in ω over the convex portion of the value function and decrease thereafter. Recall that we have chosen our primitives so that if there were no relationship between current sales and future demand price would have been $\lambda r + \alpha$ (independent of ω). So the movements in price as ω changes are totally a reflection of the experience good nature of our product.

One final point. As one might expect

$$\frac{dQ}{dr} = \frac{-\lambda}{|\mathbf{p}'_Q| + |mk(\cdot)'_Q|} < 0,$$

where $|mk(\cdot)'_Q|$ is notation for the derivative of the markup with respect to Q. So when r increases Q goes down (though not as much as it would if there were no future effect in which case the second term in the denominator would be zero).

This implies that if we ignored the dynamic aspects of the problem and did the usual linear regression of $\mathbf{p}$ against r and a constant to estimate the marginal cost parameter (λ) and the markup (α), there would be a left out variable which is negatively related to r. Consequently we would expect this regression to produce an *underestimate* of λ and an *overestimate* of α. It would, therefore, underestimate costs and overestimate markups.

4.2.1 Parameters and Policy Output.

The parameters we choose are as follows: $M = 100,000$, $\kappa = .7$, $\alpha = .4$, $\lambda = 1$, r uniform over the set $[1.01, 1.02, \ldots 2.00]$, $\beta = .98$, $a = .001$, $\delta = .6$ and $\phi = 100$.

Note that the value of β implies that we are thinking of something like monthly, perhaps bimonthly, data, and recall that the value of α determines the static markup. When we computed the value function for these parameters we found that it increases from its shutdown value of 100 (which occurs at $\omega = -7$ for most r) to at most 200,000. Its point of inflection occurs at an ω of about one and a value of about 500. The markup lies between .4 and -.3 (at its point of inflection). This is to be comparred to a marginal cost which varies between 1 and 2. The markup is pretty close to .4 for $\omega \leq -5$ and for $\omega > 10$, and the trough in the markup as a function of ω is quite steep when $-5 < \omega < 5$.

When we started 100 firms in 100 markets each at an $\omega = -2$, and let each run until the firm exited, there were 18,206 market-period observations. About half the points have

$-5 < \omega < 5$, and almost all the rest have $\omega > 5$.

4.3 Estimation.

Throughout we shall assume that all parameters but λ and possibly ϕ are known (possibly from prior estimation of the demand system and the equations of motion for ω). Define

$$y(r_t, \omega_t; \lambda, \phi) \equiv \sum_{\tau=1} \beta^\tau n_{t+\tau}(\theta) = \sum_\tau \chi_\tau \beta^\tau \pi_{t+\tau}(\lambda) + \sum_\tau [\chi_{t+\tau} - \chi_{t+\tau-1}]\beta^\tau \phi,$$

where

$$\pi_{t+\tau}(\lambda) \equiv [p_{t+\tau} - \lambda r_{t+\tau}]Q_{t+\tau},$$

and the rest of the variables are defined as above.

Now set

$$z_t(\lambda, \phi) \equiv y_t \frac{\partial log[p(\omega_{t+1}|\omega_t, Q_t)]}{\partial Q_t}$$

a variable which is known conditional on parameter values, and use the same substitutions we did in section 1 to show that

$$p(r_t, \omega_t) = \lambda r_t + \alpha - z_t(\lambda, \phi) + \epsilon_t, \tag{9}$$

where

$$\epsilon_t \equiv z_t(\lambda, \phi) - E[z_t(\lambda, \phi)|\sigma_t] \ \Rightarrow E[\epsilon_t|\sigma_t] = 0.$$

So nonlinear instrumental variable procedures will produce consistent estimators of the parameters in equation (9). Next we consider the choice of instruments.

4.3.1 Instruments and the Distribution of the Estimator.

Using results from Chamberlain (1986), but temporarily ignoring both heteroscedasticity and serial correlation, the optimal instuments are given by

$$\frac{\partial E[\lambda r_t + \alpha - z_t(\lambda, \phi)|\sigma_0]}{\partial \theta}$$

which, for $\theta = \lambda$ is

$$r_t + E[\{\sum_{t+\tau} \chi_{t+\tau}\beta^\tau r_{t+\tau} Q_{t+\tau}\} \frac{\partial log[f(\omega_{t+1}|\omega_t, Q_t]}{\partial Q_t}|\sigma_t] \equiv r_t + h_1(\omega_t, r_t),$$

and for $\theta = \phi$ is

$$E[\{\sum_{t+\tau}[\chi_{t+\tau} - \chi_{t+\tau+1}]\beta^{\tau+1}\frac{\partial log[f(\omega_{t+1}|\omega_t, Q_t]}{\partial Q_t}|\sigma_t] \equiv h_2(\omega_t, r_t).$$

We employ a semiparametric estimator. In a first step we compute a nonparametric estimate of the regression function of (the future) terms from these two equations onto (ω_t, r_t) (in forming the z_t needed for this procedure we usee the true values of all the parameters). These estimates are then used to form our instruments . Since $\{r_t\}$ is i.i.d., it is not surprising that the estimated regression functions were essentially independent of r_t. Thus our instrument for λ at $(\omega = \omega^*, r)$ is r plus

$$\hat{h}_1(\omega^*) \equiv n(\omega^*)^{-1}\sum_{\omega=\omega^*}[\frac{\partial z_t(\omega, r; \lambda_0, \phi_0)}{\partial \lambda}]$$

where $n(\omega^*)$ is the number of observations with $\omega = \omega^*$, while the instrument for ϕ is

$$\hat{h}_2(\omega^*) \equiv n(\omega^*)^{-1}\sum_{\omega=\omega^*}[\frac{\partial z_t(\omega, r; \lambda_0, \phi_0)}{\partial \phi}].$$

We will be using data on a (pseudo) panel of firms, each operating in an independent monopoly market. Moreover, at least initially, we will not utilize the fact that the data on a given firm are serially correlated, nor will we utilize the fact that the variance in our moment condition depends on ω. That is the moment condition we minimize will be constructed by simply adding all the observed moment conditons for each firm, and then summing over firms. The vector of observations on each firm will be treated as a single "observation" from an i.i.d. sample when we compute the variance-covariance matrix of our estimator.

We start by only estimating λ. Letting i index firms and

$$\rho_{i,t}(\lambda) \equiv p_{i,t} - \lambda r_{i,t} + \alpha_0 - z_{i,t}(\lambda)$$

our moment is

$$m_{i,t} \equiv \rho_{i,t}(\lambda)(r_{i,t} + \hat{h}_{1,i,t})$$

where $\hat{h}_{1,i,t} \equiv \hat{h}_1(\omega^* = \omega_{i,t})$.

Let $\vec{m}_i(\lambda, \hat{h}_1)' = (m_{i,1}, \ldots, m_{i,T(i)})$ where $T(i)$ is the length of life of the firm, and $\vec{i}_i$ be a $T(i)$ vector of ones. Then

$$G_N(\lambda) \equiv N^{-1}\sum_i \vec{i}_i' \vec{m}_i(\lambda, \hat{h}_1)$$

and our estimator is obtained by minimizing $\| G_N(\lambda, \hat{h}_1) \|$.

At times we will also present reults from an estimator which does a correction for heteroscedasticity. These estimates have an additional step which estimates the variance of $m_{i,t}(\omega, \hat{h}_1)$ as a function of ω in a manner analagous to the way we estimated h_1. We then formed the moment condition for the $(i,t)^{th}$ observation by dividing $m_{i,t}$ by the square root of the estimated variance and minimized the norm of the sum of these moment conditions.

We will report monte carlo results from repeated independently drawn pseudo samples. The report will contain the mean and the standard error of the estimates across samples. This standard error will be referred to as the "exact" standard error as it is not subject to the approximation error we usually incur in using limit distributions for the parameter of interest (though it will still contain some sampling error). For comparison we will also provide the average of the standard errors that are calculated using asymptotic approximations. We develop the formula for the estimates of these standard errors now.

It is well known that the distribution of the semiparametric estimator for this problem is the same as the distribution of the estimator that would be obtained if the true instrument function were known (i.e. if we knew $h_1(\cdot)$ and so did not have to use the estimate $\hat{h}_1(\cdot)$; see Newey, ?). Consequently if we let

$$D \equiv E[\vec{i}_i' \frac{\partial \vec{m}_i(\lambda, h_1)}{\partial \lambda}]$$

where $h_1(\cdot)$ provides the true optimal instrument and

$$\mathcal{V} \equiv E[\vec{i}_i' \vec{m}_i(\lambda, h_1) \vec{m}_i(\lambda, h_1)' \vec{i}_i]$$

then standard arguments insure

$$\sqrt{N}(\hat{\lambda} - \lambda) \rightarrow N(0, D^{-1} \mathcal{V} D^{-1}).$$

We estimate this variance by substituting $\hat{D}$ and $\hat{V}$ into the above formula where

$$\hat{D} \equiv N^{-1} \sum_i \vec{i}_i' \frac{\partial \vec{m}_i(\hat{\lambda}, \hat{h}_1)}{\partial \lambda}$$

and

$$\hat{\mathcal{V}} \equiv N^{-1} \sum_i \vec{i}_i' \vec{m}_i(\hat{\lambda}, \hat{h}_1) \vec{m}_i(\hat{\lambda}, \hat{h}_1)' \vec{i}_i.$$

4.4 Estimates.

We present monte carlo from two separate sample designs. The first design follows fifty firms in independent markets from birth to death. This sample has no truncation remainder problem. The second sample will follow fifty monopoly markets for ninety periods (this would be seven and a half years of monthly data or fifteen years of bimonthly data). If the initial monopolist in a market exits, we provide a new entrant to the market in the following period. Thus in this sample there will be a truncation remainder for one firm in each market, but not for all firms we observe. As noted the firms for which there is no truncation remainder are selected as a function of the same random draws that determine the disturbances in the estimating equations.

For each sample design we computed our estimates for each of 500 independently drawn samples. Also, to insure that the variance of our estimates across samples was not a result of an inadequate search algorithm, we obtained the estimates for each sample by a simple grid search.

4.4.1 The First Sample Design.

The estimates for the first sample design, that is the sample with no truncation remainder, are given in Table 1. The top panel of this table presents the estimates of λ when we assume ϕ is known, the second panel presents the estimates of λ when we simultaneously estimate ϕ.

The sample of fifty firms had a median lifespan of seventy periods and a mean of one hundered and fifty nine (recall that we have set $\beta = .98$ so we should either be thinking of monthly or bimonthly data, depending on your view of discount rates). The disturbances on a single firm are highly correlated over time. Moreover as the median/mean comparison above indicates, the lifespan distribution is skewed to the right and the panel is far from being balanced.

Still the table indicates that if there is any small sample bias in the estimates it is small. The standard errors indicate that our estimates are also quite precise. The estimates in the top panel that use the "regular" instruments (ω, ω^2, r, π and a constant) has a standard error of 3% of the true value. When we use our approximation to the optimal instruments the standard error gets cut to 2%, and when we used an estimated weight also it got cut to 1%.

The second panel of the table indicates that whether we estimate ϕ or not does not make much difference to the estimate of λ. Moreover the standard errors in the bottom panel are not much different than those in the first panel, so the fact that we might need to estimate ϕ does not have an inordinate effect on the variance of the estimate of λ. That is the good news.

Interestingly, the weighting procedure tends to worsen rather than improve the standard error in the second panel. Similarly the second step of the GMM procedure tends to hurt, rather than improve, the precision of the estimates (the second step of the GMM procedure has no effect on the estimates in the first panel). Apparently the impact of the estimation

Table 1: Estimates of λ
(true value is one).

	average $\hat{\lambda}$	simulated[2] s.e.
Estimate only λ.		
"optimal"[1] no weight	1.04	.018
"optimal"[1] with weight	.99	.010
inst=$(cons, \omega, \omega^2, r, \pi)$ one step GMM	.99	.029
inst=$(cons, \omega, \omega^2, r, \pi)$ 2 step GMM	.99	.029
Estimate λ & ϕ; λ-estimates reported		
"optimal"[1], also estimate ϕ	1.04	.018
"optimal"[1] with weight and ϕ	.98	.045
inst=$(cons, \omega, \omega^2, r, \pi)$ also ϕ, one step GMM	.99	.029
inst=$(cons, \omega, \omega^2, r, \pi)$ also ϕ, two step GMM	.98	.033

[1] "Optimal" refers to a nonparametric approximation to the optimal instruments, and weighted means each observation was weighted by a nonparametric approximation of the inverse of its standard error, see the text for details.
[2] Simulated s.e. refers to the standard error of the estimate across monte carlo samples.

Table 2: OLS Estimates.

parameter	value	simulated s.e.
constant	.70	(.012)
λ	.65	(.0006)
OLS with ω, ω^2		
const	.62	.016
r	.66	.014
ω	-.62	.18
ω^2	.23	.034

error in the weights more than offsets any positive effect weighting might have on the precision of the estimates in this panel.

The bad news from the estimation underlying the second panel is that there is not much information on the value of ϕ in samples with the characteristics of those drawn here. The estimate of ϕ always had a standard error which was larger than the parameter estimate, and hence we do not report them here.

We now come back to the question which motivated this example. If we ignored the dynamic aspects of pricing and assumed the static pricing equation, what would have been our estimates of marginal costs (λ) and of the markup (the constant term)? Table 2 provides the answer; the estimate of λ would have fallen from one to about .65, and the estimate of the markup would have gone up to about .7 (recall that the static markup's true value is .4, and the dynamic model generates markups that are never higher than this value, so this is an overestimate of the markup of at least 75%). Moreover these are both estimated fairly precisely.

The next row does OLS again, this time putting in ω and ω^2 as a simple test for misspecification. That is, under the static pricing model the price depends only on r. However, a researcher who was aware of the possibility of dynamic pricing might worry that the price depended also on whether the firm was in a situation in which there was a large payoff to cutting price to increase future demand, ... Then it would be reasonable to add ω and ω^2 to the equation thinking that if there were a misspecification we should expect significant coefficients on these variables. We clearly do. Note however that since r and ω are pretty much independently distributed, the estimate of λ does not change much when we add the terms in ω.

On the whole the results are encouraging enough to go on. At least when we have data that does not have a truncation remainder problem, the estimates of λ produced by our technique are quite good. This is true regardless of whether we have to estimate the sunk cost parameter simultaneously, however the data do not contain much information on the actual value of sunk costs.

4.4.2 The Second Sample Design.

This is the sampling design that followed fifty monopoly markets over ninety periods, replacing an incumbent monopolist with a new entrant if the incumbent exited. The number of firms ever active varied between 66 and 104 over the subsamples, and averaged to 83. So, in almost all samples, there were more firms with truncation remainders than without.

To get the approximation to the optimal instruments we use only the observations from the first forty periods; i.e. for each ω we find all points that reach that ω during the first forty simulation periods and then take the average discounted future value until those firms exit as the instrument for that ω value. However all observations are used in the estimation; that is there is an observation for each firm that exits until the year that it exits.

Table 3 provides the estimates that correspond to those in table 1, that is the estimates from estimation techniques that make no correction for the truncation remainder. We did all these runs twice, once adding $T - t$ to the list of instruments in Table 1, and once using only those in Table 1. The results were almost identical in the two sets of runs so we only present those with $T - t$ added to the lists of instruments.

The estimates that use the optimal instruments are not very different from the estimates that use the subsample with no truncation remainder, either in value or in variance, at least when we don't weight. The estimates from the GMM procedure which uses standard instruments is not too bad either. The one estimator that seems a bit problematic because of its relatively large estimated variance is the estimator which relies on estimated weights. Apparently the combination of the variance in the weights (partly induced by imprecise estimates of those weights) and the truncation remainder can generates some samples with θ estimates that can be far from the truth (even though they do not appear to be biased in any direction).

Since we are now using a data set that corresponds to one we might actually use, this table also provides some information on the quality of alternative estimates of the standard errors. We take the simulated standard errors as the true standard errors (though they still contain some sampling error) and compare them to the average of the standard errors obtained from the traditional asymptotic expansions, and to the average of the standard errors obtained from a bootstrap on each monte carlo sample (fifty bootstrap samples were drawn for each monte carlo sample). The average of the asymptotic expansion was always lower than the simulated standard error (except for the weighted estimate, by between 10 and 25%), while the average of the bootstrap standard errors was always above the simulated

Table 3: Estimates of λ
(sample with truncation remainder).

	average $\hat{\lambda}$	simulated[2] s.e.	average[2] asy. s.e.	nonparametric bootstrap s.e.[2]	s.e. with D^* &$\mathcal{V}^{*2}$
"optimal"[1]+(T-t) no weight	1.02	.022	.017	.031	.018
"optimal"[1] +(T-t) with weight	1.03	.11	.025	.15	n.c.
inst=$(cons, \omega, \omega^2, r, \pi, T-t)$ one step GMM	.95	.032	.029	.045	.027
inst=$(cons, \omega, \omega^2, r, \pi, T-t)$ 2 step GMM	.95	.034	.029	.046	n.c.

[1] "Optimal" refers to a nonparametric approximation to the optimal instruments, and weighted means each observation was weighted by a nonparametric approximation of the inverse of its standard error, see the text for details.
[2] Simulated s.e. refers to the standard error of the estimate across monte carlo samples. Average asy. s.e. is the average of the estimates of the asymptotic approximation to the standard error. Nonparametric bootstrap, is the average of the nonparametric bootstrap standard errors across samples (there were fifty bootstrap samples drawn for each monte carlo sample). Finally the standard error with D^*&$\mathcal{V}^*$ is the asymptotic approximation with an accurate estimate of the derivative and variance matrix. n.c.=not calculated.

standard error (by 25 to 50%). The last column provides the standard errors we get from the asymptotic expansion when we use a very precise estimate of the derivative matrix (D) and the variance matrix $\mathcal{V}$ (these were obtained by combining the information from all fifty monte carlo samples). These are quite similar to the average of the asymptotic standard errors, so what error there is in the estimates of the variance obtained from the traditional expansion, is not likely to be caused from poor estimates of D or of $\mathcal{V}$ but rather from a more fundamental difference between the finite sample distribution and the asymptotic approximation.

On the whole the results are not very different than the results for the sample without a truncation remainder. On the other hand, were we just handed this data we would not know, a priori, that we did not have to worry about correcting for the truncation remainder. We now introduce a semiparametric technique which allows us to see if the truncation remainder is indeed a serious problem in a given data set, and then correct for it if it is.

Rewrite the pricing equation in (9) explicitly including the truncation remainder, i.e.

$$p(r_t, \omega_t) = \lambda r_t + \alpha - z_t^T(\lambda) + \chi_T \beta^{T+1} \frac{\partial log[p(\omega_{t+1}|\omega_t, Q_t)]}{\partial Q_t} V(\omega_T, r_T) + \epsilon_t, \tag{10}$$

where

$$z_t^T(\lambda) \equiv \frac{\partial log[p(\omega_{t+1}|\omega_t, Q_t)]}{\partial Q_t} (\sum_{\tau=1}^{T} \beta^\tau n_{t+\tau}(\theta)) =$$

$$\frac{\partial log[p(\omega_{t+1}|\omega_t, Q_t)]}{\partial Q_t} \left(\sum_{\tau=1}^{T} \chi_\tau \beta^\tau \pi_{t+\tau}(\lambda) + \sum_{\tau=1}^{T} [\chi_{t+\tau} - \chi_{t+\tau-1}] \beta^\tau \phi \right),$$

and, by construction,

$$E[\epsilon_t|\sigma_t] = 0.$$

Now use a semiparametric approximation for $V(\omega_T, r_T)$ and instrument the whole equation with functions of σ_t i.e. with functions of (ω_t, r_t, t). Note that potential instruments include t and σ_t, while the function we are trying to estimate is only a function of σ_T. This gives us the extra moments needed for identification.

We did two different semiparametric approximations for $V(\cdot)$, a polynomial in ω_T, r_T, and a "nonoverlapping kernel" (we break (ω_T, r_T) into nonoverlapping cells, and adding dummies for each cell). In each case the instruments were the corresponding variables for (ω_t, r_t) (terms in the polynomial expansion or dummies) and t, and the approximation to the optimal instrumetns.

We also computed two statistics to test for the need for adjusting our estimation procedure to account for the truncation remainder. The statistics were computed as follows. If λ^{nc} is the estimate of λ when we ignore the truncation remainder, and λ^c is the estimate when we

Table 4: Estimates with Truncation Corrections.

Estimator	coefficient value	simulated s.e.
Polynomial with $(constant, \pi(\cdot), \omega, \omega^2, r)$		
GMM	.99	.15
GMM with Weights	1.00	.13
2-stage GMM	1.01	.16
Eight Dummies		
GMM	.99	.12
GMM with Weights	1.01	.06
2-stage GMM	.99	.13

correct for it, then our first test is a test of whether $\lambda^{nc} - \lambda^c \approx 0$. We provide two estimates of the distribution of $(\lambda^{nc} - \lambda^c)/Var(\lambda^{nc} - \lambda^c)^{1/2}$. One is from the bootstrap. In this case the estimate of $Var(\lambda^{nc} - \lambda^c)$ is obtained from the empirical variance of $(\lambda^{nc} - \lambda^c)$ across bootstrapped samples. The other is from the asymptotic expansion to the moment conditions. In this case there is a different estimate of $Var(\lambda^{nc} - \lambda^c)$ for each sample taken from the asymptotic expansion for that sample. In addition we test whether the coefficients in the polynomial expansion can reasonably be assumed to be zero; again using both bootstraps and traditional asymptotic expansions. This latter test will pick up significant coefficients even if their presence does not effect the estimate of λ. Thus given that our only interest is in obtaining a good estimate of λ we focus on the first test.

Table 4 presents the results. It is clear that once we do the correction for the truncation remainder, what little bias there had been in the uncorrected estimates goes away. It is also clear however, that at least with the current sample design the correction increases the standard errors of the estimates of λ considerably, by a factor of four or more.