Inverse Problems and Nonparametric Identification of Dynamic Structural Models: The Case of Commodity Price Speculation

John Rust
University of Maryland

Inverse Problems in Econometrrics
Institute for Fiscal Studies
November, 2003

Basic statistical inference is an “inverse problem”

- **Observe:** $(\tilde{x}_1, \ldots, \tilde{x}_n)$ a random sample from unknown distribution $F(x)$.
- **Solution to inverse problem:** Form empirical CDF $\hat{F}(x)$

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I\{\tilde{x}_i \leq x\}$$

- As $n \to \infty$, $\hat{F} \to F$ with probability 1, so inverse problem is solved.

What if object of interest is the density, not the CDF?

- **Observe:** $(\tilde{x}_1, \ldots, \tilde{x}_n)$ a random sample from an unknown density $f(x)$.
- **Solution to inverse problem:** Form kernel density estimator $\hat{f}(x)$ with bandwidth $h_n \rightarrow 0$

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \phi\left(\frac{x - \tilde{x}_i}{h_n}\right)$$

  where ϕ is a kernel satisfying $\int_{-\infty}^{\infty} \phi(x)dx = 1$.
- As $n \rightarrow \infty$ $\hat{f}(x) \rightarrow f(x)$ with probability 1, so inverse problem is solved.

Why is the “Inverse Problem” trivial in these two simple examples?

- There is a $1:1$ mapping between the unknown object of interest and the object we can estimate non-parametrically.

$$f \longleftrightarrow f \quad F \longleftrightarrow F$$

- In the case of the CDF, there is a $1:1$ mapping between the unknown CDF F and the probabiliity limit of the nonparametric estimator, $\hat{F}$. This $1:1$ mapping means that we can use the “analogy principle” and use $\hat{F}$ as our estimator of F.
- In the case of the density, there is a $1:1$ mapping between the unknown density f and the probability limit of the non-parametric estimator $\hat{f}$. The $1:1$ mapping means that we can use $\hat{f}(x)$ as our estimator of $f(x)$.

However there are problems where the inverse problem is nontrivial

- **Example:** Structural estimation of discrete choice dynamic programming models.
- **Definition:** the *structural objects* are given by $\{\beta, u(x,a), p(x'|x,a), q(\varepsilon|x)\}$ where

$$P(a|x) = \frac{\partial}{\partial v(x,a)} G\left(\{v(x,a)|a \in A(x)\}\right)$$

where G is the *social surplus function*

$$G\left(\{v(x,a)|a \in A(x)\}\right) = \int \max_{a \in A(x)} \left[v(x,a) + \varepsilon(a)\right] q(d\varepsilon|x)$$

$$v(x,a) = u(x,a) + \beta \int G\left(\{v(x',a')|a' \in A(x')\}\right) p(dx'|x,a)$$

- **Theorem:** $\{x_t, a_t\}$ is a Markov process with transition density

$$\pi(x_{t+1}, a_{t+1} | x_t, a_t) = P(a_{t+1} | x_t) p(x_{t+1} | x_t, a_t)$$

- **Special case:** If $q(\varepsilon|x)$ is a multivariate Type III extreme value distribution with scale parameter σ, we have

$$G\left(\{v(x,a) | a \in A(x)\}\right) = \sigma \log \left[\sum_{a \in A(x)} \exp\{v(x,a)/\sigma\} \right]$$

- The choice probabilities are *multinomial logits*

$$P(a|x) = \frac{\exp\{v(x,a)/\sigma\}}{\sum_{a' \in A(x)} \exp\{v(x,a')/\sigma\}}$$

where

$$v(x,a) = u(x,a) + \beta \int \sigma \log \left[\sum_{a' \in A(x')} \exp\{v(x',a')/\sigma\} \right] p(dx'|x,a)$$

What is the inverse problem in this case?

- Note that we can estimate the transition density $\pi(x_{t+1}, a_{t+1}|x_t, a_t)$ non-parametrically (using kernel density estimator). Thus the *reduced form* is non-parametrically identified and estimable.
- Further we can non-parametrically estimate the components of the transition density π, i.e. $P(a|x)$ and $p(x'|x,a)$.
- It follows that the structure $\{\beta, u(x,a), p(x'|x,a), q(\varepsilon|x)\}$ is *partially non-parametrically identified*
- Can we estimate the remaining structural objects $\{\beta, u(x,a), q(\varepsilon|x)\}$ by "inverting" the choice probabilities $\{P(a|x)\}$?
- **Unfortunately, not in general.** *The discrete dynamic programming model is non-parametrically unidentified*

The Hotz-Miller Inversion Theorem

- **Theorem:** *There is a $1:1$ mapping between the conditional choice probabilities $\{P(a|x)|a \in A(x)\}$ and the normalized value functions $\{v(x,a) - v(x,0)|a \in A(x), 0 \in A(x)\}$ where $0 \in A(x)$ is some arbitrary standardied element in the choice sets, $A(x), x \in X$.*

- **Example:** In the extreme value case, we have

$$v(x,a) - v(x,0) = \log\left(P(a|x)/P(0|x)\right)$$

- Thus, we conclude that the normalized value functions are non-parametrically identified via the Hotz-Miller inverse mapping:

$$\hat{v}(x,a) - \hat{v}(0,x) = \log(\left(\hat{P}(a|x)/\hat{P}(0|x)\right)$$

- Can we identify the remaining structural objects $\{\beta, u(x,a), q(\varepsilon|x)\}$?

Rust's Non-Identification Result (*Handbook of Econometrics* vol. 4)

- **Theorem:** *Fix an arbitrary $\beta \in (0,1)$ and let $q(\varepsilon|x)$ be Type III extreme value. There is an equivalence class containing infinitely many (actually a continuum of) utility functions, $\Lambda(\beta,q,p,P)$, which are all consistent with the observed choice probabilities $\{P(a|x)\}$. A typical element of $\Lambda(\beta,q,p,P)$ is given by:*

$$u(x,a)+k(x)-\beta\int k(x')pdx'|x,a)$$

  where k is an arbitrary function.

- **Proof:** Since the inversion from choice probabilities to value functions only identifies the normalized value functions $\{v(x,a)-v(x,0)|a\in A(x)\}$, we only identify the individual value functions $\{v(x,a)|a\in A(x)\}$ up to an arbitrary function of x, $k(x)$. That is, the inverse mapping only identifies an equivalence class of value functions $\{v(x,a)+k(x)|a\in A(x)\}$.

Rust's Non-Identification Result (continued)

- All of the value functions in the equivalence class $\{v(x,a)+k(x)|a \in A(x)\}$ are consistent with the observed choice probabilities $\{P(a|x)|a \in A(x)\}$.
- From the Bellman equation, it follows that the equivalence class of utility functions consistent with this equivalence class of value functions is given by

$$\begin{aligned} v(x,a)+k(x) &= u(x,a)+k(x)-\beta\int k(x')p(dx'|x,a)+ \\ & \beta\int \sigma\log\left[\sum_{a'\in A(x')}\exp\{[v(x',a')+k(x')]/\sigma\}\right]p(dx'|x,a) \end{aligned}$$

- It follows that any utility function of the form

$$w(x,a)=u(x,a)+k(x)-\beta\int k(x')p(dx'|x,a)$$

is consistent with the observed choice probabilities $\{P(a|x)|a \in A(x)\}$.

Is there any hope for identification?

- Suppose there is a decision $0 \in A(x)$ that takes one into a permanent absorbing state where we can identify the value function $v(x,0)$ via a separate argument.
- **Example:** Suppose the agent is a firm and action $0 \in A(x)$ corresponds to a permanent exit or scrapping decision, and the exit or scrap valuecan be observed or consistently estimated under weak assumptions.
- In this case, knowledge of $v(x,0)$ enables us to identify all of the value functions non-parametrically using the Hotz-Miller inversion result:

$$v(x,a) = \log[P(a|x)/P(0|x)] + v(x,0) = v(x,a) - v(x,0) + v(x,0)$$

- Then, using the Bellman equation we can uniquely identify the utility functions (i.e. up to an additive constant k that does not depend on x)

$$v(x,a) = u(x,a) + \beta\sigma \int \log \left[\sum_{a' \in A(x')} \exp\{[v(x',a') + k(x')]/\sigma\} \right] p(dx'|x,a)$$

Inverse Mappings and Identification (continued)

- This suggests that (nearly) full non-parametric identification may be possible when the agents are firms rather than consumers, since with firms it can be possible to observe payoffs (i.e. expected profits vs. expected utilities).

- **Caveat:** The above result still assumed an arbitrary discount factor $\beta \in (0,1)$ and an arbitrary distribution of unobservables $\{q(\varepsilon|x)\}$ and there was a strong *additive separability, conditional independence* (ASCI) assumption on which previous results depend.

- **Observation:** More can be identified if we assume we have arbitrary power to *experiment* and make arbitrary modifications to agents' beliefs.

- **Proposition:** *If the econometrician can undertake arbitrary experiments on subjects, then the underlying structure* $\{\beta, u(x,a), p(x'|x,a), q(\varepsilon|x)\}$ *is non-parametrically identified.*

MOTIVATION

- New database on a single steel wholesaler (steel broker)
- Daily data on every transaction between July 1, 1997 to September 3, 1999 (550 business days)
- 2200+ products.
- We observe the quantity (both in units and in weight) of steel bought or sold, the sales price, the shipping costs, and the identity of the buyer or seller
- *We only observe purchase prices on days purchases occur.*

A Sketch of our Model

1. At the beginning of each day t the firm knows
   - q_t – inventory on hand,
   - p_t – per unit spot purchase price
   - z_t – observed state variable
   - ε_t – unobserved state variable
2. Given $(q_t, p_t, z_t, \varepsilon_t)$ the firm orders additional inventory $q_t^o \geq 0$ for immediate delivery.
3. Given (q_t, q_t^o, p_t, z_t) the firm sets a retail price p_t^r that is modeled as a random draw from a density $\gamma(p_t^r | q_t + q_t^o, p_t, z_t)$.
4. Given $(q_t, q_t^o, p_t, p_t^r, z_t, \varepsilon_t)$ the firm observes a realized retail demand for its steel, q_t^r, modeled as a draw from a distribution $H(q_t^r | p_t, p_t^r, z_t)$ with a point mass at $q_t^r = 0$.

5. The firm cannot sell more steel than it has on hand: $q_t^s = \min[q_t + q_t^o, q_t^r]$.
6. Inventories follow: $q_{t+1} = q_t + q_t^o - q_t^s$.
7. New values of (p_{t+1}, z_{t+1}) are drawn from $g(p_{t+1}, z_{t+1} | p_t, z_t)$.
8. A new value of the unobservable state variable ε_{t+1} is drawn from the density $\phi(\varepsilon)$.

Definition: *An (S,s) policy is a decision rule of the form:*

$$q^o(p,q,x) = \begin{cases} 0 & \text{if} \quad q \geq s(p,x) \\ S(p,x) - q & \text{otherwise} \end{cases} \tag{1}$$

where S and s are functions satisfying $S(p,x) \geq s(p,x)$ for all p and x.

Endogenous Sampling Rule:

$$p_t \quad \text{is observed} \quad \iff q^o_t > 0 \iff q_t < s(p_t, x_t).$$

Proposition 1: *The firm's optimal inventory investment policy $q^o(p,q,x)$ takes the form of an (S,s) rule. That is, there exist a pair of functions (S,s) satisfying $S(p,x) \geq s(p,x)$ where $S(p,x)$ is the desired or target inventory level and $s(p,x)$ is the inventory order threshold, i.e.*

$$q^o(p,q,x) = \begin{cases} 0 & \text{if} \quad q \geq s(p,x) \\ S(p,x) - q & \text{otherwise} \end{cases} \tag{2}$$

where $S(p,x)$ is given by:

$$S(p,x) = \textit{argmax}_{0 \leq q^o \leq \overline{q} - q} \left[W(p,q^o,z) - c^o(q^o,p,\varepsilon) \right] \tag{3}$$

and the lower inventory order limit, $s(p,x)$ is the value of q that makes the firm indifferent between ordering and not ordering more inventory:

$$s(p,x) = \inf_{q \geq 0} \{ q | W(p,q,z) - [p+\varepsilon]q \geq W(p,S(p,x),z) - [p+\varepsilon]S(p,x) - K \}. \tag{4}$$

where

$$W(p,q,z) = ES(p,q,z) - c^h(p,q,z) + \beta EV(p,q,z). \tag{5}$$

Euler Equation for optimal order quantity q^o

$$\frac{\partial W}{\partial q}(p,S(p,z,\varepsilon),z) - p = \varepsilon.$$

It is tempting to assume that ε has mean zero and use the Euler equation as a basis for GMM estimation of the parameters of the model.

Obstacles to the GMM/Euler equation approach:

1. There is no convenient analytical formula for the partial derivative of the value function, $\partial W/\partial q(p,S(p,z,\varepsilon),z)$.
2. As we show below, even if the unconditional mean of ε is zero, the conditional mean of ε over those values of (p,ε) for which it is optimal to purchase (i.e. for which $q < s(p,z)$), is generally nonzero.

3. The GMM approach has no easy way to deal with problems caused by endogenous sampling, and the fact that we observe purchases only an a relatively small subset of business days in our overall sample.

Lemma 1: *Let $\{\varepsilon_t\}$ be an IID process whose density ϕ is continuous and strictly positive over the entire real line. Then we have:*

$$
\begin{aligned}
f(q^o|p,q,z) = & \quad \frac{d}{dq^o}\int I\{q^o(p,q,z,\varepsilon)\phi(\varepsilon)d\varepsilon \\
= & \quad \begin{cases} \int_{s^{-1}(p,z,q)}^{\infty} \phi(\varepsilon)d\varepsilon & \text{if} \quad q^0 = 0 \\ \int_{-\infty}^{S^{-1}(p,z,\overline{q})} \phi(\varepsilon)d\varepsilon & \text{if} \quad q^0 = \overline{q} - q \\ \frac{-\phi(\partial W/\partial q(p,q+q^o,z)-p)}{\partial^2 W(p,z,q+q^o)/\partial^2 q} & \text{otherwise} \end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
S^{-1}(p,z,q) &= \inf\{\varepsilon|S(p,z,\varepsilon) = q\} \\
s^{-1}(p,z,q) &= \inf\{\varepsilon|s(p,z,\varepsilon) = q\}. \qquad (6)
\end{aligned}
$$

$$f(q^o|p,q,z) = -\phi(S^{-1}(p,z,q+q^o))\frac{\partial S^{-1}}{\partial q^o}(p,z,q+q^o). \tag{7}$$

$$\frac{\partial W}{\partial q}(p,S(p,z,\varepsilon),z) = p+\varepsilon, \tag{8}$$

$$\frac{\partial S^{-1}}{\partial q^o}(p,z,q+q^o) = \frac{1}{\partial S/\partial \varepsilon(p,z,S^{-1}(p,z,q+q^o))} = \frac{1}{\partial^2 W/\partial^2 q(p,q+q^o,z)}. \tag{9}$$

- Conditional density of next period inventory q_{t+1}

$$\mu(q'|p,p^r,q,q^o,z) = \begin{cases} (1-\eta(p^r,p,z))\delta(p^r,p,q+q^o,z) & \text{if} \quad q'=0 \\ \eta(p^r,p,z) & \text{if} \quad q'=q+q^o \\ (1-\eta(p^r,p,z))h(q+q^o-q^r|p^r,p,z) & \text{otherwise.} \end{cases} \tag{10}$$

- **Proposition:** $\{p_t, p^r_t, z_t, q_t, q^o_t\}$ is a Markov process with transition density

$$\begin{aligned} \rho(p_{t+1},p^r_{t+1},q_{t+1},q^o_{t+1},z_{t+1}|p_t,p^r_t,q_t,q^o_t,z_t,\theta) \quad = \quad & g(p_{t+1},z_{t+1}|p_t,z_t) \\ & \times\mu(q_{t+1}|p_t,p^r_t,q_t,q^o_t,z_t) \\ & \times f(q^o_{t+1}|p_{t+1},q_{t+1},z_{t+1}) \\ & \times\gamma(p^r_{t+1}|p_{t+1},q_{t+1}+q^o_{t+1},z_{t+1}). \end{aligned}$$

- **Full information maximum likelihood (FIML)**

$$l_f(\{p_t,p^r_t,q_t,q^o_t,z_t\}_{t=1}^T|p_0,p^r_0,q_0,q^o_0,z_0\theta) = \prod_{t=1}^{T} \rho(p_t,p^r_t,q_t,q^o_t,z_t|p_{t-1},p^r_{t-1},q_{t-1},q^o_{t-},z_{t-1} \tag{11}$$

Partial Information Maximum Likelihood Estimator (PIML)

- Convert the endogenous sampling problem into a *censored sampling problem* by defining an observed censored price sequence $\{p_t\}$ by:

$$p_t = \begin{cases} p_t^* & \text{if} \quad q_t^o > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Define the Partial Information Maximum Likelihood (PIML) estimator as:

$$\hat{\theta}_T^p = \underset{\theta \in \Theta}{\operatorname{argmax}}\, l_p(\{p_t, p_t^r, q_t, q_t^o, z_t\}_{t=1}^T | p_0, p_0^r, q_0, q_0^o, z_0, \theta),$$

where l_p is given by:

$$l_p(\{p_t, p_t^r, q_t, q_t^o, z_t\}_{t=1}^T | p_0, p_0^r, q_0, q_0^o, z_0, \theta) =$$
$$\int \cdots \int \prod_{t=1}^T \rho(p_t, p_t^r, q_t, q_t^o, z_t | p_{t-1}, p_{t-1}^r, q_{t-1}, q_{t-}^o, z_{t-1}, \theta) \prod_{t \notin T_n} dp_t.$$

and $T_n = \{t_1, \ldots, t_n\}$ are the days purchases occur, i.e, $q_t^o > 0$ for $t \in T_n$

Asymptotic Properties of the FIML and PIML estimators

- Let $\xi_t = (p_t, p_t^r, q_t, q_t^o, z_t)$.
- *Let the* **purchase set** Γ *be given by:*

$$\Gamma = \{\xi \,|\, q^o = 0\}$$

then the set of purchase dates $T_n = \{t_1, \ldots, t_n\}$ *can be defined recursively as:*

$$t_{i+1} = \inf\{t > t_i \,|\, \xi_t \in \Gamma\}\,.$$

- *Let* $\{\zeta_i\}$ *denote the* **embedded process** *associated with* $\{\xi_t\}$ *and* Γ*. This is the discrete time Markov process which is observed at successive purchase dates* $t \in T_n$*, i.e.,*

$$\{\zeta_i\} = \{\xi_{t_i}\} = \{\xi_t \,|\, \xi_t \in \Gamma\}\,.$$

- A simple application of the **Chapman-Kolmogorov equation** allows us to derive the transition density ν for the embedded process $\{\zeta_i\}$ as a $t_i - t_{i-1}$-step transition density for successive visits to the purchase set Γ.

Lemma 4: *The embedded process* $\{\zeta_i\}$ *is a Markov chain with transition density* ν *given by:*

$$\nu(\zeta_i|\zeta_{i-1}) = \rho(\xi_{t_i}|\xi_{t_{i-1}}) = \int \cdots \int \prod_{t=t_i+1}^{t_{i+1}} \rho(\xi_{t+1}|\xi_t) d\xi_{t_i+1} \cdots d\xi_{t_{i+1}-1}.$$

Definition 5: *Let* $\{\omega_i\}$ *be the* **segmented process** *associated with* $\{\xi_t\}$*, i.e. the process for which* ω_i *is defined as the realized values of* $\{\xi\}$ *for the sequence of* $t_i - t_{i-1}$ *time periods following the purchase at* t_{i-1} *until the purchase at* t_i*:*

$$\omega_i = (\xi_{t_{i-1}+1}, \ldots, \xi_{t_i}).$$

Notice that the number of components in the segment ω_i is a random variable, equal to the difference $t_i - t_{i-1}$ in the successive times that $\{\xi_t\}$ visits the purchase set Γ.

Lemma 5: *The segmented process* $\{\omega_i\}$ *is a Markov chain with transition density* ν *given by:*

$$\nu(\omega_i|\omega_{i-1},\theta) = \rho(\xi_{t_i},\xi_{t_i-1},\dots,\xi_{t_{i-1}+1}|\xi_{t_{i-1}},\theta) = \prod_{t=t_i+1}^{t_{i+1}} \rho(\xi_{t+1}|\xi_t,\theta).$$

Notice that due to the Markov property, only the last element of the segment ω_{i-1}, $\xi_{t_{i-1}}$, is needed to fully determine the conditional probability of $\omega_i = (\xi_{1+t_{i-1}},\dots,\xi_{t_i})$.

- Let $\tau = t_{i+1} - t_i$, be the **recurrence time** for successive visits to Γ.
- If the mean recurrence time to Γ is finite, $E\{\tau\} < \infty$, the process $\{\xi_t\}$ will visit Γ infinitely often and the number of visits n observed over any horizon T tends to infinity with probability 1 as $T \to \infty$.
- **Assumption 3:** *The Markov chain $\{\xi_t\}$ is ergodic (i.e. it possesses a unique stationary distribution), and the purchase set Γ is recurrent (i.e. $E\{\tau\} < \infty$), and the embedded and segmented processes $\{\zeta_i\}$ and $\{\omega_i\}$ are also ergodic Markov chains.*
- Rewrite l_f and l_p as follows:

$$l_f(\{p_t, p_t^r, q_t, q_t^o, z_t\}_{t=1}^T | p_0, p_0^r, q_0, q_0^o, z_0, \theta) = \prod_{i=1}^{n-1} \left[\prod_{t=t_i+1}^{t_{i+1}} \rho(\xi_t | \xi_{t-1}, \theta) \right],$$

$$l_p(\{p_t, p_t^r, q_t, q_t^o, z_t\}_{t=1}^T | p_0, p_0^r, q_0, q_0^o, z_0, \theta) = \prod_{i=1}^{n-1} \left[\int \cdots \int \prod_{t=t_i+1}^{t_{i+1}} \rho(\xi_t | \xi_{t-1}, \theta) d p_{t_i+1} \cdots d p_{t_i}$$

- By Assumption 3 and the Renewal Theorem for Markov chains (see, e.g. Resnick 1992), we have with probability 1

$$\lim_{T \to \infty} \frac{n}{T} = \frac{1}{E\{\tau\}}.$$

  Thus, as long as $E\{\tau\} < \infty$, the process $\{\xi_t\}$ visits Γ infinitely often and $n \to \infty$ with probability 1 as $T \to \infty$.

- We carry out the asymptotic analysis indexing the sample size by the number of purchases n rather than the total number of days T.

Consistency of the FIML and PIML estimators

- It is convenient to work with the normalized log-likelihood functions. Multiply and divide the likelihoods l_f and l_p by the product term

$$\prod_{i=1}^{n-1} Pr\left\{t_{i+1} - t_i | \xi_{t_i}, \theta\right\} = \prod_{i=1}^{n-1} \left[\int \cdots \int \prod_{t=t_i+1}^{t_{i+1}} I\left\{\xi_t \in \Gamma(\{t_{i+1}\})\right\} \rho(\xi_t | \xi_{t-1}, \theta) d\xi_t \right],$$

where

$$\Gamma(\{t_{i+1}\}) = \{\xi_t | q_t^o = 0, t \neq t_{i+1}\} \cup \{\xi_t | q_t^o > 0, t = t_{i+1}\}.$$

- Take logs and divide by $n-1$ to obtain for $j=f,p$

$$\begin{aligned}
&\frac{1}{n-1}\log l_j(\{p_t,p_t^r,q_t,q_t^o,z_t\}_{t=1}^T|p_0,p_0^r,q_0,q_0^o,z_0,\theta)\\
&\quad=\frac{1}{n-1}\sum_{i=1}^{n-1}\log\rho_j(\omega_{i+1}|\omega_i,t_{i+1}-t_i,\theta)+\frac{1}{n-1}\sum_{i=1}^{n-1}\log Pr\{t_{i+1}-t_i|\omega_i,\theta\}\\
&\quad=\frac{1}{n-1}\sum_{i=1}^{n-1}v_1(\omega_{i+1},\omega_i,\theta)+\frac{1}{n-1}\sum_{i=1}^{n-1}v_2(\omega_{i+1},\omega_i,\theta),
\end{aligned}$$

where

$$\rho_f(\omega_{i+1}|\omega_i, t_{i+1}-t_i, \theta) = \frac{\prod_{t=t_i+1}^{t_{i+1}} \rho(\xi_t|\xi_{t-1}, \theta)}{\int \cdots \int \prod_{t=t_i+1}^{t_{i+1}} I\{\xi_t \in \Gamma(\{t_{i+1}\})\} \rho(\xi_t|\xi_{t-1}, \theta) d\xi_t}$$

and

$$\rho_p(\omega_{i+1}|\omega_i, t_{i+1}-t_i, \theta) = \frac{\int \cdots \int \prod_{t=t_i+1}^{t_{i+1}} \rho(\xi_t|\xi_{t-1}, \theta) dp_{t_i+1} \cdots dp_{t_{i+1}-1}}{\int \cdots \int \prod_{t=t_i+1}^{t_{i+1}} I\{\xi_t \in \Gamma(\{t_{i+1}\})\} \rho(\xi_t|\xi_{t-1}, \theta) d\xi_t}.$$

- Note that both ρ_p and ρ_f are conditional densities for the segment ω_{i+1} given ω_i and the length of the segment $t_{i+1}-t_i$.
- Due to censoring, ρ_p can be regarded as the marginal density associated with ρ_f, where we have integrated out the unobserved prices over the interval $(t_i+1, t_i+2, \ldots, t_{i+1}-1)$ when no purchases occurred.
- Under suitable regularity conditions on the moments of the functions v_j, $j=1,2$, Assumption 3 and the Ergodic Theorem for Markov processes we have with probability 1 as $n \to \infty$

$$\frac{1}{n-1}\sum_{i=1}^{n-1} v_j(\omega_{i+1}, \omega_i, \theta) \Longrightarrow E\left\{v_j(\omega', \omega, \theta)\right\},$$

where the expectation is taken with respect to the invariant distribution for (ω', ω) and is given by

$$E\left\{v_j(\omega', \omega, \theta)\right\} = \int\int v_j(\omega', \omega, \theta) d\nu(\omega' | \omega, \theta^*) d\psi(\omega, \theta^*),$$

where $\nu(\omega'|\omega, \theta^*)$ is the transition density for the segmented process $\{\omega_i\}$ and $\psi(\omega, \theta^*)$ is its invariant distribution.

- Note that as $n \to \infty$ we have with probability 1

$$\frac{1}{n-1}\sum_{i=1}^{n-1} \log \rho_j(\omega_{i+1}|\omega_i, t_{i+1} - t_i, \theta) \Longrightarrow E\left\{\log \rho_j(\omega'|\omega, \tau, \theta)\right\},$$

and

$$\frac{1}{n-1}\sum_{i=1}^{n-1} \log Pr\left\{t_{i+1} - t_i|\xi_{t_i}, \theta\right\} \Longrightarrow E\left\{\log Pr\{\tau|\omega, \theta\}\right\},$$

where τ is the recurrence time to the purchase set Γ.

$$\begin{aligned}
&E\left\{\log \rho_j(\omega'|\omega, \tau, \theta)\right\} \\
&= E\left\{\log \rho(\xi_1, \ldots, \xi_\tau|\xi_0, \tau, \theta)\right\} \\
&= \int_{\xi_0} \sum_{\tau=1}^{\infty} \left[\int_{\xi_1} \cdots \int_{\xi_\tau} \log \rho(\{\xi_t\}_{t=1}^{\tau}|\xi_0, \tau, \theta) p(\{\xi_t\}_{t=1}^{\tau}|\xi_0, \tau, \theta^*) \prod_{t=1}^{\tau} d\xi_t\right] \mathrm{Pr}\{\tau|\xi_0, \theta^*\} d\psi(
\end{aligned}$$

- Given any ξ_0 and τ, the *Information Inequality* guarantees that the expression in brackets above is maximized at $\theta = \theta^*$. Similarly we have

$$E\left\{\log\left[Pr\left\{\tau|\xi_0,\theta\right\}\right]\right\} = \int_{\xi_0} \sum_{\tau=1}^{\infty} \log\left[\text{Pr}\{\tau|\xi_0,\theta\}\right] \text{Pr}\{\tau|\xi_0,\theta^*\} d\psi(\xi_0)$$

will also be maximized at $\theta = \theta^*$ given any (ξ_0).

- This implies that the limiting log likelihood is maximized at θ^*, and standard uniform consistency arguments can be used to show that with probability 1 we have $\hat{\theta}_n^j \to \theta^*$ as $n \to \infty$.

Asymptotic Normality of the FIML and PIML estimators

- If the model is correctly specified and appropriate regularity conditions hold, the first order conditions for $\hat{\theta}_n^p$ and $\hat{\theta}_n^f$ can be expanded in Taylor series about the true parameter θ^*.
- Applying a Central Limit Theorem for mixing processes to the key score term in this Taylor series expansion one can show that:

$$\sqrt{n}\left[\hat{\theta}_n^j - \theta^*\right] \longrightarrow N(0, I_j^{-1}(\theta^*)), \quad j = p, f$$

where

$$I_j(\theta^*) = I_j^1(\theta^*) + I^2(\theta^*)$$

and

$$I_j^1(\theta^*) = E\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\log\left[\rho_j(\omega'|\omega,\tau,\theta)\right]\right\} \quad \theta = \theta^*$$

and

$$I^2(\theta^*) = E\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\log\left[Pr\{\tau|\xi_0,\theta\}\right]\right\} \quad \theta = \theta^*.$$

- It is not difficult to show that since ρ_p is a marginal density of ρ_f, we must have that the difference between the informations $I_f^1(\theta^*) - I_P^1(\theta^*)$ is a positive semi-definitive matrix.
- Thus, it is not surprising that there is a loss of information, and therefore an increase in variance, caused by the endogenous sampling problem.

Drawbacks of the PIML estimator

- Computational burden due to the high dimensional integrations that are required to evaluate l_p.
- Since no purchases of steel are observed on the majority of business days in our sample, the mean time between sales is about 10 business days, so that on average 10 dimensional integrals must be calculated for each term entering the likelihood.
- Although there have been important advances in simulation estimation and low discrepancy methods for computing high dimensional integrals (see, e.g. Rust, Traub and Woźniakowski, 1999), the PIML will still be a fairly computationally burdensome estimator.
- A second drawback is that if our interest is primarily on making inferences about the law of motion for $\{p_t, z_t\}$, the other structural parameters that must

be estimated to adjust for the endogeneity of the sampling process amount to nuisance parameters.

- Errors in the specification of the firm's optimal investment and speculation problem will result in inconsistent estimates of the parameters of interest in the transition density $g(p_{t+1}, z_{t+1} | p_t, z_t)$.

Identification issues:

- It is possible to consider the use of flexible reduced-form specifications for the densities entering the overall decomposition of the transition density ρ given in Theorem 3.
- However without some strong prior parametric restrictions on some of these densities, it is doubtful that an unrestricted model where the densities (g, μ, f, γ) are treated as unknown objects to be estimated non-parametrically is even identified.
- The (S, s) model combined with the observations of retail transaction prices provides strong identifying restrictions, limiting how far the wholesale price process $\{p_t\}$ can drift away from observed retail price for a given sequence of observed purchases.

- In particular, as the implied markup gets larger or smaller, the (S,s) model predicts that the number of orders should be increasing and decreasing in a corresponding fashion. Given the observed sequence of purchases, this property enables us to separately identify the parameters of $g(p',z'|p,z)$ and the structural parameters of the (S,s) model.
- A non-parametric model does not impose any sort of profit maximizing or loss minimizing behavioral motivation on the part of the intermediary, so it appears that the wholesale market price $\{p_t\}$ could drift arbitrarily far away from the retail prices $\{p_t^r\}$ without there being any strong effect on the likelihood of the observed sequences of purchases. This suggests that it is impossible to non-parametrically identify the form of $g(p',z'|p,z)$ and the trading rule used by the firm when we only have access to endogenously sampled data.

Simulated Minimum Distance (SMD) Estimation

- Let $\{\xi_t\}$ denote the censored process (i.e. with $p_t = 0$ when $q_t^o = 0$), and let θ denote the $K \times 1$ vector of parameters to be estimated.
- The SMD estimator is based on finding a parameter value that best fits a $J \times 1$ vector of moments of the observed process:

$$h_T \equiv \frac{1}{T} \sum_{t=1}^{T} h(\xi_t, \xi_{t-1}), \quad J \geq K.$$

- By Assumption 2, the process $\{\xi_t\}$ is ergodic so that, with probability 1, h_T converges to a limit $E\{h(\xi', \xi)\}$ where the expectation is taken with respect to the ergodic distribution of (ξ', ξ).
- Under suitable additional regularity conditions, a central limit theorem will hold for h_T, i.e. we have

$$\sqrt{T}\,[h_T - E\{h\}] \Longrightarrow N(0, \Omega(h)),$$

where

$$\Omega(h) = E\left\{(h(\xi',\xi) - E\{h\})(h(\xi',\xi) - E\{h\})'\right\}.$$

- Now assume it is possible to generate simulated realizations of the $\{\xi_t\}$ process for any candidate value of θ, and that this process is censored in exactly the same way as the observed $\{\xi_t\}$ process is censored, i.e., with $p_t = 0$ when $q_t^o = 0$.
- The simulations depend on a $T \times 1$ vector, u, of *IID* $U(0,1)$ random variables that are drawn once at the start of the estimation process and held fixed thereafter in order for the estimator to satisfy stochastic equicontinuity conditions necessary to establish asymptotic normality of the SMD estimator.
- We will consider simulated processes of the form

$$\{\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)\}, \quad t = 2, \ldots, T$$

where for each $t > 1$, $\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)$ is a continuously differentiable function of θ.

- The notation $\{u_s\}_{s\le t}$ reflects the fact that the simulated process is *adapted* to the realization of the $\{u_t\}$ process.

- Note that we allow the simulated process to depend on the first value ξ_0 of the observed data as an initializing condition.

Examples of smooth simulators: Unidimensional case $(\xi_t \in R^1)$

- Let $\rho(\xi_{t+1}|\xi_t,\theta)$ denote its transition density and $P(\xi_{t+1}|\xi_t,\theta)$ be the corresponding conditional CDF.
- The first value of the simulated process is simply set to the observed value ξ_0.
- Using the probability integral transform, define $\xi_1(u_1,\theta,\xi_0)$ by:

$$\xi_1(u_1,\theta,\xi_0) = P^{-1}(u_1|\xi_0,\theta).$$

  Clearly $\xi_1(u_1,\theta,\xi_0)$ will be a continuously differentiable function of θ if $P^{-1}(u_1|\xi_0,\theta)$ is a continuously differentiable function of θ.

- Now define recursively for $t = 2,4,\ldots$

$$\xi_t(\{u_s\}_{s\le t},\theta,\xi_0) = P^{-1}(u_t|\xi_{t-1}(\{u_s\}_{s\le t-1},\theta,\xi_0),\theta).$$

We can see recursively that $\xi_t(\{u_s\}_{s\leq t}, \theta, \xi_0)$ will be a continuously differentiable function of θ provided that $P^{-1}(u|\xi, \theta)$ is a continuously differentiable function of ξ and θ.

Multidimensional case $\xi_t = (p_t, p_t^r, q_t, q_t^o, z_t)$.

- We can do a similar simulation as in the univariate case described above, using a factorization of the transition density of $\{x_t\}$ into a product of univariate conditional densities such as given in Theorem 3. For example, if ξ_t has two components, $\xi_t = (\xi_{1,t}, \xi_{2,t})$, suppose that its transition density ρ can be factored as

$$\rho(\xi_{t+1}|\xi_t,\theta) = \rho_2(\xi_{2,t+1}|\xi_{1,t+1},\xi_t,\theta)\rho_1(\xi_{1,t+1}|\xi_t,\theta),$$

  with corresponding conditional CDFs denoted by P_1 and P_2.

- Generate simulations of $\{\xi_t\}$ that are smooth function of θ just as in the univariate case, except that in the two-dimensional case we need to generate two random $U(0,1)$ variables $u_t = (u_{1,t}, u_{2,t})$ for each time period simulated.

- To generate $\xi_1 = (\xi_{1,1}, \xi_{2,1})$ we compute

$$\begin{aligned}\xi_{1,1} &= P_1^{-1}(u_{1,1}|\xi_0, \theta) \\ \xi_{2,1} &= P_2^{-1}(u_{2,1}|\xi_{1,2}, \xi_0, \theta).\end{aligned}$$

  Clearly the resulting realization for ξ_1 is of the form $\xi_1(u_1, \xi_0, \theta)$ and will be a smooth function of θ provided that P_1 and P_2 are smooth functions of (ξ, θ).

- Continuing recursively we have:

$$\begin{aligned}\xi_{1,t+1} &= P_1^{-1}(u_{1,t+1}|\xi_t, \theta) \\ \xi_{2,t+1} &= P_2^{-1}(u_{2,t+1}|\xi_{1,t+1}, \xi_t, \theta).\end{aligned}$$

  The resulting simulations take the form $\{\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)\}$ and will be smooth functions of θ provided that P_1 and P_2 are smooth functions of (ξ, θ).

Construction of simulated moments

- Use a single simulated realization of $\{\xi_t(\{u_s\}_{s\leq t},\theta,\xi_0)\}$ to form a simulated sample moment $h_T(\{u_s\}_{s\leq T},\xi_0,\theta)$ given by

$$h_T(\{u_s\}_{s\leq T},\xi_0,\theta) = \frac{1}{T}\sum_{t=1}^{T} h\left(\xi_t(\{u_s\}_{s\leq t},\theta,\xi_0),\xi_{t-1}(\{u_s\}_{s\leq t-1},\theta,\xi_0)\right).$$

  Let $(\{u_s^1\}_{s\leq T},\ldots,\{u_s^S\}_{s\leq T})$ denote S *IID* $T\times 1$ sequences of $U(0,1)$ random vectors used to generate the S independent realizations of the endogenously sampled process $\{\xi_t(\{u_s^i\}_{s\leq t},\theta,\xi_0)\}$, $i=1,\ldots,S$.

- Define $h_{S,T}(\theta)$ as the average of S independent time averages $h_T(\{u_s^i\}_{s\leq T},\xi_0,\theta)$

$$h_{S,T}(\theta) = \frac{1}{S}\sum_{i=1}^{S} h_T(\{u_s^i\}_{s\leq T},\xi_0,\theta).$$

Definition: the simulated minimum distance estimator $\hat{\theta}_T$ is defined by:

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}} (h_{S,T}(\theta) - h_T)' W_T (h_{S,T}(\theta) - h_T),$$

where W_T is a $J \times J$ positive definite weighting matrix.

- Note: the SMD estimator is in the class of *Simulated Moments Estimators (SME)* studied by Lee and Ingram (1991) and Duffie and Singleton (1993).
- **Assumption 4:** *the parametric model is correctly specified, i.e., there is a* $\theta^* \in \Theta$ *such that:*

$$\{\xi_t(\{u_s\}_{s \leq t}, \theta^*, \xi_0)\} \sim \{\xi_t\}$$

  i.e. when $\theta = \theta^*$, the simulated sequence initialized from the observed value ξ_0 has the same probability distribution as the observed sequence $\{\xi_t\}$.
- On the “to-do” list: work out asymptotics of the SMD estimator in the misspecified case.

Sketch of the derivation of the asymptotic distribution of the SMD estimator.

- **Assumption 5:** *For any* $\theta \in \Theta$ *the process* $\{\xi_t(\{u_s\}_{s \le t}, \theta, \xi_0)\}$ *is ergodic with unique invariant density* $\psi(\xi|\theta)$ *given by:*

$$\psi(\xi'|\theta) = \int \rho(\xi'|\xi,\theta) d\psi(\xi|\theta).$$

- Define the functions $E\{h|\theta\}$, $\nabla E\{h|\theta\}$, and $\nabla h_{S,T}(\theta)$ by:

$$\begin{aligned} E\{h|\theta\} &= \int h(\xi',\xi) d\rho(\xi'|\xi,\theta) d\psi(\xi|\theta) \\ \nabla E\{h|\theta\} &= \frac{\partial}{\partial\theta} E\{h|\theta\} \\ \nabla h_{S,T}(\theta) &= \frac{\partial}{\partial\theta} h_{S,T}(\theta). \end{aligned}$$

- **Assumption 6:** θ^* *is* **identified**, *i.e. if* $\theta \neq \theta^*$ *then* $E\{h|\theta\} \neq E\{h|\theta^*\} = E\{h\}$. *Furthermore, rank*$(\nabla E\{h|\theta\}) = K$ *and*

$\lim_{T\to\infty} W_T = W$ with probability 1 where W is a $J \times J$ positive definite matrix.

- **Consistency of the SMD estimator** Under appropriate regularity conditions the simulated process is uniformly ergodic, i.e., with probability 1 we have

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \left|(h_{S,T}(\theta) - h_T)' W_T (h_{S,T}(\theta) - h_T) - (E\{h|\theta\} - E\{h|\theta^*\})' W (E\{h|\theta\} - E\{h|\theta^*\})\right| \tag{12}$$

- Assumption 5 guarantees that the unique minimizer of $(E\{h|\theta\} - E\{h|\theta^*\})' W (E\{h|\theta\} - E\{h|\theta^*\})$ is θ^*, and this combined with the uniform consistency result implies the consistency of $\hat{\theta}_T$.

Asymptotic normality of $\hat{\theta}_T$

- Do a Taylor series expansion of the first order condition for $\hat{\theta}_T$:

$$(h_{S,T}(\hat{\theta}_T) - h_T)' W_T \nabla h_{S,T}(\hat{\theta}_T) = 0.$$

expanding $h_{S,T}(\hat{\theta}_T)$ about $\theta = \theta^*$ to get

$$h_{S,T}(\hat{\theta}_T) = h_{S,T}(\theta^*) + \nabla h_{S,T}(\tilde{\theta}_T)(\hat{\theta}_T - \theta^*),$$

where $\tilde{\theta}_T$ denotes a vector that is (elementwise) on the line segment between $\hat{\theta}_T$ and θ^*. Substituting this into the first order condition for $\hat{\theta}_T$ and solving we obtain:

$$(\hat{\theta}_T - \theta^*) = -\left[\nabla h_{S,T}(\tilde{\theta}_T)' W_T \nabla h_{S,T}(\hat{\theta}_T)\right]^{-1} \nabla h_{S,T}(\hat{\theta}_T)' W_T \left[h_{S,T}(\theta^*) - h_T\right],$$

Multiplying both sides of this equation by $\sqrt{T}$ and invoking the Central Limit theorem for the difference $\sqrt{T}[h_{S,T}(\theta^*) - h_T]$ we obtain

$$\sqrt{T}\left[h_{S,T}(\theta^*) - h_T\right] \Longrightarrow N\left(0, (1 + 1/S)\Omega(h, \theta^*)\right).$$

- Note that $h_{S,T}(\theta^*)$ is an average of S independent realizations of $\{\xi_t(\{u_s\}_{s\leq t}, \theta, \xi_0)\}$, which by Assumption 4 has the same distribution as $\{\xi_t\}$.
- As a result each of the terms entering $h_{S,T}(\theta^*)$, $h_T(\{u^i\}, \theta^*)$, has the same probability distribution as h_T and are distributed independently of h_T. The Central Limit Theorem applied to h_T yields

$$\sqrt{T}\left[h_T - E\{h|\theta^*\}\right] \Longrightarrow N\left(0, \Omega(h, \theta^*)\right).$$

- Similarly, for each $i = 1, \ldots, S$ we have

$$\sqrt{T}\left[h_T(\{u_s^i\}_{s\leq T}, \theta^*) - E\{h|\theta^*\}\right] \Longrightarrow N\left(0, \Omega(h, \theta^*)\right).$$

Note that

$$\left[h_{S,T}(\theta^*) - h_T\right] = \left[\frac{1}{S}\sum_{i=1}^{S}\left[h_T(\{u_s^i\}_{s\leq T}, \theta^*) - E\{h|\theta^*\}\right] + E\{h|\theta^*\} - h_T\right],$$

so that we have

$$\sqrt{T}\left[h_{S,T}(\theta^*) - h_T\right] \Longrightarrow \left[\frac{1}{S}\sum_{i=1}^{S}\tilde{X}_i + \tilde{X}_0\right],$$

where $(\tilde{X}_0, \tilde{X}_1, \ldots, X_S)$ are *IID* $N(0, \Omega(h, \theta^*))$ random vectors. It follows that

$$\sqrt{T}\left[h_{S,T}(\theta^*) - h_T\right] \Longrightarrow N(0, (1 + 1/S)\Omega(h, \theta^*)).$$

- Thus, we have

$$\sqrt{T}\left[\hat{\theta}_T - \theta^*\right] \Longrightarrow N(0,(1+1/S)\Lambda_1^{-1}\Lambda_2\Lambda_1^{-1}),$$

where

$$\begin{aligned}\Lambda_1 &= [\nabla E\{h|\theta^*\}]' W \nabla \left[E\{h|\theta^*\}\right] \\ \Lambda_2 &= [\nabla E\{h|\theta^*\}]' W \Omega(h,\theta^*) W \left[\nabla E\{h|\theta^*\}\right].\end{aligned}$$

- The optimal weight matrix $W = [\Omega(h,\theta^*)]^{-1}$ results in an SMD estimator with minimal variance. In this case the asymptotic distribution of $\hat{\theta}_T$ simplifies to:

$$\sqrt{T}[\hat{\theta}_T - \theta^*] \Longrightarrow N(0,(1+1/S)\Lambda^{-1})$$

where:

$$\Lambda = \left[\nabla E\{h|\theta^*\}'[\Omega(h,\theta^*)]^{-1}\nabla E\{h|\theta^*\}\right].$$

- Note that the penalty to forming an SMD estimator using only a single realization $S=1$ of the endogenously sampled process $\{\xi_t(\{u_s\}_{s\leq t},\theta,\xi_0)\}$

is fairly small, the variance of the resulting estimator is only twice as large as an estimator that computes the expectation of $h_T(\{u\}, \theta)$ exactly, such as would be done via Monte Carlo integration when $S \rightarrow \infty$.

- The SMD estimator can be implemented in practice by solving

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}} (h_{S,T}(\theta) - h_T)' \left[\hat{\Omega}(h,\theta)\right]^{-1} (h_{S,T}(\theta) - h_T),$$

where

$$\hat{\Omega}(h,\theta) = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t(\theta) \varepsilon_t(\theta)'$$

where

$$\varepsilon_t(\theta) = h\left(\xi_t(\{u_s\}_{s \le t}, \theta, \xi_0), \xi_{t-1}(\{u_s\}_{s \le t-1}, \theta, \xi_0)\right) - h_T(\{u_s\}_{s \le T}, \xi_0, \theta).$$

- Thus, an estimate of the optimal weighting matrix $\Omega(h,\theta)$ is recomputed each time the parameter θ is updated.

How to select the moments h?

- More efficient estimators can be obtained by selecting “efficient moment functions” h such as the score of the partial information maximum likelihood function derived in section 2. Such an estimator can attain the Cramer-Rao efficiency bound derived for the PIML estimator. However the score involves a ratio of integrals, and it is not clear that these integrals can be replaced by simulation estimates and still obtain a consistent SMD estimator.
- If accurate numerical integrals are required, the computational advantage of the SMD estimator is lost and it may be less computationally burdensome to compute the PIML estimator directly. This is a topic for future work.
- The definition of the SMD estimator can be extended to allow moments formed from the segmented Markov chain $\{\omega_i\}$. This formulation would be required in the case where h is the score of the partial information likelihood

function, since the components of the score involve the segmented chain as shown above.

- Using moments from the segmented chain involves some minor modifications of the arguments given above. We now do the asymptotics as a function of the number of purchases n rather than the total number of time periods T over which the process is observed. In this case we define the sample moments h_n by

  $$\frac{1}{n-1}\sum_{i=1}^{n-1} h(\omega_{i+1}, \omega_i),$$

  and the simulated moments $h_{S,n}(\theta)$ can be defined accordingly, using the simulated process $\{\xi_t(\{u_s^i\}_{s\leq t}, \theta, \xi_0)\}$, $i = 1, \ldots, S$ to construct S *IID* realizations of the segmented process.

Relaxing the assumption that the parametric model is correctly specified.

- As long as assumptions 5 and 6 hold, there will still exist well defined limiting moments for the simulated process, $E\{h|\theta\}$, for each $\theta \in \Theta$.
- Define θ^* as the value that minimizes the distance between the simulated model and the true data generating process:

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[E\{h|\theta\} - E\{h\}\right]' W \left[E\{h|\theta\} - E\{h\}\right],$$

  where $E\{h\}$ denotes the limit of h_T as $T \to \infty$ for the true data generating process.
- If the value of θ^* that minimizes this distance is interior to the parameter space Θ, then the following first order condition must hold at θ^*:

$$(E\{h|\theta^*\} - E\{h\})' W \nabla E\{h|\theta^*\} = 0,$$

where $E\{h\}$ denotes the long run or ergodic expectation of h with respect to the true data generating process.

- This implies that as $t \longrightarrow \infty$ the random vector

$$X_t \equiv \nabla E\{h|\theta^*\}' W(h(\xi_t(\{u_s\}_{s\leq t}, \theta^*, \xi_0)) - h(\xi_t)),$$

satisfies

$$\begin{aligned} \lim_{t\to\infty} E\{X_t\} &= 0 \\ \lim_{t\to\infty} \text{cov}(X_t) &= \Lambda_2 \end{aligned}$$

for some $J \times J$ covariance matrix Λ_2.

- Note that in the misspecified case, Λ_2 may not equal the same formula as the Λ_2 when the model is correctly specified.
- Using a suitable Central Limit theorem for mixing processes, we should have

$$\sqrt{T}\nabla E\{h|\theta^*\}'W\left[h_T(\{u\},\theta^*)-h_T\right] \Longrightarrow N(0,\Lambda_2).$$

Following a Taylor expansion argument just as in the correctly specified case above, we should be able to derive the same general form for the asymptotic distribution of $\hat{\theta}_T$ in the misspecified case, i.e.

$$\sqrt{T}\left[\hat{\theta}_T-\theta^*\right] \Longrightarrow N(0,(1+1/S)\Lambda_1^{-1}\Lambda_2\Lambda_1^{-1}),$$

where

$$\Lambda_1 = \left[\nabla E\{h|\theta^*\}'W\nabla E\{h|\theta^*\}\right]$$

and where

$$\sqrt{T}\nabla h_T(\{u_s\},\theta^*)'W_T\left[h_T(\{u_s\},\theta^*)-h_T\right] \Longrightarrow N(0,\Lambda_3).$$

- The main outstanding issue is to actually establish the limiting asymptotic distribution that we conjectured above, and relate the asymptotic covariance matrix Λ_3 to the asymptotic covariance matrix Λ_2. The similarity of the two expressions suggests that $\Lambda_2 = \Lambda_3$, but further work, including careful attention to regularity conditions, would required to determine whether this is the case.

Non-parametric identification of the commodity speculation model.

- We now consider a version of the commodity price speculation with endogenous price setting and price discrimination.
- We consider which parts of the model might be identified non-parametrically.
- Since this is a firm (expected profit maximizer, profits potentially observable) the hope is that more will be identified than would be for a consumer (expected utility maximizer, utility generally not observable).

Definition: The *structural objects* of the commodity speculation model include:

- The transition probability for the wholesale price $g(p_{t+1}, x_{t+1} | p_t, x_t)$
- The firm's beliefs about the arrival rate of buyers, $\eta(p, x)$
- Beliefs about the quantity demanded conditional on arrival, $m(q^d | p, x)$
- Beliefs about other buyer chacteristics given arrival and q^d, $\mu(z | q^d, p, x)$

Structural objects of the commodity speculation model, (continued)

- Beliefs about the reservation prices of buyers conditional on arrival, q^d and z, $f(r|q^d, p, x, z)$
- The cost of holding inventory, $c^h(p, q, x)$
- The fixed cost of placing an order to restock inventory, K
- Beliefs of per unit costs of “lost goodwill” due to unfilled orders, γ
- The firm’s interest rate, r with implied discount factor $\beta = \exp(-r/365)$

Definition: The *reduced-form objects* of the commodity speculation model are:

- The optimal purchasing rule, defined by the functions $S(p,x)$ and $s(p,x)$ given by

$$q^o(p,q,x) = \begin{cases} 0 & \text{if} \quad q \geq s(p,x) \\ S(p,x) - q & \text{otherwise} \end{cases}$$

- The optimal pricing rule, given by

$$p^r = p^r(p,q,x,z,q^d) = c(p,q,x) + \frac{1 - F(p^r|q^d,p,x,z)}{f(p^r|q^d,p,x,z)}$$

Non-parametric estimation/identification of the reduced-form

- Argue that order threshold, $s(p,x)$, is non-parametrically identified. Observe whether or not firm orders each period, so estimate non-parametric binary choice model for binary choice model $I\{q < s(p,x)\}$.
- Argue that optimal target inventory level, $S(p,x)$, is non-parametrically identified via a non-parametric regression on the order sizes on the subset of the days that the firm purchases,

$$q^0 = q^o(p,q,x) = S(p,x) - q$$

- Argue that the optimal pricing rule, $p^r(p,q,x,z,q^d)$ is non-parametrically identified via a non-parametric regression using observed retail prices on the subset of the days that the firm sells *and* purchases,

$$p^r = p^r(p,q,x,z,q^d)$$

Which structural objects can be inverted/estimated from the reduced-form?

- Inferring beliefs about customer reservation values. Note that when $q < s(p,x)$ we have

$$c(p,q,x) = p$$

and

$$p^r = p + \frac{1 - F(p^r | q^d, p, x, z)}{f(p^r | q^d, p, x, z)}$$

- Thus, using data on the subset of days that the firm sells and buys and where $q < s(p,x)$ we observe $p^r - p$ and can thus “recover” the markup term in the right hand side of the equation above.

- Thus, via non-parametric regression we can estimate the markup function $e(p^r, q^d, p, x, z)$

Inverting structural objects from the reduced-form (continued)

- Using the markup function, we can solve the ODE to get the conditional density of reservation prices

$$f(r|q^d,p,x,z) = F'(r|q^d,p,x,z) = \frac{1-F(r|q^d,p,x,z)}{e(r,q^d,p,x,z)}$$

- Fixing (q^d,p,x,z), the solution to this ODE is

$$F(r|q^d,p,x,z) = 1-\exp\{-\int_0^r [e(u,q^d,p,x,z)]^{-1}du\}$$

- Thus, we argue that the CDF and thus the density of consumer reservation values can be recovered from (i.e. inverted from) the reduced form, and is thus non-parametrically identified.
- This strategy has been used in auction theory (Vuong *et. al.* and in recent work by Pesendorfer and Jofre-Bonet on dynamic auctions.

Inverting structural objects from the reduced-form (continued)

- What about the distribution of quantity demanded by a customer, conditional on arrival, $m(q^d|p,x)$?
- Note that we observe quantity sold, q_t^s, a censored observation of q_t^d

$$q_t^s = \min[q_t + q_t^o, q_t^d]$$

- Notice the censoring threshold, $q_t + q_t^o$, is quantity on hand, exhibits quite a lot of variability over time. We do not have fixed censoring.
- When $q_t^d < q_t + q_t^o$ we have $q_t^d = q_t^s$
- We observe whether or not q_t^d is censored: q_t^d is censored *iff* $q_{t+1} = 0$ (i.e. a stockout occurs).

Inverting structural objects from the reduced-form (continued)

- **Conjecture:** *If there is no upper bound on possible inventory levels, $m(q^d|p,x)$ is non-parametrically identified, using a time series version of semi-parametric sample selection models. (cites?)*
- **Non-parametric identification of the arrival probability** We only observe whether or not firm makes a sale: we do not observe cases where customer arrives and is quoted a price, but decides not to take it.
- We can non-parametrically estimate the conditional probability that a sale occurs, i.e. conditional on (p,p^r,q^d,x,z), $Pr\{q_t^s > 0|p,p^r,q^d,x,z\}$ via non-parametric binary choice model on the binary event $I\{q_t^s > 0\}$.
- However from the model we know that the probability of a sale is given by

$$Pr\{q_t^s|p,p^r,q^d,x,z\} = \eta(p,x)\left[1-F(p^r|q^d,p,x,z)\right]$$

since we can estimate the left hand side non-parametrically, and $F(r|q^d,p,x,z)$ non-parametrically, we can unocver $\eta(p,x)$ as a ratio of these two conditional probabilities.

Inverting structural objects from the reduced-form (continued)

- **Non-parametric identification of the conditional distribution of customer characteristics,** $\mu(z|q^d,p,x)$ **conditional on arrival and** q^d
- This probabiility is conditional on *arrival.* We can non-parametrically estimate a related conditional probability $\mu_S(z|q^d,p,p^r,x)$ conditional on arrival *and* a sale occurs at the quoted price p^r

$$\mu_S(z|q^d,p,p^r,x,q^s>0) = \frac{\mu(z|q^d,p,x)}{[1-F(p^r|q^d,p,x,z)]}$$

- Thus, since we can non-parametrically estimate the left hand side of the above equation, and we can non-parametrically estimate the denominater term on the right hand side, we can non-parametrically estimate of $\mu(z|q^d,p,x)$ as a product of the other two objects,

$$\mu(z|q^d,p,x) = [1-F(p^r|q^d,p,x,z)]\mu_S(z|q^d,p,p^r,x,q^s>0)$$

Inverting structural objects from the reduced-form (continued)

- Which of the structural objects is left? $\{r, K, \gamma, c^h(p,q,x), g(p',x'|p,x)\}$
- Estimating the “forcing process” $g(p_{t+1}, x_{t+1}|p_t, x_t)$. This is a key part of the speculation model, reflects the firm’s beliefs about the equilibrium dynamics in the wholesale market.
- This would not be a problem if we always observe $\{p_t, x_t\}$, but we only observe p_t on days the firm buys, not on days it does not buy.
- That is, we have

$$p_t \quad \text{is observed} \quad \textit{iff} \quad q_t < s(p_t, x_t)$$

- We call this the *endogenous sampling problem.*