Pesendorfer and Schmidt-Dengler (2007): Asymptotic Least Squares Estimators for Dynamic Games

Jason Blevins
Duke University Applied Microeconomics Reading Group

June 11, 2008

Outline

- Considers the class of asymptotic least squares estimators for dynamic games.
- Estimation is based on equilibrium conditions.
- Discuss identification and provide sufficient conditions for exact identification.
- Characterize the efficient asymptotic least squares estimator.
- Several well-known estimators are members of this class.
- Monte Carlo experiments.

Framework

- Dynamic games in discrete time with $t \in 1, \ldots, \infty$.
- N players, $K+1$ actions, L states per player, common discount factor β.
- States: $s_{i,t} \in S_i = \{1, \ldots, L\}$, $\varepsilon_{i,t} \sim F(\varepsilon \,|\, s_{i,t}, s_{-i,t})$ on $\mathbb{R}^K$. Let $S = S_1 \times \cdots \times S_N$.
- The payoff shocks $\varepsilon_{i,t}$ are private information, independent across players and time, and independent of the actions of other players.
- Actions $a_{i,t} \in A_i = \{0, 1, \ldots, K\}$ are made simultaneously. Let $A = A_1 \times \cdots \times A_N$.
- State transitions follow some density $g(a_t, s_t, s_{t+1})$. Let G denote the $m_a m_s \times m_s$ matrix of these probabilities where $m_s = \#S = L^N$ and $m_a = \#A = (K+1)^N$.
- Period payoffs are given by

$$\pi_i(a_t, s_t) + \sum_{k=1}^{K} \varepsilon_{i,t,k} \mathbb{1}\{a_{i,t} = k\}$$

Markov Perfect Equilibria

Consider pure Markovian strategies $a_i(s_t, \varepsilon_{i,t})$. A collection $(a, \sigma) = (a_1, \dots, a_N, \sigma_1, \dots \sigma_N)$ is a *Markov perfect equilibrium* if

1. for all i, a_i is a best response to a_{-i} given the beliefs σ_i at all states $s \in S$,
2. all players use Markovian strategies,
3. for all i the beliefs σ_i are consistent with the strategies a.

Define the ex-ante Value function

$$V_i(s;\sigma_i) = \sum_a \sigma_i(a\,|\,s)\left[\pi_i(a,s) + \beta \sum_{s'} g(a,s,s')V_i(s';\sigma_i)\right] + \sum_{k=1}^{K} \mathbb{E}_\varepsilon\left[\varepsilon_{i,k}\,|\,a_i = k\right]\sigma_i(a_i = k\,|\,s).$$

We can write this in matrix notation as

$$\begin{aligned} V_i(\sigma_i) &= \sigma_i \Pi_i + D_i(\sigma_i) + \beta \sigma_i G V_i(\sigma_i) \\ &= [I - \beta \sigma_i G]^{-1}[\sigma_i \Pi_i + D_i(\sigma_i)]. \end{aligned}$$

Equilibrium Characterization

The continuation value net of payoff shocks under a_i with beliefs σ_i is

$$u_i(a_i;\sigma_i,\theta) = \sum_{a_{-i}} \sigma_i(a_{-i}\,|\,s)\left[\pi_i(a_{-i},a_i,s) + \beta \sum_{s'} g(a_{-i},a_i,s,s')V_i(s';\sigma_i)\right].$$

It is optimal to choose a_i under the beliefs σ_i if

$$u_i(a_i;\sigma_i,\theta) + \varepsilon_{i,a_i} \geq u_i(a_i';\sigma_i,\theta) + \varepsilon_{i,a_i'} \quad \forall a_i' \in A_i.$$

Ex ante, in expectation we have

$$p(a_i\,|\,s,\sigma_i) = \Psi_i(a_i,s,\sigma_i;\theta) = \int \mathbb{1}\left\{u_i(a_i;\sigma_i,\theta) - u_i(k;\sigma_i,\theta) \geq \varepsilon_{i,k} - \varepsilon_{i,a_i}, k \neq a_i\right\} dF.$$

In matrix notation we have a $(N \cdot K \cdot m_s) \times 1$ system

$$p = \Psi(\sigma;\theta).$$

Equilibrium Properties

In equilibrium, beliefs are consistent and we have the fixed point problem

$$p = \Psi(p; \theta). \tag{1}$$

Thus, finding an equilibrium is a fixed point problem on $[0, 1]^{N \cdot K \cdot m_s}$.

Proposition 1. *In any Markov perfect equilibrium, the probability vector* p *satisfies* (1)*. Conversely, any* p *that satisifes* (1) *can be extended to a Markov perfect equilibrium.*

Theorem 1. *A Markov perfect equilibrium exists.*

We have the same results under symmetric equilibria: existence and necessary and sufficient conditions. Symmetry reduces the number of equations in (1) and thus the computational complexity.

Identification

The model is identified if there exists a unique set of model primitives $(\Pi_i, \ldots, \Pi_N, F, \beta, g)$ that generate any particular set of choice and state transition probabilities.

- Time series data $\{a_t, s_t\}_{t=1}^T$.
- Suppose the data allow us to characterize $p(a \,|\, s)$ and $g(a, s, s')$.
- Fix β and F. There are $m_a \cdot m_s \cdot N$ remaining unknowns in $(\Pi_1, \ldots, \Pi_N)$.

Proposition 2. *Suppose F and β are given. Then at most $K \cdot m_s \cdot N$ parameters can be identified.*

There are only $K \cdot m_s \cdot N$ equations in the equilibrium conditions but $m_a \cdot m_s \cdot N$ parameters. We need at least $(m_a \cdot m_s - K \cdot m_s) \cdot N$ restrictions in order to identify all parameters.

Identification: A Linear Representation

There is some $\bar{\varepsilon}_i^{a_i}(s)$ that makes player i indifferent between actions a_i and 0:

$$\sum_{a_{-i}\in A_{-i}} p(a_{-i}\,|\,s)\left[\pi_i(a_{-i},a_i,s)+\beta\sum_{s'\in S} g(a_{-i},a_i,s,s')V_i(s';p)\right]+\bar{\varepsilon}_i^{a_i}(s)$$
$$=\sum_{a_{-i}\in A_{-i}} p(a_{-i}\,|\,s)\left[\pi_i(a_{-i},0,s)+\beta\sum_{s'\in S} g(a_{-i},0,s,s')V_i(s';p)\right]$$

From before, $V_i(\sigma_i)=[I-\beta\sigma_i G]^{-1}[\sigma_i\Pi_i+D_i(\sigma_i)]$. Thus, we have a linear system of equations for player i:

$$X_i(p,g,\beta)\Pi_i+Y_i(p,g,\beta)=0$$

where X_i is a $(K\cdot m_s)\times(m_a\cdot m_s)$ matrix and Y_i is a $(K\cdot m_s)\times 1$ vector, both of which depend on the choice probabilities, transition probabilities, and β.

Identification: Linear Restrictions

Consider player i. Let R_i be a $(m_a \cdot m_s - K \cdot m_s) \times (m_a \cdot m_s)$ matrix of restrictions and let r_i be a $(m_a \cdot m_s - K \cdot m_s) \times 1$-dimensional vector such that $R_i \Pi_i = r_i$.

We can now form an augmented linear system of $m_a \cdot m_s$ equations in $m_a \cdot m_s$ unknowns (hence the order condition is satisfied):

$$\begin{bmatrix} X_i \\ R_i \end{bmatrix} \Pi_i + \begin{bmatrix} Y_i \\ r_i \end{bmatrix} = \overline{X}_i \Pi_i + \overline{Y}_i = 0.$$

Proposition 3. *Consider any player i and suppose that F and β are given. If $\operatorname{rank}(\overline{X}_i) = m_a \cdot m_s$, then Π_i is exactly identified.*

Example: Consider the following restrictions:

$$\begin{aligned} \pi_i(a_i, a_{-i}, s_i, s_{-i}) &= \pi_i(a_i, a_{-i}, s_i, s'_{-i}) \qquad && \forall a \in A, (s_i, s_{-i}) \in S, (s_i, s'_{-i}) \in S \\ \pi_i(0, a_{-i}, s_i) &= r_i(a_{-i}, s_i) && \forall a_{-i} \in A_{-i}, s_i \in S_i \end{aligned}$$

The first is an exclusion restriction while the second is an exogeneity restriction (e.g., payoffs for inactive firms are known to be zero). If $L \geq K + 1$, then these restrictions ensure identification (provided that the rank condition holds).

Asymptotic Least Squares Estimators

Let $\theta = (\theta_\pi, \theta_F, \beta, \theta_g) \in \Theta \subset \mathbb{R}^q$ be the parameters of interest.

There are also $H \leq (N \cdot K \cdot m_s) + (m_a \cdot m_s \cdot m_s)$ auxiliary parameters $p(\theta)$ and $g(\theta)$, related to θ through the $N \cdot K \cdot m_s$ equations

$$h(p, g, \theta) = p - \Psi(p, g, \theta) = 0. \tag{2}$$

Asymptotic least squares estimators (Gourieroux and Monfort, 1995, Section 9.1) proceed in two steps:

1. Estimate the auxiliary parameters p and g.

2. Estimate the parameters of interest using weighted least squares using (2) as estimating equations.

Asymptotic Least Squares Estimators

Assume that consistent and asymptotically normal estimators of p and g are available such that as $T \to \infty$,

$$(\hat{p}_T, \hat{g}_T) \longrightarrow (p(\theta_0), g(\theta_0)) \quad a.s.,$$

$$\sqrt{T}\,[(\hat{p}_T, \hat{g}_T) - (p(\theta_0), g(\theta_0))] \xrightarrow{d} \mathrm{N}(0, \Sigma(\theta_0)).$$

The estimation principle involves choosing θ in order to satisfy the constraints

$$h(\hat{p}_T, \hat{g}_T, \theta) = \hat{p}_T - \Psi(\hat{p}_T, \hat{g}_T, \theta) = 0.$$

Let W_T be a symmetric positive-definite weight matrix of dimension $(N \cdot K \cdot m_s) \times (N \times K \times m_s)$. The asymptotic least squares estimator corresponding to W_T is defined as

$$\tilde{\theta}_T(W_T) = \arg\min_{\theta}[\hat{p}_T - \Psi(\hat{p}_T, \hat{g}_T, \theta)]^\top W_T[\hat{p}_T - \Psi(\hat{p}_T, \hat{g}_T, \theta)].$$

Asymptotic Least Squares Estimators: Assumptions

1. Θ is a compact set.

2. θ_0 lies in the interior of Θ.

3. As $T \to \infty$, $W_T \to W_0$ a.s. where W_0 is a non-stochastic positive definite matrix.

4. θ satisfies

$$[p(\theta_0) - \Psi(p(\theta_0), g(\theta_0), \theta)]^\top W_o [p(\theta_0) - \Psi(p(\theta_0), g(\theta_0), \theta)] = 0$$

   implies that $\theta = \theta_0$.

5. The functions π, g, and F are twice continuously differentiable in θ.

6. The matrix $[\nabla_\theta \Psi(p(\theta_0), g(\theta_0), \theta_0)]^\top W_o [\nabla_\theta \Psi(p(\theta_0), g(\theta_0), \theta_0)]$ is nonsingular.

Asymptotic Least Squares Estimators: Properties

Proposition 4. *Given the assumptions above the asymptotic least squares estimator* $\tilde{\theta}_T(W_T)$ *exists,* $\tilde{\theta}_T(W_T) \xrightarrow{a.s.} \theta_0$*, and as* $T \to 0$*,*

$$\sqrt{T}\left(\tilde{\theta}_T(W_T) - \theta_0\right) \xrightarrow{d} \mathrm{N}\left(0, \Omega(\theta_0)\right)$$

where

$$\begin{aligned}\Omega(\theta_0) = \left(\nabla_\theta \Psi^\top W_0 \nabla_{\theta^\top}\right)^{-1} \nabla_\theta \Psi^\top W_0 \left[\begin{pmatrix} I & 0 \end{pmatrix} - \nabla_{(p,g)^\top}\Psi\right] \Sigma \\ \cdot \left[\begin{pmatrix} I & 0 \end{pmatrix} - \nabla_{(p,g)^\top}\Psi\right]^\top W_0 \nabla_{\theta^\top}\Psi \left(\nabla_\theta \Psi^\top W_0 \nabla_{\theta^\top}\right)^{-1}\end{aligned}$$

where 0 *is the* $(N \cdot K \cdot m_s) \times (m_a \cdot m_s \cdot m_s)$ *zero matrix and the various matrices are evaluated at* θ_0, $p(\theta_0)$, *and* $g(\theta_0)$.

Efficient Asymptotic Least Squares

Proposition 5. *Under the maintained assumptions, the best asymptotic least squares estimators exist. They correspond to sequences of matrices* W_T^* *converging to*

$$W_0^* = \left([\begin{pmatrix} I & 0 \end{pmatrix} - \nabla_{(p,g)'}\Psi]\Sigma[\begin{pmatrix} I & 0 \end{pmatrix} - \nabla_{(p,g)'}\Psi]^\top \right)^{-1}.$$

Their asymptotic covariance matrices are

$$\left(\nabla_\theta \Psi^\top \left([\begin{pmatrix} I & 0 \end{pmatrix} - \nabla_{(p,g)'}\Psi]\Sigma[\begin{pmatrix} I & 0 \end{pmatrix} - \nabla_{(p,g)'}\Psi]^\top \right)^{-1} \nabla_{\theta^\top}\Psi \right)^{-1}$$

Here, 0 *denotes a* $(N \cdot K \cdot m_s) \times (m_a \cdot m_s \cdot m_s)$ *matrix of zeros.*

Asymptotic Least Squares: Moment Estimator

The moment estimator proposed by Hotz and Miller (1993) is an asymptotic least squares estimator with a particular weight matrix.

Let T_{is} denote the set of observations for individual i in state s and let $\alpha_{is} = (\alpha_1, \ldots, \alpha_K)$ be a vector of indicators for each choice (with zero omitted).

The moment condition is

$$\mathbb{E}\left[Z \otimes (\alpha_{is} - \Psi_{is}(\hat{p}_T, \hat{g}_T, \theta))\right] = 0$$

where Z is a $J \times 1$-dimensional vector of instruments.

Suppose $Z_t = Z_{is}$. Then the corresponding sample analog becomes

$$\frac{1}{NT} \sum_{\substack{1 \le i \le N \\ s \in S}} \sum_{t \in T_{is}} Z_t \otimes (\alpha_t - \Psi_{is}(\hat{p}_T, \hat{g}_T, \theta)) = \frac{1}{NT} \sum_{\substack{1 \le i \le N \\ s \in S}} n_{is} \left[Z_{is} \otimes (\hat{p}_{is} - \Psi_{is}(\hat{p}_T, \hat{g}_T, \theta))\right].$$

Thus, the moment estimator in this case is an asymptotic least squares estimator with estimating equation $\hat{p} - \Psi(\hat{p}_T, \hat{g}_T, \theta) = 0$.

Asymptotic Least Squares: Pseudo Maximum Likelihood

The pseudo maximum likelihood estimator of Aguirregabiria and Mira (2002, 2007) is also an asymptotic least squares estimator.

The partial pseudo log-likelihood, conditional on estimates $\hat{g}_T$ is

$$\ell = \sum_{s \in S} \sum_{i=1}^{N} \sum_{k \in A_i} n_{kis} \ln \Psi_{kis}(\hat{p}_T, \hat{g}_T, \theta).$$

The first order condition is

$$\frac{\partial \ell}{\partial \theta} = (\nabla_\theta \Psi^\top)\, \Sigma_p^{-1}(\Psi)\, [\hat{p} - \Psi(\hat{p}_T, \hat{g}_T, \theta)]$$

where $\Sigma_p^{-1}(\Psi)$ is the inverse covariance matrix of the choice probabilities.

This is equivalent to the first order condition of the asymptotic least squares estimator with weight matrix $W_T^{ml} \xrightarrow{p} \Sigma_p^{-1}$.

Monte Carlo Study

- Compare LS-E, PML, LS-I, and k-PML.
- A simple two player, two action, two state, game with five equilibria.
- Three equilibria are used for experiments with various sample sizes.
- LS-E estimator performs best overall (in eight of 12 experiments).
- LS-E performs poorly with the smallest sample size ($T = 100$).
- PML ranks second (by MSE) in seven of 12 specifications.
- PML performs better than LS-E for $T = 100$ and worse for larger sample sizes. This may be because the covariance matrix of $(\hat{p}_T, \hat{g}_T)$ is estimated better than the efficient weight matrix for small T.
- PML may be less computationally burdensome for large state spaces.

References

Aguirregabiria, V. and Mira, P. (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica*, 70:1519–1543.

Aguirregabiria, V. and Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 75:1–53.

Gourieroux, C. and Monfort, A. (1995). *Statistics and Econometric Models*, volume 1. Cambridge University Press.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60:497–529.