ALTERNATIVE APPROACHES TO ESTIMATION OF DYNAMIC NONLINEAR MODELS

Related to work with Elie Tamer and Ekaterini Kyriazidou

Consider

$$y_{it} = 1\left\{x_{it}'\beta + y_{i,t-1}\gamma + \alpha_i + \varepsilon_{it} \geq 0\right\}$$
$$t = 1, \ldots, T, \quad i = 1, \ldots, N$$

Interested in estimation of (γ, β).

N Large. T small.

So will consider asymptotics with $N \to \infty$ for T fixed.

More generally

$$f\left(y_{it} \middle| y_{it-1}, \ldots, y_{it-k}, x_i, \alpha_i\right)$$

Fully parametric approach ("random effects"):

Specify $f(\alpha_i)$ and $f(y_{i1}|\, x_i, \alpha_i)$

$$\mathcal{L} = \int f(y_{i1}|\, x_i, \alpha_i) \prod_{t=2}^{T} f(y_{it}|\, y_{it-1}, x_i, \alpha_i)\, f(\alpha_i)\, d\alpha_i$$

But what is

$$f(y_{i1}|\, x_i, \alpha_i) \quad ?$$

With stationarity and time–invariant x_i, one can sometimes solve for it. But these are strong assumptions.

Less parametric approach ("fixed effects approach"):

For $T = 4,$ consider distribution of y_{i2} given $(y_{i1}, y_{i2} + y_{i3} = 1, y_{i4}, x_{i3} = x_{i4})$. This is informative about (β, γ) without assumptions on α_i.

For example if ε_{it} is i.i.d. logistic, then

$$P(y_{i2} = 1 | y_{i1}, y_{i2} + y_{i3} = 1, y_{i4}, x_{i3} = x_{i4})$$
$$= \frac{\exp((x_{i2} - x_{i3})\beta + \gamma(y_{i1} - y_{i4})}{1 + \exp((x_{i2} - x_{i3})\beta + \gamma(y_{i1} - y_{i4}))}$$

which does *not* depend on α_i.

This (and other) observations can be used to estimate (γ, β).

Problems

- Often impossible
  - No General Approach
- Sometimes weak results when possible.
  - Matching.
    - Asymptotics similar to that in nonparametric regression.
    - Not known how to deal with discrete explanatory variables such as time–dummies or trends.
- Interesting?
  - Cannot calculate marginal effects.

Back to basics

$$y_{it} = 1\left\{x_{it}'\beta + y_{i,t-1}\gamma + \alpha_i + \varepsilon_{it} \geq 0\right\}$$

Let $p_0\left(\alpha, x^T\right) = P\left(y_{i0} = 1 \middle| x_i^T, \alpha_i\right)$ and let θ be all the parameters of the model (incl. parameters in distribution of ε_{it} and α_i).

The set of $(p_0(\cdot,\cdot), \theta)$ that are consistent with the data–generating process, is

$$\left\{(p_0(\cdot,\cdot), \theta) : P\left(\pi\left(A; p_0\left(\cdot, x^T\right), \theta\right) = P\left(A \middle| x^T\right)\right) = 1 \text{ for all } A\right\}$$

and the sharp bounds on θ is given by

$$\{\theta : \exists p_0(\cdot,\cdot) \text{ such that } P\left(\pi\left(A; p_0\left(\cdot, x^T\right), \theta\right) = P\left(A \middle| x^T\right)\right) = 1 \text{ for all } A\}$$

The identified region is the solution to a number of optimization problems.

For example

$$\min_{p_0(\cdot,\cdot),\theta} E\left[w\left(x^T\right)\left\|\pi\left(\mathcal{A};p_0\left(\cdot,x^T\right),\theta\right)-P\left(\mathcal{A}\middle|\, x^T\right)\right\|\right]$$

where $\mathcal{A}$ is the set of all outcomes.

Or

$$\max_{p_0(\cdot,\cdot),\theta} E\left[w\left(x^T\right)\log\left(\pi\left(y_i;p_0\left(\cdot,x^T\right),x^T,\theta\right)\right)\right] =$$

$$\max_{p_0(\cdot,\cdot),\theta} \qquad E\left[\log\left(\int p_0\left(\alpha,x^T\right)^{y_{i1}}\left(1-p_0\left(\alpha,x^T\right)\right)^{1-y_{i1}}\right.\right.$$

$$\left.\left.\prod_{t=2}^{T} P\left(y_{it}\middle|\, x_i^T, y_{it-1};\theta\right) dG\left(\alpha\middle|\, x_i^T;\theta\right)\right)\right]$$

Where is this going???

What is is all good for?

Suppose that α has a discrete distribution with known points of support, a_m, and unknown probabilities ρ_m. Ignore x_i^T.

Then

$$\pi\left(\mathcal{A};p_0\left(\cdot\right),\theta\right)=\sum_{m=1}^{M}\rho_m\left(p_0\left(a_m\right)\pi\left(\mathcal{A}\middle|\,y_0=1,a_m;\theta\right)\right.$$

$$\left.+\left(1-p_0\left(a_m\right)\right)\pi\left(\mathcal{A}\middle|\,y_0=0,a_m;\theta\right)\right)$$

$$=\sum_{m=1}^{M}z_{m,1}\pi\left(\mathcal{A}\middle|\,y_0=1,a_m;\theta\right)+\sum_{m=1}^{M}z_{m,0}\pi\left(\mathcal{A}\middle|\,y_0=0,a_m;\theta\right)$$

where

$$z_{m,1}=\rho_m p_0\left(a_m\right)\quad\text{and}\quad z_{m,0}=\rho_m\left(1-p_0\left(a_m\right)\right)$$

($\{z_m\}$ gives probabilities in the joint distribution of y_0 and α)

Θ is the values of θ for which the equations

$$\sum_{m=1}^{M}\sum_{\ell=0}^{1} z_{m,\ell}\pi\left(A \middle| y_0=\ell, a_m;\theta\right) = P\left(A\right) \tag{1}$$

$$\sum_{m=1}^{M}\sum_{\ell=0}^{1} z_{m,\ell} = 1 \tag{2}$$

$$z_{m,\ell} \geq 0 \tag{3}$$

have a solution for $\{z_m\}_{m=1}^{2M}$.

$$\Theta = \arg\max_{\theta} \operatorname*{maximize}_{\{z_m\},\{v_i\}} \sum_i -v_i$$

$$P(A) - \sum_{m=1}^{M} \sum_{\ell=0}^{1} z_{m,\ell} \pi \left(A \middle| y_0 = \ell, a_m; \theta \right) = v_A$$

$$\begin{aligned} 1 - \sum_{m=1}^{M} \sum_{\ell=0}^{1} z_{m,\ell} &= v_0 \\ z_{m,\ell} &\geq 0 \\ v_i &\geq 0 \end{aligned}$$

(The optimal function value is 0).

Example:

$$y_{it} = 1\{y_{i,t-1}\gamma + t\beta + \alpha_i + \varepsilon_{it} - 0.35 \geq 0\} \qquad \text{for} \qquad t = 1, 2, \ldots$$

with ε_{it} i.i.d. standard normal.

Not known whether the parameters of interest are point identified.

$$P\left(\alpha_i = a_j\right) =$$

$$\begin{cases} \Phi\left(\frac{a_j + a_{j+1}}{2}\right) & \text{for} \quad a_j = -4.0 \\ \Phi\left(\frac{a_j + a_{j+1}}{2}\right) - \Phi\left(\frac{a_j + a_{j-1}}{2}\right) & \text{for} \quad a_j = -3.9, -3.8, .., 3.9 \\ 1 - \Phi\left(\frac{a_j + a_{j-1}}{2}\right) & \text{for} \quad a_j = 4.0 \end{cases}$$

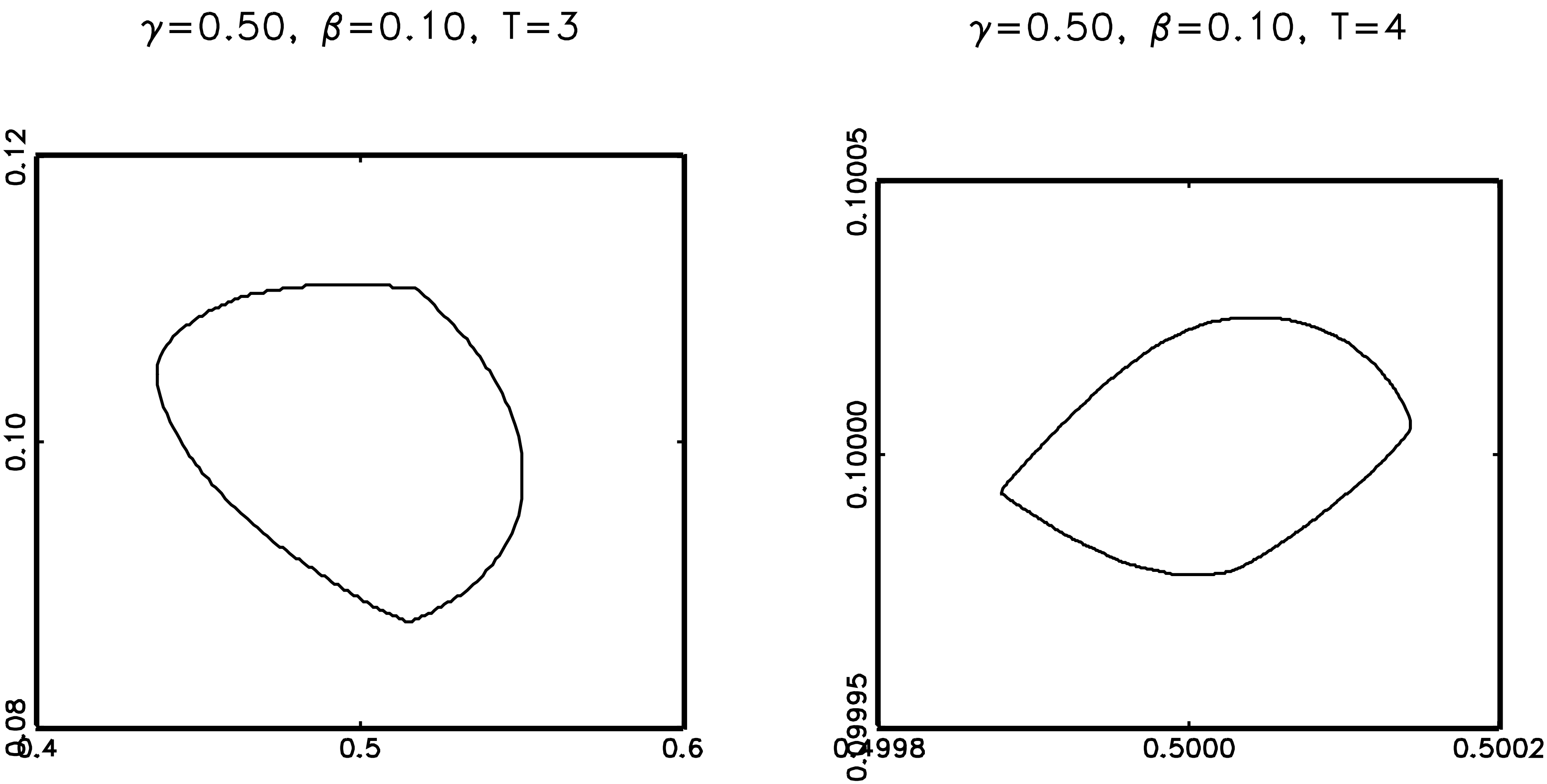


Figure 1: Identified region for (γ, β).

Marginal Effects.

Can calculate bounds on objects like

$$
\begin{aligned}
& E\left[\Phi\left(t^{\star}\beta+\gamma+\alpha\right)-\Phi\left(t^{\star}\beta+\alpha\right)\right] \\
= & \sum_{m}\left(\Phi\left(t^{\star}\beta+\gamma+a_{m}\right)-\Phi\left(t^{\star}\beta+a_{m}\right)\right)P\left(\alpha=a_{m}\right) \\
= & \sum_{m}\left(\Phi\left(t^{\star}\beta+\gamma+a_{m}\right)-\Phi\left(t^{\star}\beta+a_{m}\right)\right)\left(z_{m,1}+z_{m,0}\right)
\end{aligned}
$$

for some $t^{\star}$.

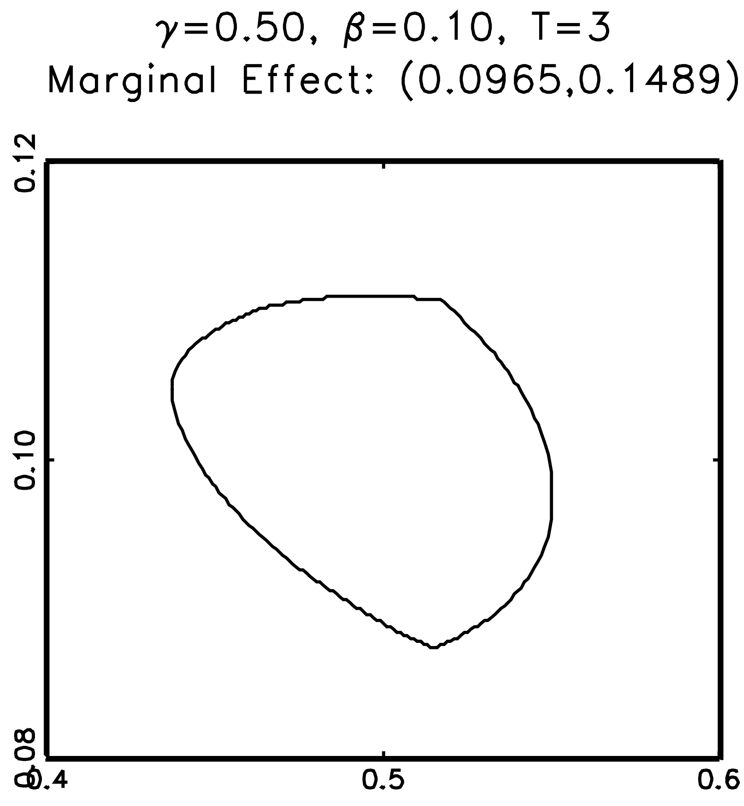
To get lower bound, minimize $MEFF(\beta,\gamma)$ over (β,γ) in the identified region, where

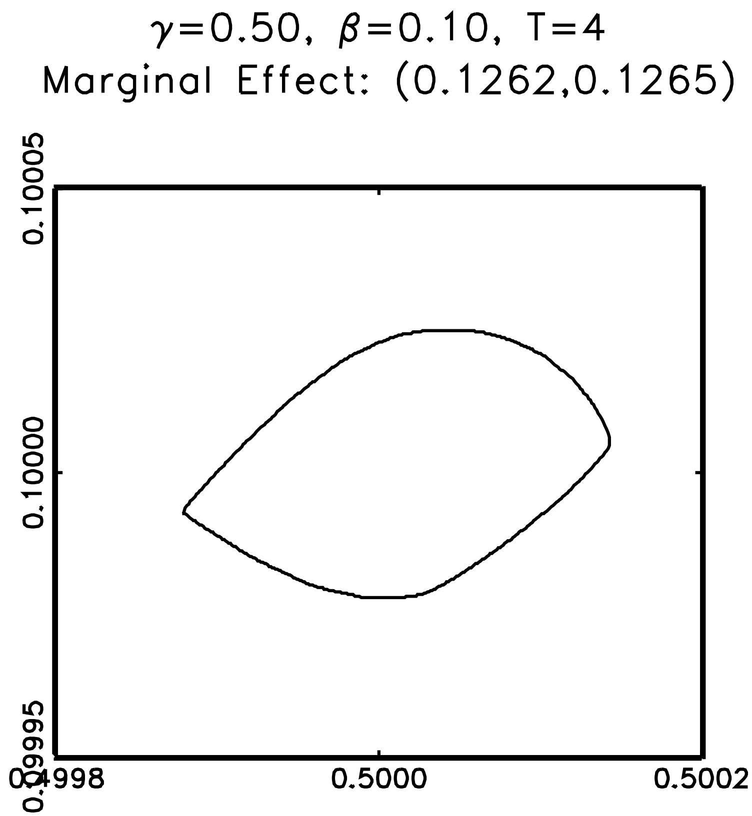
$$MEFF(\beta,\gamma) = \min_{\{z_{m,\ell}\}} \sum_m \left(\Phi\left(t^\star\beta+\gamma+a_m\right) - \Phi\left(t^\star\beta+a_m\right)\right)\left(z_{m,1}+z_{m,0}\right)$$

subject to

$$\sum_{m=1}^{M}\sum_{\ell=0}^{1} z_{m,\ell}\pi\left(A \middle| y_0=\ell, a_m;\theta\right) = P(A)$$

$$\begin{aligned}\sum_{m=1}^{M}\sum_{\ell=0}^{1} z_{m,\ell} &= 1\\ z_{m,\ell} &\geq 0\end{aligned}$$





- Not interesting in itself
- But illustrates that the approach might be interesting

Conclusions

- Seems that identification in some dynamic discrete choice models is tricky. Some “unnatural” assumptions in the literature might actually be necessary

- But non–identification might not matter much.