Ec. 2120 November 20, 2021

FINAL REVIEW PROBLEMS AND PREVIOUS FINAL EXAM

ANNOUNCEMENTS

FINAL EXAMINATION
The Final Examination will be held on Wednesday 1 at Noon. It will be a 70 minutes long. The Final is closed notes - closed internet - closed books of all kinds etc. But you may use a single 8.5 by 11 inch sheet of paper with notes on both sides.

Coverage: Lecture Notes 7 - 15.

REMINDER (again)
Grades in graduate school largely do NOT matter. Please do your best to learn the material. Most importantly, I hope this is not going to cause extra stress. If you find yourself overwhelmed, please reach out to me.

OFFICE hours will be announced soon.

1. This question is based on autoregression with random effects using longitudinal data. Let

$$Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}$$

and assume that (Y_i, A_i) is i.i.d. according to some unknown distribution F. We observe Y_i for $i = 1, \ldots, n$ but A_i is not observed. The structural regression model, which involves the latent variable A_i, is

$$E_F(Y_{it} \mid Y_{i1}, \ldots, Y_{i,t-1}, A_i) = \lambda_t^* + \rho^* Y_{i,t-1} + A_i \qquad (t = 2, \ldots, T). \tag{1}$$

We shall impose the normalization that $E_F(A_i) = 0$.

(a) Suppose that $T = 3$. Define

$$V_i = (Y_{i3} - Y_{i2}) - (\lambda_3^* - \lambda_2^*) - \rho^*(Y_{i2} - Y_{i1})$$

and show that

$$E_F(V_i) = E_F(Y_{i1} V_i) = 0. \tag{2}$$

Use the moment conditions in (2) to construct an estimator for ρ^*. Discuss the conditions needed for this estimator to be consistent as $n \to \infty$.

(b) Construct a parameter space Θ and restriction functions $\beta(\theta)$, $\Sigma(\theta)$ such that

$$E_F(Y_i) = \beta(\theta^*), \quad \mathrm{Cov}_F(Y_i) = \Sigma(\theta^*) \tag{3}$$

for some $\theta^* \in \Theta$. The parameter vector θ should include $\lambda = (\lambda_2 \quad \dots \quad \lambda_T)'$ and ρ, and will contain additional parameters that you define. The restriction functions in (3) should impose all the restrictions on $E_F(Y_i)$ and $\mathrm{Cov}_F(Y_i)$ that are implied by the model in (1), but should not impose any additional restrictions.

(c) Suppose that $T \geq 2$. Set up a quasi-likelihood function based on a multivariate normal distribution. Provide enough detail so that a research assistant, who knows linear algebra and numerical optimization in Matlab, can obtain the quasi-maximum likelihood (QML) estimate.

(d) Discuss the conditions needed for the QML estimate of ρ^* to be consistent as $n \to \infty$.

2. Suppose we have data W_i for a random sample of n individuals. We have the scores on two tests taken in the third grade: T_{i1} and T_{i2}, years of schooling S_i, and log earnings Y_i. Consider the following latent variable model:

$$\begin{aligned}
E^*(T_{i1} \,|\, 1, A_i) &= \alpha_1 + \lambda_1 A_i \\
E^*(T_{i2} \,|\, 1, A_i) &= \alpha_2 + \lambda_2 A_i \\
E^*(S_i \,|\, 1, A_i) &= \alpha_3 + \lambda_3 A_i \\
E^*(Y_i \,|\, 1, S_i, A_i) &= \alpha_4 + \gamma S_i + A_i.
\end{aligned}$$

We observe

$$W_i = (T_{i1}, T_{i2}, S_i, Y_i) \qquad (i = 1, \dots, n);$$

we do not observe A_i, but, because of the random sampling, we assume that (W_i, A_i) is i.i.d. according to some unknown distribution. Define prediction errors so that

$$\begin{aligned}
T_{i1} &= \alpha_1 + \lambda_1 A_i + V_{i1} \\
T_{i2} &= \alpha_2 + \lambda_2 A_i + V_{i2} \\
S_i &= \alpha_3 + \lambda_3 A_i + V_{i3} \\
Y_i &= \alpha_4 + \gamma S_i + A_i + V_{i4},
\end{aligned}$$

and assume that the prediction errors are mutually uncorrelated with each other. (That V_{i3} and V_{i4} are uncorrelated with each other is by construction, but the rest of the assumption is restrictive.)

(a) Consider the substitution

$$Y_i = \alpha_4 + \gamma S_i + (T_{i1} - \alpha_1 - V_{i1})/\lambda_1 + V_{i4}.$$

Show that S_i and T_{i2} are orthogonal to $V_{i4} - (V_{i1}/\lambda_1)$.

(b) Use the orthogonality conditions in (a) and the linear GMM framework to provide an estimator of γ. Give enough detail so that a research assistant, who knows how to use Matlab but does not know any probability or statistics, could program your method.

(c) Set up a quasi-likelihood function based on a multivariate normal distribution for W_i. Provide enough detail so that the RA can maximize the function. (The RA will use nonlinear optimization routines in Matlab and just needs a formula for the function to be maximized.)

(d) How can the limit distribution results in the generalized method of moments (GMM) framework be used with the quasi-likelihood function in (c) to obtain a .95 confidence interval for γ?

(e) How can the Bayesian bootstrap be used with the quasi-likelihood function in (c) to obtain a .95 credible interval for γ?

3. Suppose that you have data from a random sample $\{Y_i, X_i\}_{i=1}^n$ from some unknown distribution F, where Y_i is scalar and X_i is $K \times 1$. Assume that

$$E_F(Y_i \mid X_i) = X_i'\beta^*, \quad \text{Var}(Y_i \mid X_i) = \exp(X_i'\gamma^*).$$

(a) Set up a quasi-likelihood function based on a normal distribution for Y_i conditional on X_i. Provide enough detail so that the RA can maximize the function.

(b) Does maximizing the quasi-likelihood function provide consistent estimates of β^* and γ^* even if the distribution is not normal? Discuss.

4. Consider panel data with a binary dependent variable:

$$Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}, \quad Z_i = \begin{pmatrix} Z_{i1} \\ \vdots \\ Z_{iT} \end{pmatrix},$$

with $Y_{it} = 0$ or 1, and Z_{it} is $J \times 1$. We assume random sampling for the cross-section units: (Y_i, Z_i) i.i.d. for $i = 1, \ldots, N$. For each t, the conditional distribution of Y_{it} is given by the following probit model:

$$\text{Prob}(Y_{it} = 1 \mid Z_{it}) = \Phi(X_{it}'\gamma) \qquad (t = 1, \ldots, T), \tag{4}$$

where X_{it} (which is $K \times 1$) is a given, known function of Z_{it}: $X_{it} = g_t(Z_{it})$, and Φ is the cumulative distribution function for the standard normal distribution.

(a) Consider estimating γ with a maximum-likelihood program for a cross-section probit model, pooling the $n = NT$ observations as if they came from a single cross section:

$$\hat{\gamma} = \arg\max_a \sum_{i=1}^{N} \sum_{t=1}^{T} \Big(Y_{it} \log[\Phi(X_{it}'a)] + (1 - Y_{it}) \log[1 - \Phi(X_{it}'a)] \Big).$$

This estimator is not based on a correct likelihood function, since we are only assuming (4) and random sampling over i. Explain why, nevertheless, $\hat{\gamma}$ is a consistent estimate of γ as $N \to \infty$ with T fixed.

(b) Explain how to obtain the asymptotic covariance matrix for $\hat{\gamma}$ in (a). (You do not need to simplify the derivatives.)

(c) Define

$$\delta = \text{Prob}(Y_{i1} = 1 \,|\, Z_{i1} = d) - \text{Prob}(Y_{i1} = 1 \,|\, Z_{i1} = c).$$

Explain how to use the Bayesian bootstrap to obtain a .95 interval for δ.

5. This question is based on Note 18, Section 5, which discusses the monotonicity assumption in Imbens and Angrist (1994). Suppose that the treatment can take on three values:

$$T_i \in \mathcal{T} = \{t_0, t_1, t_2\}.$$

Define the indicator variables:

$$T_{i0} = 1(T_i = t_0), \quad T_{i1} = 1(T_i = t_1), \quad T_{i2} = 1(T_i = t_2).$$

The instrumental variable S_i can take on values in the set

$$\mathcal{S} = \{s_1, \ldots, s_L\}.$$

The potential treatment function $T_i(\,\cdot\,)$ maps each $s \in \mathcal{S}$ into a random variable, giving

$$T_i(s_1), \ldots, T_i(s_L),$$

and the observed treatment T_i is the potential treatment function evaluated at the observed S_i:

$$T_i = T_i(S_i).$$

Define the indicator functions:

$$T_{i0}(s) = 1(T_i(s) = t_0), \quad T_{i1}(s) = 1(T_i(s) = t_1), \quad T_{i2}(s) = 1(T_i(s) = t_2) \qquad (s \in \mathcal{S}).$$

Then the treatment indicator variables are obtained by evaluating the indicator functions at the observed S_i:

$$T_{ij} = T_{ij}(S_i) \qquad (j = 0, 1, 2).$$

Note that

$$T_{i0}(s) + T_{i1}(s) + T_{i2}(s) = 1 \qquad (s \in \mathcal{S}).$$

The potential outcome function $Y_i(\,\cdot\,)$ maps each $t \in \mathcal{T}$ into a random variable, giving

$$Y_i(t_0), Y_i(t_1), Y_i(t_2),$$

and the observed outcome Y_i is the potential outcome function evaluated at the observed treatment T_i:

$$Y_i = Y_i(T_i).$$

(We start with a potential outcome function $Y_i(t, s)$ and then impose the exclusion restriction that this depends only upon t, so we can replace $Y_i(t, s)$ by $Y_i(t)$.) The IV is randomly assigned:

$$\{\{Y_i(t),\ t \in \mathcal{T}\}, \{T_i(s) : s \in \mathcal{S}\}\} \perp\!\!\!\perp S_i.$$

(a) Show that

$$Y_i = \sum_{j=0}^{2} T_{ij}(S_i) Y_i(t_j).$$

(b) Show that

$$E(Y_i \,|\, S_i = s) = \sum_{j=0}^{2} E[T_{ij}(s) Y_i(t_j)].$$

(c) For any two values a and b in $\mathcal{S}$, show that

$$\begin{aligned} E(Y_i \,|\, S_i = b) - E(Y_i \,|\, S_i = a) &= \sum_{j=0}^{2} E[(T_{ij}(b) - T_{ij}(a)) Y_i(t_j)] \\ &= \sum_{j=1}^{2} E[(T_{ij}(b) - T_{ij}(a))(Y_i(t_j) - Y_i(t_0))]. \end{aligned}$$

(d) Suppose that $\mathcal{S} = \{a, b\}$ and

$$\begin{aligned} &\mathrm{Prob}(T_{i1}(b) - T_{i1}(a) \geq 0) = 1 \\ &\mathrm{Prob}(T_{i2}(b) - T_{i2}(a) \geq 0) = 1. \end{aligned}$$

Show that

$$\begin{aligned} &E(Y_i \,|\, S_i = b) - E(Y_i \,|\, S_i = a) \\ &\quad = \sum_{j=1}^{2} E[Y_i(t_j) - Y_i(t_0) \,|\, T_{ij}(b) - T_{ij}(a) = 1] \Pr(T_{ij}(b) - T_{ij}(a) = 1). \end{aligned}$$

What, if anything, is identified here that is related to treatment effects? In particular, can we identify an average treatment effect for a subpopulation (a conditional average treatment effect), or perhaps some average of such effects?

(e) In (d), the application could be to a population at risk for high blood pressure, with $S_i = b$ corresponding to some encouragement to be aware of the risk, perhaps through a blood pressure test, and $S_i = a$ provides a control group. The treatments could be take drug 1 (t_1), take drug 2 (t_2), or no drug (t_0). The monotonicity assumption in (d) would be plausible if initially people were not taking a drug, and after receiving the encouragement ($S_i = b$) some took drug 1, some took drug 2, and the rest stayed with no drug. But suppose that initially some were taking one of the drugs, and after receiving the encouragement ($S_i = b$) some of the people taking drug 1 switched to drug 2, and some taking drug 2 switched to drug 1. What, if anything, is identified now that is related to treatment effects?