ECONOMETRICS II (ECO 2401S)

University of Toronto. Department of Economics. Spring 2009
Professor Victor Aguirregabiria

TEST (April 8, 2009)

QUESTION 1 (40 points): Consider the life-cycle consumption model in Hansen and Singleton (Ectca, 1982), and the corresponding Euler equation:

$$\beta \left(\frac{C_{t+1}}{C_t} \right)^{\alpha} R_{t+1} - 1 = \varepsilon_{t+1}$$

where ε_{t+1} is an expectational error such that under the rational expectations assumption $E_t(\varepsilon_{t+1}) = 0$. Given a time series dataset $\{C_t, R_t : t = 1, 2, ..., T\}$ we are interested in the estimation of the vector of parameters $\theta_0 \equiv (\alpha, \beta)$.

(a) [10 points] Describe a GMM estimator of θ_0 based on the unconditional moment restrictions $E(\varepsilon_{t+1}) = 0$, $E(R_t\varepsilon_{t+1}) = 0$, and $E(C_t\varepsilon_{t+1}) = 0$.

(b) [15 points] Given the conditional moment restrictions $E(\varepsilon_{t+1}|C_t, R_t) = 0$, derive the moment conditions of the optimal GMM estimator of θ_0.

(c) [15 points] Given the optimal moment conditions obtained in (b), describe a procedure to obtain a feasible GMM estimator that exploits those optimal moment conditions.

(a) [10 points] Define the vector of sample moment conditions:

$$m_T(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} Z_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{\alpha} R_{t+1} - 1 \right]$$

where Z_t is the column vector $(1, R_t, C_t)'$. Then, a GMM estimator of θ_0 is:

$$\hat{\theta} = \arg\min_{\theta} \; m_T(\theta)' A_T m_T(\theta)$$

where A_T is a weighting matrix. Given these unconditional moment restrictions, the optimal GMM estimator is the one with weighting matrix $A_T^* = E(\varepsilon_{t+1}^2 Z_t Z_t')^{-1}$.

(b) [15 points] Following Newey (1990), the optimal set of moment conditions is $E\left(Z_t^* \varepsilon_{t+1}\right) = 0$, where:

$$Z_t^* = \frac{1}{E\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)} \begin{bmatrix} E\left(\dfrac{\partial \varepsilon_{t+1}}{\partial \alpha} \mid C_t, R_t \right) \\ E\left(\dfrac{\partial \varepsilon_{t+1}}{\partial \beta} \mid C_t, R_t \right) \end{bmatrix}$$

Taking into account that $\varepsilon_{t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{\alpha} R_{t+1} - 1$, we have that:

$$\begin{aligned}
\frac{\partial \varepsilon_{t+1}}{\partial \alpha} &= \log\left(\frac{C_{t+1}}{C_t}\right) \beta \left(\frac{C_{t+1}}{C_t}\right)^{\alpha} R_{t+1} &&= \log\left(\frac{C_{t+1}}{C_t}\right) \varepsilon_{t+1} \\
\frac{\partial \varepsilon_{t+1}}{\partial \beta} &= \left(\frac{C_{t+1}}{C_t}\right)^{\alpha} R_{t+1} &&= \frac{\varepsilon_{t+1}}{\beta} + 1
\end{aligned}$$

Therefore,

$$Z_t^* = \frac{1}{E\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)} \begin{bmatrix} E\left(\log\left(\frac{C_{t+1}}{C_t}\right) \varepsilon_{t+1} \mid C_t, R_t\right) \\ 1 \end{bmatrix}$$

because $E(\varepsilon_{t+1} | C_t, R_t) = 0$.

(c) [15 points] Let $\hat{\theta}$ be the GMM estimator described in point (a). And let $\{\hat{\varepsilon}_{t+1}\}$ be the residuals based on this estimator, i.e., $\hat{\varepsilon}_{t+1} = \hat{\beta} \left(\frac{C_{t+1}}{C_t}\right)^{\hat{\alpha}} R_{t+1} - 1$. Using these residuals, we obtain nonparametric estimates of the expectations $E\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)$ and $E\left(\log\left(\frac{C_{t+1}}{C_t}\right) \varepsilon_{t+1} \mid C_t, R_t\right)$. For instance, we can use a kernel (Nadaraya-Watson) estimator. Or alternatively, we can run a regression of $\hat{\varepsilon}_{t+1}^2$ on the terms of a high order polynomial in (C_t, R_t) and use the fitted values of that regression as nonparametric estimates of $E\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)$. Similarly, we can run a regression of $\log\left(\frac{C_{t+1}}{C_t}\right) \hat{\varepsilon}_{t+1}$ on the terms of a high order polynomial in (C_t, R_t) and use the fitted values of that regression as nonparametric estimates of $E\left(\log\left(\frac{C_{t+1}}{C_t}\right) \varepsilon_{t+1} \mid C_t, R_t\right)$. We represent these estimates as $\hat{E}\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)$ and $\hat{E}\left(\log\left(\frac{C_{t+1}}{C_t}\right) \varepsilon_{t+1} \mid C_t, R_t\right)$. Then, our estimate of the vector of optimal instruments is:

$$\hat{Z}_t^* = \begin{bmatrix} \dfrac{\hat{E}\left(\log\left(\frac{C_{t+1}}{C_t}\right) \varepsilon_{t+1} \mid C_t, R_t\right)}{\hat{E}\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)} \\ \dfrac{1}{\hat{E}\left(\varepsilon_{t+1}^2 \mid C_t, R_t\right)} \end{bmatrix}$$

Finally, define the vector of sample moment conditions:

$$m_T^*(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} Z_t^* \left[\beta \left(\frac{C_{t+1}}{C_t}\right)^{\alpha} R_{t+1} - 1\right]$$

The optimal GMM estimator of θ_0 is the value of θ that solves the system $m_T^*(\theta) = 0$.

QUESTION 2 (25 points): Consider the static linear panel data model:

$$y_{it} = x_{it}'\,\beta + \eta_i + u_{it}$$

where η_i and u_{it} are independent unobservables with zero mean, and x_{it} is a vector of strictly exogenous regressors such that $E[x_{it}\ u_{is}] = 0$ for any t and s. We have a panel dataset $\{y_{it}, x_{it} : i = 1, 2, ..., N; t = 1, 2, ..., T\}$, where N is large and $T = 2$. We are interested in the estimation of the vector of parameters β, and also in the variance of the individual effects η_i, and the variance of the transitory shocks u_{it}.

(a) [10 points] Describe the fixed effects or within-groups estimator of β. Explain its asymptotic properties.

(b) [5 points] What is the fixed-effect estimator of η_i. Describe its statistical properties.

(c) [5 points] Under the assumption that u_{it} is iid, explain how to obtain a consistent estimator of $Var(\eta_i)$ and of $Var(u_{it})$.

(d) [5 points] Suppose that u_{it} is not serially correlated but it is time-heteroscedastic: i.e., $Var(u_{i1}) \neq Var(u_{i2})$. Explain how to obtain consistent estimators of $Var(\eta_i)$, $Var(u_{i1})$, and $Var(u_{i2})$.

(a) [10 points] The WG estimator of β is the OLS estimator in the following WG-transformed equation:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\,\beta + (u_{it} - \bar{u}_i)$$

In a static model (i.e., $E[x_{it}\ u_{is}] = 0$ for any t and s), this estimator is consistent because $E((x_{it} - \bar{x}_i)\,(u_{it} - \bar{u}_i)) = 0$. If $E(x_{it}\eta_i) \neq 0$, and if u_{it} is iid over individuals and over time, then the WG estimator is also efficient. The reason is that the WG estimator is equivalent to the OLS estimator where we include dummy variables to control for the individual effects, and in that model the error term u_{it} is iid and therefore OLS is efficient. If $E(x_{it}\eta_i) = 0$, the WG estimator is not efficient. A more efficient estimator can be obtained if η_i is considered as part of the error term (i.e., individual dummy variables are NOT included), and a GLS estimator is obtained. That is the Balestra-Nerlove estimator.

(b) [5 points] Let $\hat{\beta}$ be the WG estimator, and let $\hat{\varepsilon}_{it}$ be the WG residuals: $\hat{\varepsilon}_{it} = y_{it} - x_{it}'\hat{\beta}$. The WG estimator of the individual effect η_i is the sample mean of $\{\hat{\varepsilon}_{it}\}$ for individual i:

$$\hat{\eta}_i = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{it}$$

In fact, this is also the estimator of η_i that we obtain if we use a "brute force" dummy-variables-estimator of the model. This estimator is not consistent for fixed T. As N goes to infinity (and T is fixed), the estimator $\hat{\eta}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - x_{it}'\hat{\beta})$ converges in probability to $\frac{1}{T}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta) = \eta_i + \bar{u}_i \neq \eta_i$.

(c) [5 points] Under the assumption that u_{it} is iid, we have that:

$$\begin{aligned} E\left(\varepsilon_{it}^{2}\right) &= Var(\eta_i) + Var(u_{it}) \\ E\left(\varepsilon_{it}\varepsilon_{it-1}\right) &= Var(\eta_i) \end{aligned}$$

Therefore, using the WG residuals $\{\hat{\varepsilon}_{it}\}$ we can obtain the following consistent estimators of $Var(\eta_i)$ and $Var(u_{it})$:

$$\begin{aligned} \widehat{Var}(\eta_i) &= \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{\varepsilon}_{it}\hat{\varepsilon}_{it-1} \\ \widehat{Var}(u_{it}) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{\varepsilon}_{it}^{2} - \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{\varepsilon}_{it}\hat{\varepsilon}_{it-1} \end{aligned}$$

With $T = 2$, these expressions become:

$$\begin{aligned} \widehat{Var}(\eta_i) &= \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_{i2}\ \hat{\varepsilon}_{i1} \\ \widehat{Var}(u_{it}) &= \frac{1}{2N} \sum_{i=1}^{N} \left[\hat{\varepsilon}_{i1}^{2} + \hat{\varepsilon}_{i2}^{2}\right] - \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1} = \frac{1}{2N} \sum_{i=1}^{N} \left[\hat{\varepsilon}_{i1} - \hat{\varepsilon}_{i2}\right]^{2} \end{aligned}$$

(d) [5 points] If $T = 2$ and u_{it} is not serially correlated but it is time-heteroscedastic, we have that:

$$\begin{aligned} E\left(\varepsilon_{i1}^{2}\right) &= Var(\eta_i) + Var(u_{i1}) \\ E\left(\varepsilon_{i2}^{2}\right) &= Var(\eta_i) + Var(u_{i2}) \\ E\left(\varepsilon_{i2}\varepsilon_{i1}\right) &= Var(\eta_i) \end{aligned}$$

Therefore, using the WG residuals $\{\hat{\varepsilon}_{it}\}$ we can obtain the following consistent estimators of $Var(\eta_i)$, $Var(u_{i1})$**,** and $Var(u_{i2})$**:**

$$\begin{aligned} \widehat{Var}(\eta_i) &= \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_{i2}\ \hat{\varepsilon}_{i1} \\ \widehat{Var}(u_{i1}) &= \frac{1}{N} \sum_{i=1}^{N} \left[\hat{\varepsilon}_{i1}^{2} - \hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1}\right] \\ \widehat{Var}(u_{i2}) &= \frac{1}{N} \sum_{i=1}^{N} \left[\hat{\varepsilon}_{i2}^{2} - \hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1}\right] \end{aligned}$$

QUESTION 3 (35 points): Consider the following binary choice model for an individual's unit demand of a good. A model of utility maximization implies that:

$$Y = 1 \;\; \Leftrightarrow \;\; \{(M - P) + \alpha_0 + \alpha_1 Z - \varepsilon + \alpha_2 (M - P)\} > M$$

$Y \in \{0, 1\}$ is the binary variable that indicates whether the individual purchases the good ($Y = 1$) or not ($Y = 0$). M is the individual's disposable income. P is the price of the good in the market where the individual lives. Z is a vector of individual characteristics observable to the econometrician, such as age and education. α_0, α_1 and α_2 are preference parameters. And ε is a zero mean random variable that is individual-specific. Re-arranging terms, we have the following BCM:

$$Y = 1 \;\; \Leftrightarrow \;\; \{\alpha_0 + \alpha_1 Z + \alpha_2 (M - P) - P - \varepsilon\} > 0$$

Given a sample $\{Y_i, Z_i, M_i, P_i : i = 1, 2, ..., n\}$, we are interested in the estimation of the vector of parameters $\theta \equiv (\alpha_0, \alpha_1, \alpha_2, \sigma_\varepsilon)$, where σ_ε^2 is the variance of ε.

(a) [10 points] Suppose that ε_i is independent of (Z_i, M_i, P_i) and it is normally distributed with mean zero and variance σ_ε^2. Describe the MLE of θ. Show that all the parameters in θ are identified.

(b) [10 points] We are concerned with the endogeneity of the price P_i. For instance, in markets populated by consumers with relatively high values of ε we expect to find higher prices, and this implies a positive correlation between P_i and ε_i. Describe in detail a method to deal with the endogeneity of prices in this BCM. Try to justify the identifying exclusion restriction.

(c) [15 points] Define the Maximum Score Estimator and the Smooth Maximum Score Estimator of $\theta \equiv (\alpha_0, \alpha_1, \alpha_2, \sigma_\varepsilon)$.

(a) [10 points] The log-likelihood function of this model is:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} Y_i \;\log \Phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \alpha_2 (M_i - P_i) - P_i}{\sigma_\varepsilon}\right) \\ &\quad + (1 - Y_i)\;\log \Phi\left(\frac{-\alpha_0 - \alpha_1 Z_i - \alpha_2 (M_i - P_i) + P_i}{\sigma_\varepsilon}\right) \end{aligned}$$

where Φ is the CDF of the standard normal. The MLE is the value of θ that maximizes $l(\theta)$.

All the parameters in $(\alpha_0, \alpha_1, \alpha_2, \sigma_\varepsilon)$ are identified. To see this, note that we can write $\frac{\alpha_0 + \alpha_1 Z_i + \alpha_2 (M_i - P_i) - P_i}{\sigma_\varepsilon}$ as $\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i$, where $\tilde{\alpha}_0 = \alpha_0 / \sigma_\varepsilon$, $\tilde{\alpha}_1 = \alpha_1 / \sigma_\varepsilon$, $\tilde{\alpha}_2 = \alpha_2 / \sigma_\varepsilon$, and $\tilde{\alpha}_3 = -1/\sigma_\varepsilon$. In a Probit model, the parameters $(\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3)$ are identified. And given $(\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3)$, we can uniquely obtain θ as:

$$\begin{aligned} \alpha_0 &= -\tilde{\alpha}_0 / \tilde{\alpha}_3 \\ \alpha_1 &= -\tilde{\alpha}_1 / \tilde{\alpha}_3 \\ \alpha_2 &= -\tilde{\alpha}_2 / \tilde{\alpha}_3 \\ \sigma_\varepsilon &= -1 / \tilde{\alpha}_3 \end{aligned}$$

(b) [10 points] We can apply Rivers-Vuong method. To apply this method we have to make several assumptions. Consider the following model:

$$(1) \quad Y_i = I\{\alpha_0 + \alpha_1 Z_i + \alpha_2 (M_i - P_i) - P_i - \varepsilon_i > 0\}$$

$$(2) \quad P_i = W_i' \delta + u_i$$

where equation (1) is our BCM of demand, and equation (2) is a reduced form equation for prices. The first assumption is that W_i contains exogenous observable variables which are independent of u_i and ε_i. The second assumption is that W_i contains at least one variable (with parameter $\delta \neq 0$) that does not enter as explanatory variable in equation (1). This assumption is a "exclusion restriction". In the context of estimating this demand model, possible exclusion restrictions come from variables that affect production costs but not demand. For instance, if the good is produced locally at different local markets, production costs could vary across these local markets. The third assumption is that the unobservables u_i and ε_i are jointly normal. Under joint normality, we have that:

$$\varepsilon_i = \pi \; u_i + \xi_i$$

where π is the parameter $\sigma_{\varepsilon u}/\sigma_u^2$, and ξ is a variable that is normally distributed, independent of u, and $\sigma_\xi^2 = \sigma_\varepsilon^2 (1 - \rho^2)$, where where ρ is the correlation between ε and u. Solving this expression in equation (1), we get:

$$Y_i = I\{\alpha_0 + \alpha_1 Z_i + \alpha_2 (M_i - P_i) - P_i - \pi \; u_i - \xi_i > 0\}$$

And it is convenient to represent this BCM as:

$$Y_i = I\{\alpha_0^* + \alpha_1^* Z_i + \alpha_2^* (M_i - P_i) + \alpha_3^* P_i + \alpha_4^* \; u_i - \xi_i^* > 0\}$$

where ξ_i^* is the standard normal random variable ξ_i/σ_ξ, and

$$\begin{aligned} \alpha_0^* &= \alpha_0/\sigma_\xi \\ \alpha_1^* &= \alpha_1/\sigma_\xi \\ \alpha_2^* &= \alpha_2/\sigma_\xi \\ \alpha_3^* &= -1/\sigma_\xi \\ \alpha_4^* &= -\pi/\sigma_\xi \end{aligned}$$

Given this model, we can estimate consistently θ using a two-step procedure. In the first step, we estimate the reduced form equation (2) by OLS and obtain the OLS residuals $\{\hat{u}_i\}$. In the second step, we estimate the Probit model:

$$Y_i = I\{\alpha_0^* + \alpha_1^* Z_i + \alpha_2^* (M_i - P_i) + \alpha_3^* P_i + \alpha_4^* \; \hat{u}_i - \xi_i^* > 0\}$$

where we include the residual variable $\hat{u}_i$ as an explanatory variable. Given the estimate of $\{\alpha_0^*, \alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*\}$, we can uniquely obtain an estimate of θ as:

$$\begin{aligned} \alpha_0 &= -\alpha_0^*/\alpha_3^* \\ \alpha_1 &= -\alpha_1^*/\alpha_3^* \\ \alpha_2 &= -\alpha_2^*/\alpha_3^* \\ \sigma_\xi &= -1/\alpha_3^* \\ \pi &= \alpha_4^*/\alpha_3^* \end{aligned}$$

(c) [15 points] Note: For notational convenience, I present here estimators of the vector of parameters $\tilde{\alpha} \equiv (\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3)$, where these $\tilde{\alpha}$ parameters are the ones defined in (a). As explained in (a), there is a one-to-one relationship between θ and $\tilde{\alpha}$, and it is straightforward to obtain θ given $\tilde{\alpha}$. However, we can describe the following estimators directly for the vector θ.

The Maximum Score estimator (MSE) of $\tilde{\alpha}$ is defined as the value that maximizes in $\tilde{\alpha}$ the score function:

$$\begin{aligned} S(\tilde{\alpha}) &= \sum\nolimits_{i=1}^{n} Y_i \; I\{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i > 0\} \\ &+ \; (1 - Y_i) \; I\{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i < 0\} \end{aligned}$$

The score function $S(\tilde{\alpha})$ provides the number of correct predictions if we predict Y_i using the predictor $I\{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i > 0\}$. Maximizing this score function is equivalent to maximize the following function

$$\sum\nolimits_{i=1}^{n} (2Y_i - 1) \; I\{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i > 0\}$$

And it is also equivalent to maximize the LAD function:

$$\sum\nolimits_{i=1}^{n} |Y_i - I\{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i > 0\}|$$

Therefore, the MSE can be defined as the maximizer of any of these three criterion functions.

The smooth-MSE is defined as the value of $\tilde{\alpha}$ that maximizes the following smooth score function:

$$SS(\tilde{\alpha}) \quad \sum\nolimits_{i=1}^{n} (2Y_i - 1) \; F\left(\frac{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i}{b_n}\right)$$

F is the CDF of random variable with support the real line and with symmetric distribution around zero, and it is a continuously differentiable and strictly increasing function (e.g., the CDF of the standard normal). b_n is a bandwidth parameter that goes to zero as the sample size n goes to infinity. Therefore, as n goes to infinity, $F\left(\frac{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i}{b_n}\right)$ converges to the indicator function $I\{\tilde{\alpha}_0 + \tilde{\alpha}_1 Z_i + \tilde{\alpha}_2 (M_i - P_i) + \tilde{\alpha}_3 P_i > 0\}$, and the smooth score function converges to the standard score function.