Economics Letters 73 (2001) 227–231

economics letters

www.elsevier.com/locate/econbase

Consistent estimation of the random structural coefficient distribution from the linear simultaneous equations system

Jinyong Hahn*

Brown University, Department of Economics, Box B, Providence, RI 02912, USA

Received 28 January 2001; accepted 24 May 2001

Abstract

For the linear simultaneous equations model (LSEM) with random structural coefficients, it is sometimes of interest to identify the distribution of those structural coefficients. It is shown that the structural parameter distribution can be identified if the distribution of the exogenous variables is reasonably smooth.

Keywords: Linear simultaneous equations; Identification; Structural parameter

JEL classification: C3

1. Introduction

Recently, Foster and Hahn (2000) considered estimation of the Marshallian demand function linear in random coefficients $f_i(p, y) = \alpha_i + \beta_i \cdot p + \gamma_i \cdot y$, where $f_i(p, y)$ is the ith individual's (log) demand for certain good given the (log) price p and the (log) income y, and $\theta_i = (\alpha_i, \beta_i, \gamma_i)$ is a random vector. Under the assumption that θ_i is independent of (p_i, y_i), it was shown that the *distribution* of θ_i can be consistently estimated. Though the assumption that θ_i is independent of p_i may be reasonable under the assumption that consumers are price takers, the assumption that θ_i is independent of y_i may not be plausible. This is especially true when the expenditure is used as a proxy for income, in which case Summers (1959) recommends using the simultaneous equations model (SEM) to avoid the inconsistency of the OLS. His analysis would carry over *if* (β_i, γ_i) is constant across individuals. If (β_i, γ_i) is heterogeneous, standard IV estimator may not necessarily converge to the parameter of interest, say $E[(\beta_i, \gamma_i)]$. It is thus of interest to see if the distribution of θ_i can be identified under the SEM.

Angrist et al. (2000) argued that the probability limit of the IV estimator can be interpreted as some

*Tel.: +1-401-863-3883; fax: +1-401-863-1970.

E-mail address: Jinyong-ahn@brown.edu (J. Hahn).

0165-1765/01/$ – see front matter
PII: S0165-1765(01)00492-X

weighted average of the parameter. The interpretation of the IV estimator as a weighted average of the random coefficient motivates another interesting question. Suppose that there are two groups of individuals, where the distributions of structural parameter may differ from each other. Suppose, however, the distributions of some *component* of the structural parameter are the same in both groups. It may well happen that the IV estimator of this component for the first group will be different from that for the second group even asymptotically solely because those IV estimators estimate weighted averages with different weights. It may be of interest to identify the distribution of this component of the structural parameter. As an example, consider Pitt and Khandker's (1995) finding that the credit program for the poor on the well-being of the household is greater when women are the program participants. If the treatment effect is heterogeneous across individuals, then the standard IV estimates for men and women might be different solely because the program participation is an endogenous variable governed by parameter whose distribution is different for different sexes. It is thus of interest to see if the program effect parameter has the same distribution for both men and women.

2. Identification

Consider the reduced form

$$y_i = G(\theta_i) \cdot x_i \tag{1}$$

of some simultaneous equations system. Here, θ_i denotes the structural parameters, and x_i is the exogenous variable. The mapping $\theta \mapsto G(\theta_i)$ completely characterizes the relation between the structural parameter and the reduced form parameter.

The structural parameter θ_i does not have to be literally structural. For example, we might be interested in one equation $y_{1i} = \gamma'_{1i} Y_{1i} + \gamma'_{2i} X_{1i}$ in SEM, where y_{1i} is an endogenous variable, Y_{1i} and X_{1i} are vectors consisting of endogenous variables and exogenous variables which appear on the right hand side of this structural equation. Consider the reduced form equation $Y_{1i} = \Gamma_{1i} X_{1i} + \Gamma_{2i} X_{2i}$, where X_{2i} is the vector consisting of exogenous variables not appearing in the above structural equation. Defining $y_i \equiv (y_{1i}, Y'_{1i})'$, $x_i \equiv (X'_{1i}, X'_{2i})'$, and letting θ_i consisting of the structural parameters γ_{1i}, γ_{2i}, and the reduced form parameters Γ_{1i}, Γ_{2i}, we find that Eq. (1) holds with

$$G(\theta_i) = \begin{bmatrix} \gamma'_{1i}\Gamma_{1i} & \gamma'_{1i}\Gamma_{2i} + \gamma'_{2i} \\ \Gamma_{1i} & \Gamma_{2i} \end{bmatrix}.$$

Assumption. (i) (θ_i, x_i) $i = 1, \ldots n$ is *i.i.d.*; (ii) θ_i is independent of x_i; (iii) The map $\theta \mapsto g(\theta) \equiv \mathrm{vec}(G(\theta))$ is a homeomorphism between $\Theta = \text{support}\ (\theta_i)$ and $g(\Theta) \equiv \{g(\theta) : \theta \in \Theta\}$; (iv) Θ is compact; (v) $\{t'x : x \in \text{support}(F_{x,0})\}$ contains an open set for every $t \neq 0$.

Remark 1. From Assumption (iii), it necessarily follows that $\dim(\theta_i) \geq \dim(g)$.

Let F_θ denote a candidate distribution of θ_i, which will induce a distribution of $g(\theta_i)$, which will be denoted by $g(F_\theta)$ with some abuse of notation. Let $F_{(y,x),0}$ and $F_{x,0}$ denote the true distribution of (y_i, x_i) and x_i. Also, let $\mathscr{L}(g(F_\theta), F_{x,0})$ denote the distribution of (y_i, x_i) under Eq. (1) and F_θ. If $F_{\theta,0}$ is the true distribution of θ_i, then $\mathscr{L}(g(F_{\theta,0}), F_{x,0})$ will coincide with $F_{(y,x),0}$. Therefore, the true distribution $F_{\theta,0}$ of θ_i solves

$$\min_{F_\theta} d[\mathscr{L}(g(F_\theta), F_{x,0}), F_{(y,x),0}]$$

for any metric d. Observe that $F_{x,0}$ and $F_{(y,x),0}$ are consistently estimated by corresponding empirical distributions $\hat{F}_x$ and $\hat{F}_{(y,x)}$, which motivates an estimator $\hat{F}_\theta$ that solves

$$\min_{F_\theta \in \mathbb{C}} d[\mathscr{L}(g(F_\theta), \hat{F}_x), \hat{F}_{(y,x)}], \tag{2}$$

where $\mathbb{C}$ denotes the collection of probabilities supported by any compact set which contains Θ. Here, $d(\cdot, \cdot)$ denote any metric which metrizes weak convergence, e.g. an L_2-metric on characteristic functions. Such estimator will be called the non-parametric minimum distance estimator.

Consistency of the minimum distance estimator is established using the following lemma:

Lemma 1. *Suppose that Assumptions* (i)–(v) *hold. Let* $\{(F_{\theta,n}, F_{x,n})\}$ *denote a sequence of candidate distributions of* θ_i *and* x_i. (i) *If* $d(F_{\theta,n}, F_{\theta,0}) = o_p(1)$, *and* $d(F_{x,n}, F_{x,0}) = o_p(1)$, *then* $d[\mathscr{L}(g(F_{\theta,n}), F_{x,n}), \mathscr{L}(g(F_{\theta,0}), F_{x,0})] = o_p(1)$. (ii) *If* $d[\mathscr{L}(g(F_{\theta,n}), F_{x,n}), \mathscr{L}(g(F_{\theta,0}), F_{x,0})] = o_p(1)$, *then* $d(F_{\theta,n}, F_{\theta,0}) = o_p(1)$.

Proof. (i) By the continuous mapping theorem, we have $d(g(F_{\theta,n}), g(F_{\theta,0})) \to 0$. The conclusion then easily follows by Beran and Millar (1994), Proposition 2.1. (ii) By Beran and Millar (1994), Proposition 2.2, it suffices to show that, if $d(g(F_{\theta,n}), g(F_{\theta,0})) \to 0$, then $d(F_{\theta,n}, F_{\theta,0}) \to 0$. Suppose not. Because Θ is compact, the sequence $\{F_{\theta,n}\}$ is tight. Therefore, we can find some subsequence $\{n'\}$ which converges to $F^* \neq F_{\theta,0}$. Because g is one-to-one, we have $g(F^*) \neq g(F_{\theta,0})$. But then, we would have $d(g(F_{\theta,n'}), g(F^*)) \to 0$, a contradiction. □

Theorem 1. *Under Assumptions* (i)–(v), $d(\hat{F}_\theta, F_{\theta,0}) = o_p(1)$.

Proof. By the law of large numbers, we have $d(\hat{F}_x, F_{x,0}) = o_p(1)$. Thus, by Lemma 1 (i), we obtain

$$d(\mathscr{L}(g(F_{\theta,0}), \hat{F}_{x,0}), \mathscr{L}(g(F_{\theta,0}), F_{x,0})) = o_p(1). \tag{3}$$

By the application of the law of large numbers again, we obtain,

$$d(\hat{F}_{(y,x)}, \mathscr{L}(g(F_{\theta,0}), F_{x,0})) = o_p(1). \tag{4}$$

Triangle inequality applied to Eqs. (3) and (4) yields

$$d(\hat{F}_{(y,x)}, \mathscr{L}(g(F_{\theta,0}), \hat{F}_{x,0})) = o_p(1). \tag{5}$$

Combining Eq. (5) with the definition (2) of $\hat{F}_\theta$, we obtain

$$d(\hat{F}_{(y,x)}, \mathscr{L}(g(\hat{F}_\theta), \hat{F}_{x,0})) = o_p(1). \tag{6}$$

Triangle inequality applied to Eqs. (4) and (6) yields

$$d\big(\mathscr{L}\big(g\big(\hat{F}_{\theta}\big), \hat{F}_{x,0}\big), \mathscr{L}\big(g\big(F_{\theta,0}\big), F_{x,0}\big)\big) = o_p(1).$$

Using Lemma 1 (ii), we obtain the desired conclusion. □

Computation of the non-parametric minimum distance estimator using Eq. (2) is computationally intractable. We thus consider the multinomial approximation. Let $\mathbb{C}(m)$ denote a subset of $\mathbb{C}$ which contains every multinomial distribution with m points with mass at each point being m^{-1}. Let m_n denote any strictly increasing sequence of positive integers such that $\lim_{n\to\infty} m_n = \infty$, and define $\tilde{F}_{\theta,n}$ as a solution to

$$\tilde{F}_{\theta,n} \equiv \underset{F_\theta \in \mathbb{C}(m_n)}{\operatorname{argmin}}\, d\big[\mathscr{L}\big(g(F_\theta), \hat{F}_x\big), \hat{F}_{(y,x)}\big].$$

Because the multinomial distribution is dense in $\mathbb{C}$ in the topology defined by the weak convergence, we have

Theorem 2. $d\big(\tilde{F}_{\theta,n}, F_{\theta,0}\big) \to 0$ *in probability*.

The estimation of the minimum distance estimator requires a choice of $d(\cdot,\cdot)$. For this purpose, we can use the L_2-norm on the characteristic functions:

$$d(P_1,P_2) \equiv \left(\int |\phi_1(t) - \phi_2(t)|^2 \, \mathrm{d}Q(t)\right)^{1/2},$$

where $\phi_1(t)$ and $\phi_2(t)$ are characteristic functions of P_1 and P_2, and Q is some probability with the support equal to the whole Euclidean space. It is often impossible to obtain an analytic expression of $d(\cdot,\cdot)$ when Q has the whole Euclidean space as its support. But we may instead use

$$d_N(P_1, P_2) = \left(\int |\phi_1(t) - \phi_2(t)|^2 \, \mathrm{d}Q_N(t)\right)^{1/2},$$

where Q_N is the empirical distribution of a random sample of size N from Q. As long as $N \to \infty$ as $n \to \infty$, the corresponding simulated minimum distance estimator is would still be consistent.

Acknowledgements

I appreciate Andrew Foster for helpful discussion.

References

Angrist, J., Grady, K., Imbens, G., 2000. The Interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. Review of Economic Studies 67, 499–527.

Beran, R., Millar, P., 1994. Minimum distance estimation in random coefficient regression models. Annals of Statistics 22, 1976–1992.

Foster, A., Hahn, J., 2000. A consistent semiparametric estimation of the consumer surplus distribution. Economics Letters 69, 245–251.

Pitt, M., Khandker, S., 1994. Household and Intrahousehold Impacts of the Grameen Bank and Similar targeted Credit Programs in Bangladesh, Brown University, unpublished manuscript.

Summers, R., 1959. A note on least squares bias in household expenditure analysis. Econometrica 27, 121–126.