CAM

Centre for Applied Microeconometrics

Department of Economics
University of Copenhagen

http://www.econ.ku.dk/CAM/

A Note on the Correlated Random Coefficient Model

Christophe Kolodziejczyk

2006-10

The activities of CAM are financed by a grant from
The Danish National Research Foundation

A Note on the Correlated Random Coefficient Model

Christophe Kolodziejczyk*

Center for Applied Microeconometrics,

Department of Economics, University of Copenhagen

Abstract

In this note we derive the bias of the OLS estimator for a correlated random coefficient model with one random coefficient, but which is correlated with a binary variable. We provide set-identification to the parameters of interest of the model. We also show how to reduce the bias of the estimator.

Keywords: Correlated random coefficient model, Discrete choice, Bias, Cross-section data

JEL Codes: C13, C21

*I thank Martin Browning and Ramses Abul Naga for their very useful suggestion. I am responsible for any error.

A Note on the Correlated Random Coefficient Model

Christophe Kolodziejczyk

1159 words

1 Introduction

Many individual economic decisions involve private information which by definition are not available to the researcher. Unobserved heterogeneity may influence decisions over the life-cycle and both the timing of these events and the size of these idiosyncratic parameters will be intimately linked. This poses some methodological problems for the estimation of models where we have unobserved heterogeneity correlated with variables that reflect past decisions. Often we will have to estimate correlated random coefficient models. For relevant examples of this phenomenon, one can refer to Card [2] and Browning and Lechene [1]. Card, for example, shows with a life-cycle model that the number of years of schooling depends not on only on the ability of the individuals but also on their specific returns to education. If the returns to schooling vary among the population we will have to estimate an earnings equation where the idiosyncratic returns to education are correlated with the number of schooling years. If we estimate such a model by OLS, we will get biased estimate because people with lower returns will tend to invest less in education.

The econometrics literature has recently been interested in the treatment of correlated random coefficient models. Card proposes instrumental variables estimators for his model. Heckman and Vitlacyl [3] and Wooldridge [5], [6] provide identifying assumptions if one is willing to use instrumental variables techniques. Unfortunately, the assumptions proposed might be difficult to fulfill in the presence of endogenous binary variables. It seems on the other hand to be fruitful to investigate the size of the bias of traditional estimators and provide bounds to the parameters of interest.

The idea of this note is to derive the bias of the OLS estimator of a correlated random coefficient model. The model we investigate involves one correlated random coefficient and a binary variable associated with it. We use this latter characteristic of the model and the economic predictions to provide set-identification of the parameters of interest. The results provided in this paper may be useful for practitioners as these kind of models appear frequently in economics.

In section 2, we present the model which consists of a demand equation conditioned on a binary variable, and we discuss the identification of the model with instrumental variables. In section 4, we derive the bias of the OLS estimator. We show that we can give upper or lowerbounds to the parameters of interest of the model. We also show that we can reduce the bias. Section 5 gives some concluding comments.

2 A Correlated Random Coefficient model

In this section, we discuss the estimation of a correlated random coefficient model (CRCM) with cross-section data. We consider the case where there is only one correlated coefficient. Moreover, the random coefficient is correlated with a discrete variable. The aim of the estimation is to recover the structural parameters of the model, particularly the mean of the random coefficient. Consider for example the following model

$$c_{w,i} = \kappa + \alpha x_i + \delta_i h_i + \epsilon_i \tag{1}$$

with $E\left[\epsilon_i | 1, x_i, h_i, \delta_i h_i\right] = 0$ and h_i is a discrete endogenous variable. Equation (1) is a conditional demand equation where c_i^w is work-related expenses, x_i is total expenditures and h_i the work status (see Kolodziejczyk [4] for further details on this model). Since δ_i varies among the population, we can write

$$\delta_i = \delta + v_i \tag{2}$$

with $E\left(\delta_i\right) = \delta$ and $E\left(v_i\right) = 0$. Since v_i is correlated with h_i, then the OLS estimator of a regression of c_i^w on x_i and h_i will be biased.

To see this we can rewrite the model as

$$c_{w,i} = \kappa + \alpha x_i + \delta h_i + \omega_i + \epsilon_i, \tag{3}$$

where we have defined $\omega_i \equiv v_i h_i$. The variable h_i is endogenous, and we assume it can be expressed by the following reduced form equation

$$h_i = 1\left[g\left(v_i, \mathbf{q}_i'\right) > 0\right] \tag{4}$$

where $g(\cdot)$ is some non-linear function and $\mathbf{q}_i$ is a vector of explanatory variables. Equation (4) summarizes the comparison of the utility level obtained in both states and the net gain of taking action $h = 1$ depends on v_i. We assume that that the net gain of action $h = 1$ g is monotonically decreasing with v_i, i.e.

$$\partial g / \partial v_i < 0. \tag{5}$$

What are the assumptions needed to identify δ? Heckman and Vitlacyl [3] and Wooldridge [5], [6] find identifying assumptions for linear versions of this model which allow to use instrumental variables techniques. In order to estimate the model with standard instrumental variables techniques, we have to assume the covariance between the fixed costs and the decision to retire is constant. Rewrite the model as

$$\begin{aligned} \delta_i &= \mathbf{z}_i'\alpha_\delta + v_i \\ h_i &= \mathbf{w}_i'\alpha_h + e_i \end{aligned}$$

where $\mathbf{z}_i$ and $\mathbf{w}_i$ are instruments for δ_i and h_i respectively. We need the following assumption to identify δ

$$E\left[v_i e_i \middle| \mathbf{z}_i, \mathbf{w}_i\right] = k \tag{6}$$

where k is a constant. Heckman and Vytlacil (op. cit.) propose GMM estimators. Unfortunately, because of the non-linearity of the model assumption (6) will probably not hold. Since h_i is a non-linear function, the model will be nonseparable. Furthermore, it might be difficult in some context to find good instruments. Therefore, we prefer to consider the OLS estimator of this random coefficient model, compute its bias and try to see whether we can sign it.

3 Bias of the OLS estimator

In this section we derive the bias of the OLS estimator of model (1). We rewrite equation (3) as

$$c_{w,i}^{*} = \alpha x_i^{*} + \delta h_i^{*} + \omega_i^{*} + \epsilon_i^{*}, \tag{7}$$

where the stars denotes variables in deviations from their mean. We assume x_i is measured without error and we define $\mathbf{x}_i' \equiv (x_i^{*}, h_i^{*})$ and $\beta' \equiv (\alpha, \delta)$. We maintain the assumption that $E[\epsilon | x, h, \omega] = E[\epsilon] = 0$.

The probability limit of the OLS estimator is equal to

$$\text{plim}\, \beta_{OLS} - \beta = E(\mathbf{x}'\mathbf{x})^{-1} E(\mathbf{x}'\omega)$$

where

$$E(\mathbf{x}'\mathbf{x}) = \begin{pmatrix} V(x) & \text{cov}(x,h) \\ \text{cov}(x,h) & V(h) \end{pmatrix}$$

$$E(\mathbf{x}'\omega) = \begin{pmatrix} \text{cov}(x,\omega) \\ \text{cov}(h,\omega) \end{pmatrix}.$$

The elements of these two matrices can be rewritten as

$$\begin{aligned}
V(h) &= P(h=1)(1-P(h=1)),\\
\operatorname{cov}(x,h) &= [E(x)-E(x|\,h=1)]\,P(h=1),\\
\operatorname{cov}(x,\omega) &= \{\operatorname{cov}(x,v|\,h=1)-[E(x)-E(x|\,h=1)]\,E(v|\,h=1)\}\,P(h=1),\\
\text{and }\operatorname{cov}(h,\omega) &= E(v|\,h=1)(1-P(h=1))\,P(h=1)
\end{aligned}$$

which result from the following probability limits

$$\begin{aligned}
\operatorname{plim} N_1/N &= P(h=1),\\
\operatorname{plim}\frac{1}{N}\sum x_i^2 &= E\left(x^2\right),\\
\operatorname{plim}\frac{1}{N_1}\sum v_i h_i &= E(v|\,h=1),\\
\operatorname{plim}\frac{1}{N_1}\sum x_i h_i &= E(x|\,h=1),\\
\text{and }\operatorname{plim}\frac{1}{N_1}\sum v_i x_i h_i &= E(v'x|\,h=1)
\end{aligned}$$

where we have defined N as the number of observations and N_1 as the number of observations where $h=1$.

Using the fact that $E(x)=E(x|\,h=1)\,P(h=1)+E(x|\,h=0)\,P(h=0)$, the bias can be rewritten as

$$\operatorname{plim}\alpha_{OLS}-\alpha=\frac{\operatorname{cov}(x,v|\,h=1)\operatorname{Pr}(h=1)\,P(h=0)}{V(x)\,P(h=0)-[E(x)-E(x|\,h=1)]^2\,P(h=1)} \tag{8}$$

and

$$\operatorname{plim}\delta_{OLS}-\delta=E(v|\,h=1)+\frac{\operatorname{cov}(x,v|\,h=1)\,[E(x)-E(x|\,h=1)]\operatorname{Pr}(h=1)}{V(x)\,P(h=0)-[E(x)-E(x|\,h=1)]^2\,P(h=1)} \tag{9}$$

We make the following additional assumptions.

$$E\left(x|\,h=1\right) > E\left(x|\,h=0\right) \tag{10}$$

and

$$\operatorname{cov}\left[vx|\,h=1\right] > 0 \tag{11}$$

We also assume that v is symmetrically distributed around its mean.

Proposition 1 *Under assumptions (5),(10) and (11), the asymptotic biases of* α_{OLS} *and* δ_{OLS} *are respectively positive and negative.*

Proof. By the Cauchy-Schwarz inequality we have

$$V\left(x\right) - \left[E\left(x|\,h=0\right) - E\left(x|\,h=1\right)\right]^2 P\left(h=0\right) P\left(h=1\right) \geq 0. \tag{12}$$

Under assumptions (11), the right hand-side of (8) is positive. Furthermore, under assumptions (5) and the symmetry assumption for v_i we have $E\left[v|\,h=1\right] < 0$. In addition with assumption (10) this implies that the right hand-side of (9) is negative. ■

These conditions imply that the OLS estimator of δ will be downward biased and will give a lower bound to this parameter. The OLS estimator of α will be upward biased. The OLS estimator cannot identify α and δ, but we can give a lower bound to δ and an upper bound to α.

If we estimate the demand system on the population where $h=0$, we will obtain an estimator of α, which we denote by α^0_{OLS}. Since we no longer have a correlated random coefficient model, this estimator is consistent. Then we can rewrite equation (9) as

$$\operatorname{plim}\delta_{OLS} - \frac{\left[E\left(x\right) - E\left(x|\,h=1\right)\right]}{\Pr\left(h=0\right)} \operatorname{plim}\left(\alpha_{OLS} - \alpha^0_{OLS}\right) = \delta + E\left(v|\,h=1\right). \tag{13}$$

All the objects on the left-hand side of equation (13) are identified and can be estimated consistently. This allows a reduction of the asymptotic bias of δ_{OLS} by using the following estimator

$$\delta_{OLS} - \frac{N}{N - N_1} \left(\overline{x} - \overline{x}_1\right) \left(\alpha_{OLS} - \alpha^0_{OLS}\right)$$

where $\overline{x} = N^{-1} \sum_{i=1}^{N} x_i$ and $\overline{x}_1 = N_1^{-1} \sum_{i=1}^{N} x_i h_i$.

4 Conclusion

In this note we have discussed the estimation of a **correlated** random coefficient model with a unique random coefficient correlated with a binary variable. We have shown that the instrumental variables methods will unlikely give consistent estimates of the parameters of interest. Therefore we have derived the bias of the OLS estimator of the model and provided some set identification, i.e. we were able to give upper-bounds and lower-bounds to these parameters. Furthermore the economics of the problem helped us to provide an estimator whose bias is lower. This may provide useful information for practitioners who would like to estimate such models which often appear in economics.

References

[1] Browning, Martin, Lechene, Valérie (2003), "Children and Demand: Direct and Non-Direct Effects", Review of Economics of the Household 1, 9-31.

[2] Card, David (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", Econometrica 69, 1127-1160

[3] Heckman, James, Vytlacil, Edward (1998), "Instrumental Variables Methods for the Correlated Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling", Journal of Human Resources 30, 974, 987.

[4] Kolodziejczyk, Christophe (2005), "Retirement, Fixed Costs to Work and Home-production: an empirical analysis.", Mimeo

[5] Wooldridge, Jeffrey, "Econometric Analysis of Cross-section and Panel Data", Cambridge, MIT Press, 2002

[6] Wooldridge, Jeffrey M., "Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model", Economics Letters 79, 185-191.