Journal of Econometrics 188 (2015) 135–149

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Identification and estimation in a correlated random coefficients binary response model

Stefan Hoderlein [a,*], Robert Sherman [b,*]

[a] Boston College, Department of Economics, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA

[b] California Institute of Technology, Division of Humanities and Social Sciences, 1200 East California Blvd., Pasadena, CA, USA

ARTICLE INFO

Article history:
Received 29 January 2013
Received in revised form
8 April 2014
Accepted 24 March 2015
Available online 30 April 2015

JEL classification:
C13
C14
C25
C26

Keywords:
Correlated random coefficients binary response model
Instrumental variables
Control variables
Conditional median restrictions
Localize-then-average strategy

ABSTRACT

We study a linear index binary response model with random coefficients B allowed to be correlated with regressors X. We identify the mean of the distribution of B and show how the mean can be interpreted as a vector of expected relative effects. We use instruments and a control vector V to make X independent of B given V. This leads to a localize-then-average approach to both identification and estimation. We develop a $\sqrt{n}$-consistent and asymptotically normal estimator of a trimmed mean of the distribution of B, explore its small sample performance through simulations, and present an application.

1. Introduction

In this paper, we study a linear index binary response model with random coefficients allowed to be correlated with regressors. We call such a model a correlated random coefficients binary response model, or a CRCBR model, for short. We show how to identify the mean of the distribution of random coefficients, and show that this mean can be interpreted as a vector of expected relative effects. We also develop a $\sqrt{n}$-consistent and asymptotically normal estimator of a trimmed mean of this distribution.

One economic motivation for the CRCBR model is similar in spirit to the motivation for a correlated random coefficients wage model used by Heckman and Vytlacil (1998) to estimate the mean rate of return to schooling. These authors argue that schooling is correlated with its random coefficient, the rate of return. Moreover, they state that "a correlated random coefficients model is central to the human capital earnings model". By analogy, consider the binary decision of a married woman to work or not. Level of education may be positively correlated with this decision. However, the weight given to education in making the decision may vary from person to person due to the influence of unobserved factors like ability and motivation. Women of higher ability and motivation may not only seek more education but also give more weight to education in making their work decisions. The result is that the weight given to education is not the same positive constant for each woman, but tends to be higher for women with more education. In other words, the weight given to education is positively correlated with education. A natural way to model this type of heterogeneity is with a linear index binary response model where the regression coefficients, or weights, are random variables allowed to be correlated with the observed regressors.

A similar economic motivation is derived from Wooldridge (2010, pp. 73–76). He shows that correlated random coefficients arise in linear index models whenever an observed covariate interacts with a correlated unobserved covariate. He gives an example of a wage model where education, an observed covariate, interacts with ability, an unobserved covariate likely to be correlated with education. By analogy, consider modeling a binary work decision

* Corresponding authors.
E-mail addresses: stefan_hoderlein@yahoo.com (S. Hoderlein), sherman@amdg.caltech.edu (R. Sherman).

http://dx.doi.org/10.1016/j.jeconom.2015.03.044
0304-4076/

where the decision to work depends on education and ability and there is an interaction between education and ability. This situation can be modeled with a linear index binary response model allowing regressors to be correlated with random coefficients.

We give an application to health economics in Section 5. We model the occurrence of abnormal birth conditions in newborns whose mothers smoked during pregnancy. The average number of cigarettes the mother smoked each day during pregnancy is one of the regressors in the model. We argue that the coefficient of smoking is likely to be positively correlated with smoking, since the level of smoking may be positively correlated with the level of other unobserved risky behaviors (e.g., use of drugs, neglect of proper diet) that may make the occurrence of birth abnormalities more likely.

We now turn to a formal development of the CRCBR model.

Let the latent variable $Y^* = XB^*$, where $X = (X_1, X_2, \ldots, X_k)$ is a $1 \times k$ vector of random explanatory variables and $B^* = (B_1^*, B_2^*, \ldots, B_k^*)'$ is a $k \times 1$ vector of random coefficients allowed to be correlated with X. Take $X_2 \equiv 1$, the intercept term. Then $XB^* = X_1B_1^* + B_2^* + X_3B_3^* + \cdots + X_kB_k^*$. Define $Y = \{Y^* > 0\} = \{XB^* > 0\}$, where $\{A\}$ denotes the indicator function of the event A. Note that $Y = \{XB > 0\}$ where $B = \lambda B^*$ for any $\lambda > 0$. That is, Y is invariant to positive scale normalizations of B^*, and so the distribution of B^* can only be identified up to scale.[1] We assume that $B_1^* > 0$ and take $\lambda = 1/B_1^*$, so that the vector of random coefficients of interest is $B = B^*/B_1^* = (1, B_2, \ldots, B_k)$. For simplicity, from now on, we take $Y^* = XB = X_1 + B_2 + X_3B_3 + \cdots + X_kB_k$.

For ease of exposition, we assume the mean of B exists and write $\beta = (1, \beta_2, \beta_3, \ldots, \beta_k)' = \mathbb{E}B$.[2] Define the CRCBR model

$$Y = \{XB > 0\} \quad (1)$$

$$= \{\epsilon < X\beta\} \quad (2)$$

where $\epsilon = -X(B-\beta) = (\beta_2 - B_2) + (\beta_3 - B_3)X_3 + \cdots + (\beta_k - B_k)X_k$.

Note that the CRCBR model reduces to the traditional nonrandom coefficients binary response model when B_2, the coefficient of the intercept term, is the only random coefficient in the model. In this case, $\epsilon = \beta_2 - B_2$. When B_2 is allowed to be correlated with components of X, we have an endogenous binary response model like those of Rivers and Vuong (1988) and Blundell and Powell (2004). When B_2 is uncorrelated with X, we have an exogenous binary response model like those of Manski (1975, 1985) and Horowitz (1992).

Refer to (2) and note that if any component of X is correlated with any component of B, then $\mathbb{E}X'\epsilon \neq 0$. Thus, correlation between regressors and random coefficients implies endogeneity. We say that a component of X is endogenous if it is correlated with at least one component of B. To handle this endogeneity, we require instruments and control variables.

Let $\mathcal{X}_1$ denote an endogenous component of X. We assume that there exists a vector of instruments, Z, for $\mathcal{X}_1$. This means that at least one component of Z not in X is correlated with $\mathcal{X}_1$, and Z is unrelated to B in a sense to be defined shortly. In addition, we assume that $\mathcal{X}_1 = \phi_1(Z, V_1)$ where V_1 is a random variable independent of Z, and ϕ_1 is a real-valued function invertible in V_1 for each possible value of Z. If Z is independent of B given V_1, then $\mathcal{X}_1$ is independent of B given V_1, and so V_1 is a control variable for $\mathcal{X}_1$ with respect to B, as defined in Imbens and Newey (2009). Under these restrictions, $\mathcal{X}_1$ can be either a separable or nonseparable function of Z and V_1. For example, we allow the standard separable case $\mathcal{X}_1 = \phi_1(Z, V_1) = M(Z) + V_1$ where $M(Z) = \mathbb{E}(\mathcal{X}_1 \mid Z)$ and $V_1 = \mathcal{X}_1 - M(Z)$, as in Blundell and Powell (2004).

Let e denote the number of endogenous components of X and let $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_e)$ denote the vector of endogenous components of X. For simplicity, assume that Z is a vector of instruments for each component of $\mathcal{X}$. Let $V = (V_1, \ldots, V_e)$ denote the vector of control variables for $\mathcal{X}$. That is, as above, for each j, $\mathcal{X}_j = \phi_j(Z, V_j)$ where V_j is independent of Z and ϕ_j is invertible in V_j conditional on Z. If Z is independent of B given V, then $\mathcal{X}$ is independent of B given V. If, in addition, the exogenous components of X are independent of (B, V), then X is independent of B given V. In this setting, conditioning on V, the source of endogeneity, produces conditional independence. This suggests a localize-then-average approach to both identification and estimation.

If X is independent of B given V, and (B, V) satisfies a type of joint symmetry condition, then conditional median restrictions hold, which generalize the median independence restriction of Manski (1975, 1985). These conditional median restrictions are sufficient to identify $\mathbb{E}(B \mid V)$ in a distribution-free way that allows arbitrary forms of heteroscedasticity in a conditional error term. Averaging over V then identifies $\beta = \mathbb{E}B$.

We note that as in the exogenous set-up analyzed by Manski (1975, 1985) and Horowitz (1992), our weak conditional median restrictions, by themselves, preclude identifying features like the average structural function of Blundell and Powell (2004) as well as the associated average marginal effects. See also the discussion in Hoderlein (2010). Identifying such features may be possible under stronger assumptions, but we choose to follow the maximum score approach in this paper, and leave such extensions for future work. However, we do show in the next section that $\beta = \mathbb{E}B$ can be interpreted as a vector of expected relative effects, which can provide useful interpretations in practice, as we illustrate in an application section.

We estimate $\mathbb{E}(B \mid V)$ with a localized version of the smoothed maximum score estimator of Horowitz (1992), and then average the trimmed estimated conditional expectations to obtain an estimator of a trimmed mean of the distribution of B.[3] The conditional expectation estimators suffer from a curse of dimensionality, but the estimator of the trimmed mean does not. The averaging overcomes the curse and yields a $\sqrt{n}$-consistent and asymptotically normal estimator of the trimmed mean. An interesting aspect of the estimation procedure is the localization step. We do not localize directly on V, which would require estimating V, a difficult task, in general. Rather, we localize on an invertible transformation of V, which is easily estimated with kernel regression methods. This simplified localization step is made possible by a simple generalization of a result in Matzkin (2003).

Conditional median restrictions are used to identify the localizing function $\mathbb{E}(B \mid V)$. Specifically, we require a conditional median independence (CMI) condition and a conditional median zero (CMZ) condition. The CMI condition holds whenever X is independent of B given V, a plausible sufficient condition. The CMZ condition is implied by symmetry conditions on the distribution of (B, V) which guarantee that certain conditional means and conditional medians coincide. These symmetry conditions are nontrivial, but they are satisfied for a wide range of possible (B, V) distributions, as we show. As stated earlier, we focus on $\mathbb{E}(B \mid V)$ as the localizing function for ease of exposition, so that averaging over V gives the familiar object $\mathbb{E}B$ as the parameter of interest. However, other reasonable measures of center of the distribution of B can be identified using other localizing functions which satisfy CMI and CMZ. In the next section we describe examples where $\mathbb{E}(B \mid V)$ (and therefore

[1] For related discussion, see Gautier and Kitamura (2009) and Ichimura and Thompson (1998).

[2] As we show in Section 2, our methods do not require that the mean of B exists, and even if the mean exists, β need not be the mean, but can denote any reasonable measure of center of the distribution of B.

[3] As discussed in Section 3 and in online Appendix B, we use trimming as a purely technical device to assist in asymptotic arguments. There need be no practical difference between the trimmed mean and the mean β. The trimmed mean can be made arbitrarily close to β by choosing a large enough trimming constant.