Economics Letters 113 (2011) 12–15

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

A simple method for estimating unconditional heterogeneity distributions in correlated random effects models

Jeffrey M. Wooldridge *

Department of Economics, Michigan State University, 110 Marshall-Adams Hall, East Lansing, MI 48824-1038, United States

ARTICLE INFO

Article history:
Received 3 April 2011
Received in revised form
11 May 2011
Accepted 23 May 2011
Available online 6 June 2011

JEL classification:
C23
C33

Keywords:
Unobserved heterogeneity
Correlated random effects
Unconditional distribution
Conditional distribution

ABSTRACT

I propose a general, simple approach to recovering an unconditional heterogeneity distribution when a conditional distribution has been estimated. The approach can be applied to cross section models and panel data models – both static and dynamic – with unobserved heterogeneity.

1. Introduction

The correlated random effects (CRE) approach to panel data models dates back to Mundlak (1978) in the linear case and Chamberlain (1980) in the nonlinear case. Wooldridge (2010) illustrates how several popular models – including linear models, binary response models, ordered response models, models for corner solution responses, and count data models – can be specified and estimated using a CRE approach.

The CRE approach has several benefits, including that it is usually straightforward to implement and that quantities of interest, such as average partial (or marginal) effects, are easy to recover. The starting point for a CRE approach is typically the same as the so-called "fixed effects" approaches. Namely, one specifies a marginal or joint distribution for a response or a set of responses conditional on observed covariates and unobserved heterogeneity. The second step, unique to the CRE approach, is to model the conditional distribution of heterogeneity given a set of observable covariates.

As shown by Altonji and Matzkin (2005) and Wooldridge (2005a), one need not be able to fully identify the conditional or unconditional heterogeneity distributions in order to estimate average partial effects (APEs). Wooldridge (2010) covers several common parametric examples – including linear, probit, and Tobit CRE models – that demonstrate how the APEs are identified even when the distribution of heterogeneity is not. ((Chamberlain, 1984), first demonstrated identification of APEs for the CRE probit model in the presence of serial correlation of unknown form.) Thus, in many applications one can estimate quantities of interest without making extra assumptions needed to characterize the heterogeneity distribution.

Nevertheless, one is sometimes interested in estimating the unconditional distribution of heterogeneity in the population — even if it means imposing extra assumptions. In this paper, I use a simple result from probability theory that relates an unconditional density to the expected value of a conditional density. This result allows one to identify and consistently estimate an unconditional heterogeneity distribution when a distribution of heterogeneity conditional on observed variables is identified. Inference for the unconditional density is relatively straightforward.

Before continuing, I should emphasize that, when applied to panel data models, I am thinking of cases where there are relatively few time periods for each cross section unit. With many time periods per unit, a "fixed effects" approach – where unobserved heterogeneity is estimated along with the population parameters – is common, in which case those estimates can be used to estimate the distribution of heterogeneity. For a recent example, see Jones and Schurer (2011).

* Tel.: +1 517 353 5972; fax: +1 517 432 1068.
E-mail address: wooldri1@msu.edu.

0165-1765/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2011.05.019

2. A simple relationship between marginal and conditional densities

Let $(\mathbf{U}, \mathbf{W})$ be a random vector with marginal densities $f_{\mathbf{U}}(\cdot)$ and $f_{\mathbf{W}}(\cdot)$. Technically, these are with respect to σ-finite measures, say $\eta_{\mathbf{U}}$ and $\eta_{\mathbf{W}}$. Let $f_{\mathbf{U}|\mathbf{W}}(\cdot|\cdot)$ be the conditional density with respect to the measure $\eta_{\mathbf{U}}$. For each $\mathbf{w} \in \mathcal{W}$ (the support of $\mathbf{W}$), $f_{\mathbf{U}|\mathbf{W}}(\cdot|\mathbf{w})$ can be discrete, continuous, or have both features. We assume that $f_{\mathbf{U}|\mathbf{W}}(\cdot|\mathbf{w})$ is a density with respect to the measure $\eta_{\mathbf{U}}$, which simply means that the nature of $\mathbf{U}$ – for example, whether it is discrete or continuous – does not change with the conditioning values $\mathbf{w}$.

It is well known (for example Billingsley, 1995) that we can obtain $f_{\mathbf{U}}(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{U}$ by integrating (in the measure-theoretic sense) the conditional density against $f_{\mathbf{W}}(\mathbf{w})$:

$$f_{\mathbf{U}}(\mathbf{u}) = \int_{\mathcal{W}} f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}) f_{\mathbf{W}}(\mathbf{w})\eta_{\mathbf{W}}(\mathrm{d}\mathbf{w}). \tag{2.1}$$

For the purposes of this paper, the key result is expressing (2.1) as an expected value with respect to the distribution of $\mathbf{W}$. In particular, for any given $\mathbf{u} \in \mathcal{U}$ (the support of $\mathbf{U}$), (2.1) can be written as

$$f_{\mathbf{U}}(\mathbf{u}) = E_{\mathbf{W}}[f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{W})], \tag{2.2}$$

where the notation $E_{\mathbf{W}}[\cdot]$ is used to emphasize that the expectation is with respect to the distribution of $\mathbf{W}$. The expression in (2.2) is useful for estimating $f_{\mathbf{U}}(\mathbf{u})$ quite generally when we replace $E_{\mathbf{W}}[\cdot]$ with its sample analog.

3. A general estimation strategy

Now let $f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}; \boldsymbol{\gamma})$ be a parametric model for the conditional density, and assume that it is correctly specified: for some $\boldsymbol{\gamma}_o \in \Gamma$, $f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}; \boldsymbol{\gamma}_o)$ is the true conditional density. Let $\hat{\boldsymbol{\gamma}}$ be a $\sqrt{N}$-consistent, asymptotically normal estimator of $\boldsymbol{\gamma}_o$ based on a random sample $\{(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{W}_i) : i = 1, \ldots, N\}$. Typically, $\hat{\boldsymbol{\gamma}}$ is obtained along with an estimator $\hat{\boldsymbol{\theta}}$ that indexes some feature of $D(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{W}_i, \mathbf{U}_i)$ – sometimes a complete density and other times a conditional expectation.

Under weak regularity conditions, a $\sqrt{N}$-consistent, asymptotically normal estimator of $f_{\mathbf{U}}(\mathbf{u})$ is obtained as

$$\hat{f}_{\mathbf{U}}(\mathbf{u}) = N^{-1}\sum_{i=1}^{N} \hat{f}_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{W}_i) \equiv N^{-1}\sum_{i=1}^{N} f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{W}_i; \hat{\boldsymbol{\gamma}}). \tag{3.1}$$

Underlying consistency is the weak law of large numbers because

$$N^{-1}\sum_{i=1}^{N} f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{W}_i; \boldsymbol{\gamma}_o) \xrightarrow{p} E_{\mathbf{W}}[f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\boldsymbol{W}; \boldsymbol{\gamma}_o)] = f_{\mathbf{U}}(\mathbf{u}), \tag{3.2}$$

where the equality follows from Eq. (2.2). Replacing $\boldsymbol{\gamma}_o$ with $\hat{\boldsymbol{\gamma}}$ generally preserves consistency. Continuity of $f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}; \boldsymbol{\gamma})$ in $\boldsymbol{\gamma}$ and a boundedness condition on the moments $E_{(\mathbf{U},\mathbf{W})}[f_{\mathbf{U}|\mathbf{W}}(\mathbf{U}|\boldsymbol{W}; \boldsymbol{\gamma})]$, across all possible values of $\boldsymbol{\gamma}$, suffice. See, for example, Wooldridge (2010, Lemma 12.1). Asymptotic normality can be obtained from Wooldridge (2010, Problem 12.17); in particular, the delta method can be used to obtain a valid asymptotic standard error for any $\mathbf{u}$. Bootstrapping can be used, although it could be computationally intensive to obtain a standard error for numerous values of $\mathbf{u}$.

When $\mathbf{W}_i$ is a discrete random vector taking on G values, say $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_G\}$, the estimator in (3.1) is simple to characterize:

$$\begin{aligned}\hat{f}_{\mathbf{U}}(\mathbf{u}) &= N^{-1}\sum_{\mathbf{W}_i=\mathbf{w}_1}^{N} f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}_1; \hat{\boldsymbol{\gamma}}) + \cdots \\ &\quad + N^{-1}\sum_{\mathbf{W}_i=\mathbf{w}_G}^{N} f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}_G; \hat{\boldsymbol{\gamma}}) \\ &= \hat{\rho}_1 f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}_1; \hat{\boldsymbol{\gamma}}) + \cdots + \hat{\rho}_G f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}_G; \hat{\boldsymbol{\gamma}}),\end{aligned} \tag{3.3}$$

where $\hat{\rho}_g = N_g/N$ is the fraction of the sample with $\mathbf{W}_i = \mathbf{w}_g$. The expression in (3.3) shows that $\hat{f}_{\mathbf{U}}(\mathbf{u})$ is a finite mixture of the estimated conditional densities, with mixing probabilities equal to the sample proportions. Note that this has nothing to do with the nature of the distribution of $\mathbf{U}$; $f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{w}; \boldsymbol{\gamma})$ can be discrete, continuous, or have both features.

By construction, the moments of the estimated distribution can be obtained by a sample version of the law of iterated expectations. For a function $q(\mathbf{U}_i)$, suppose $E[q(\mathbf{U}_i)|\mathbf{W}_i] = m(\mathbf{W}_i; \boldsymbol{\gamma})$. Then the estimator of $\mu_q \equiv E[q(\mathbf{U}_i)]$ obtained from integrating (3.1) is simply

$$\begin{aligned}\hat{\mu}_q &= N^{-1}\sum_{i=1}^{N}\int_{\mathcal{U}} q(\mathbf{u}) f_{\mathbf{U}|\mathbf{W}}(\mathbf{u}|\mathbf{W}_i; \hat{\boldsymbol{\gamma}})\eta_{\mathbf{U}}(\mathrm{d}\mathbf{u}) \\ &= N^{-1}\sum_{i=1}^{N} m(\mathbf{W}_i; \hat{\boldsymbol{\gamma}}),\end{aligned} \tag{3.4}$$

where $m(\mathbf{w}; \hat{\boldsymbol{\gamma}})$ is the conditional mean function obtained from $f_{\mathbf{U}|\mathbf{W}}(\cdot|\mathbf{w}; \hat{\boldsymbol{\gamma}})$.

4. Application to cross section models

As one application of the estimator on Section 3, consider a standard heteroskedastic probit model, written as

$$y_i = 1[\alpha_o + \mathbf{x}_i\boldsymbol{\beta}_o + u_i > 0]$$
$$D(u_i|\mathbf{x}_i) = Normal[0, \exp(\mathbf{x}_i\boldsymbol{\gamma}_o)],$$

where y_i is the binary response. The parameters $(\alpha_o, \boldsymbol{\beta}_o, \boldsymbol{\gamma}_o)$ are easily estimated via maximum likelihood using the response probabilities

$$P(y_i = 1|\mathbf{x}_i) = \Phi[\exp(-\mathbf{x}_i\boldsymbol{\gamma}_o/2)(\alpha_o + \mathbf{x}_i\boldsymbol{\beta}_o)]. \tag{4.1}$$

The MLE is programmed into popular software packages.

As shown by Wooldridge (2005b), consistent estimation of the average partial effects (APEs) is possible without estimating the unconditional heterogeneity distribution. To this end, the average structural function [Blundell and Powell (2004)] is defined by

$$ASF(\mathbf{x}) = E_{u_i}\{1[\alpha_o + \boldsymbol{x}\boldsymbol{\beta}_o + u_i > 0]\}.$$

For any fixed vector $\mathbf{x}$, a consistent estimator of $ASF(\mathbf{x})$ is

$$\widehat{ASF}(\mathbf{x}) = N^{-1}\sum_{i=1}^{N}\Phi[\exp(-\mathbf{x}_i\hat{\boldsymbol{\gamma}}/2)(\hat{\alpha} + \mathbf{x}\hat{\boldsymbol{\beta}})]. \tag{4.2}$$

Average partial effects are estimated by computing derivatives and changes of $\widehat{ASF}(\mathbf{x})$. Thus, for the purposes of estimating directions and magnitudes of the effects, the unconditional distribution of u_i is not needed.

We may be curious, though, to have some sense of the shape of the density of u_i. We know that the conditional distribution, $D(u_i|\mathbf{x}_i)$, has the familiar bell shape, centered at zero (because of the intercept included in the model) but with variances generally changing with $\mathbf{x}_i$. We can write the conditional density as

$$f_{u_i|\mathbf{x}_i}(u|\mathbf{x}) = \exp(-\boldsymbol{x}\boldsymbol{\gamma}_o/2)\phi[\exp(-\boldsymbol{x}\boldsymbol{\gamma}_o/2)u] \tag{4.3}$$

for any $\mathbf{x}$ and u, where $\phi(\cdot)$ is the standard normal density function. Since we have consistent estimators of the unknown parameters we can use Eq. (3.2) to estimate the unconditional density, $f_{u_i}(\cdot)$, at each point:

$$\begin{aligned}\hat{f}_{u_i}(u) &= N^{-1}\sum_{i=1}^{N}\exp(-\mathbf{x}_i\hat{\boldsymbol{\gamma}}/2)\phi[\exp(-\mathbf{x}_i\hat{\boldsymbol{\gamma}}/2)u] \\ &= N^{-1}\sum_{i=1}^{N}(2\pi)^{-1/2}\exp(-\mathbf{x}_i\hat{\boldsymbol{\gamma}}/2) \\ &\quad \times \exp[-\exp(-\mathbf{x}_i\hat{\boldsymbol{\gamma}})u^2/2].\end{aligned} \tag{4.4}$$