Journal of Econometrics 151 (2009) 47–55

Contents lists available at ScienceDirect

Journal of Econometrics

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

GMM redundancy results for general missing data problems[☆]

Artem Prokhorov[a], Peter Schmidt[b,c,*]

[a] Department of Economics, Concordia University, Montreal, PQ H3G1M8, Canada

[b] Department of Economics, Michigan State University, East Lansing, MI 48824, United States

[c] School of Economics and Business, Yonsei University, Seoul 120-749, Republic of Korea

ARTICLE INFO

Article history:
Received 25 June 2006
Received in revised form
3 March 2009
Accepted 4 March 2009
Available online 20 March 2009

JEL classification:
C13

Keywords:
Generalized method of moments
Inverse probability weighting
Missing at random
Selectivity

ABSTRACT

We consider questions of efficiency and redundancy in the GMM estimation problem in which we have two sets of moment conditions, where two sets of parameters enter into one set of moment conditions, while only one set of parameters enters into the other. We then apply these results to a selectivity problem in which the first set of moment conditions is for the model of interest, and the second set of moment conditions is for the selection process. We use these results to explain the counterintuitive result in the literature that, under an ignorability assumption that justifies GMM with weighted moment conditions, weighting using estimated probabilities of selection is better than weighting using the true probabilities. We also consider estimation under an exogeneity of selection assumption such that both the unweighted and the weighted moment conditions are valid, and we show that when weighting is not needed for consistency, it is also not useful for efficiency.

1. Introduction

This paper is motivated by a puzzle in the missing data (selectivity) literature. Consider the setting of a GMM problem is which we have a set of moment conditions, with some parameters θ_1 (the "parameters of interest"), and these moment conditions hold in the unselected sample. However, we also have a selection mechanism such that the moment conditions do not hold in the selected sample. Under certain assumptions given below (typically referred to as "ignorability" or "selection on observables"), weighting the original moment conditions by the inverse of the probability of selection yields a modified set of moment conditions that do hold in the selected sample. We will follow Wooldridge (2002b, 2007) in calling the estimator based on these weighted moment conditions the "inverse probability weighting" (IPW) estimator.

Unless the probability of selection is known for each selected observation, implementation of the IPW estimator will require a model for the probability of selection. Let θ_2 be the parameters (the "selection parameters") in the moment conditions derived from this model. Typically these moment conditions will be based on the score function from the likelihood function for the selection process. A two-step IPW procedure can be considered, in which the first step is the estimation of θ_2 from the selection model, and the second step is the estimation of θ_1 by GMM on the weighted moment conditions, where the weighting is done using the estimated probabilities of selection.

In this setting, the puzzle is that it is better to estimate the selection probabilities than to use the true selection probabilities, even if the latter are known. In other words, in terms of the augmented model described above, we get a better estimator of θ_1 when we use the estimated θ_2 in the second step than if we used the true θ_2. This phenomenon has been discussed by Wooldridge (1999, 2001, 2002b, 2007), and it has also been noted in a number of previous works, including Pierce (1982), Rosenbaum (1987), Imbens (1992), Robins et al. (1992), Robins and Rotnitzky (1995), Hirano et al. (2003), Henmi and Eguchi (2004) and Hitomi et al. (2008). This is puzzling because knowledge of θ_2, if properly exploited, cannot be harmful.

To resolve this puzzle, we follow Newey and McFadden (1994) in setting up an augmented set of moment conditions, where the first subset are the weighted original moment conditions, which now contain both θ_1 and θ_2, and the second subset are the moment conditions from the selection model, which contain only θ_2. We show that the second set of moment conditions is useful (non-redundant), even when θ_2 is known. This is true because the second

☆ Helpful comments from two referees and an associate editor, J. Wooldridge and seminar participants at Michigan State, Rice and Monash Universities, University of New South Wales, the 2004 Mannheim Empirical Research Summer School and the 33rd Annual Australian Conference of Economists are gratefully acknowledged.

* Corresponding author. Tel.: +1 517 355 8381; fax: +1 517 432 1068.
E-mail address: schmidtp@msu.edu (P. Schmidt).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.03.010

set of moment conditions is correlated with the first set in the selected sample (even though it is not in the full sample). So the inefficiency of the estimator based on known θ_2 and the first set of moment conditions only is due to its failure to exploit the information in the second set of moment conditions; whereas, when θ_2 is not known, there is no choice but to include the second set of moment conditions.

This raises the question of whether, when θ_2 is known, we can improve on the two-step estimator (which uses estimated θ_2 in the second step) by using a GMM estimator based on both sets of moment conditions, but where only θ_1 is estimated. After all, this GMM estimator cannot be worse than the two-step estimator of θ_1. The answer to this question is a bit complicated. In the case that the original GMM problem (the one that contains the parameter of interest) is overidentified, the two-step estimator is dominated by a one-step estimator that estimates θ_1 and θ_2 jointly in the augmented GMM model. However, we show that, in the augmented GMM model, knowledge of θ_2 is redundant (does not improve the precision of estimation of θ_1). So, while it can never hurt to know more, if that knowledge is used properly, in this case it does not help either.

The result just quoted is given in Section 3 of the paper. In Section 2, we set the stage by giving a number of results on efficiency and redundancy of estimation in a general GMM setting, when one set of moment conditions depends on θ_1 and θ_2, while a second set of moment conditions depends only on θ_2. Some of these results are original and interesting in their own right. We consider "m-redundancy", which is redundancy of moment conditions in the sense of Breusch et al. (1999), and we also consider "p-redundancy", which is a term we propose to refer to redundancy of the knowledge of some of the parameters for estimation of the other parameters. One of our results gives an interesting connection between these two concepts: the first set of moment conditions with θ_1 known is m-redundant for estimation of θ_2 if and only if knowledge of θ_2 is p-redundant for estimation of θ_1.

In Section 4 of the paper we reconsider the selectivity model under a stronger "exogeneity of selection" assumption under which both the unweighted moment conditions and the weighted moment conditions hold in the selected population. Wooldridge (2001) has shown that in this circumstance it is better to use the unweighted moment conditions than the weighted moment conditions. However, this does not rule out the possibility that it would be better to use both. We show that, given exogeneity of selection, the optimal moment conditions in the selected population are the same as in the unselected population. If weighting was not part of the efficient estimation problem in the unselected sample, it also plays no role in the selected sample. In this sense, when we do not have to weight for reasons of consistency, we also do not have to weight for reasons of efficiency.

GMM is sufficiently general to accommodate most of the extremum and minimum distance estimators in econometrics (see, e.g., Newey and McFadden (1994), p. 2118). The arguments we present can be applied, for example, to (Q)MLE, M-estimation, WLS, and NLS. They also extend to the asymptotic equivalents of GMM such as empirical likelihood and exponential tilting estimators. Hence, our results apply quite generally. Specifically, they relate to the treatment effect estimation literature (e.g., Rosenbaum and Rubin (1983) and Heckman et al. (1998)), to the stratified-sampling literature (e.g., Manski and Lerman (1977), Manski and McFadden (1981), Cosslett (1981a,b), Imbens (1992) and Tripathi (2003)) and other similarly-structured problems (e.g., Hellerstein and Imbens (1999), Nevo (2002, 2003) and Crepon et al. (1997)). Also, our results of Section 2 apply to a number of other settings in which two-step estimators arise, including the generated regressors of Pagan (1984), the latent variables models of Zellner (1970) and Goldberger (1972), and many others. However, we do not consider semiparametric estimation of the selection model ("propensity score"), as in Hahn (1998) or Hirano et al. (2003).

2. Efficiency and redundancy results for the general estimation problem

2.1. Preliminaries

Consider a random vector $w^* \in \mathcal{M}^* \subset \mathbb{R}^{\dim(w^*)}$, the compact set $\Theta = \Theta_1 \times \Theta_2 \subset \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$, and the population condition

$$\mathbb{E}[h(w^*, \theta)] = 0, \tag{1}$$

where $h : \mathcal{M}^* \times \Theta \to \mathbb{R}^m$ is a vector of known real-valued moment functions. Under regularity conditions, Hansen (1982) established consistency and asymptotic normality of the generalized method of moments (GMM) estimator that minimizes a squared Euclidean distance of the random sample analogues of the population moments, $\bar{h}(\theta) = \frac{1}{n}\sum_{i=1}^{n} h(w_i^*, \theta)$, from their population counterparts equal to zero. Thus, the GMM estimator $\hat{\theta}$ minimizes the objective function

$$\bar{h}(\theta)'\hat{\Omega}\bar{h}(\theta), \tag{2}$$

where $\hat{\Omega}$ converges in probability to Ω, the appropriate (optimal) positive semidefinite weighting matrix.

For simplicity, we assume here that w_i^*, $i = 1, \ldots, n$, are i.i.d.

The following regularity assumptions on the moment functions are sufficiently strong to ensure both consistency and asymptotic normality of the GMM estimator.

Assumption 2.1. Let $\|\cdot\|$ denote the Euclidean norm, $\mathbb{B}(\theta, \delta) \subset \Theta$ denote an open $p_1 + p_2$-ball of radius δ with center at θ, $\nabla_\theta h(\cdot, \theta)$ denote the $m \times (p_1 + p_2)$ Jacobian of $h(\cdot, \theta)$ with respect to θ, and "w.p.1" stand for "with probability one". Assume that the moment function in (1) satisfies the following conditions:

(i) $\exists$ unique $\theta_o \in int(\Theta)$ that solves (1);
(ii) $h(w^*, \theta)$ is continuous at each $\theta \in \Theta$ w.p.1;
(iii) $h(w^*, \theta)$ is (once) continuously differentiable on $\mathbb{B}(\theta_o, \delta)$ for some $\delta > 0$ w.p.1;
(iv) $\mathbb{E}\{\sup_{\theta \in \Theta} \|h(w^*, \theta)\|^2\} < \infty$;
(v) $\mathbb{E}\{\sup_{\theta \in \mathbb{B}(\theta_o, \delta)} \|\nabla_\theta h(w^*, \theta)\|\} < \infty$ for some $\delta > 0$;
(vi) $\mathbb{E}[\nabla_\theta h(w^*, \theta_o)]$ is of full column rank.

Then it is a standard result (see, e.g., Newey and McFadden (1994), Theorems 2.6 and 3.4) that, under Assumption 2.1, the GMM estimator of θ is consistent and asymptotically normal.

2.2. The general estimation problem

Suppose that we can partition θ into subsets of parameters $(\theta_1', \theta_2')'$ and $h(\cdot)$ into subsets of functions $(h_1(\cdot)', h_2(\cdot)')'$. If we suppress w^* for notational convenience, we can write

$$\begin{aligned} &\mathbb{E}[h_1(\theta_1, \theta_2)] = 0, \quad (A) \\ &\mathbb{E}[h_2(\theta_2)] = 0, \quad (B) \end{aligned} \tag{3}$$

where $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$, $h_1(\cdot)$ and $h_2(\cdot)$ are m_1- and m_2-vectors of known functions, respectively ($m = m_1 + m_2$). We consider the general case of overidentification, i.e., $m_1 \geq p_1$ and $m_2 \geq p_2$. These identification conditions (plus the corresponding rank conditions assumed below) ensure that θ_2 is identified by (B) alone, and that, given θ_2, θ_1 is identified by (A) alone, so that two-step estimation is possible.

The optimal weighting matrix for GMM will be the inverse of the following covariance matrix or its components:

$$C = \mathbb{V}[h(\theta)] = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{4}$$