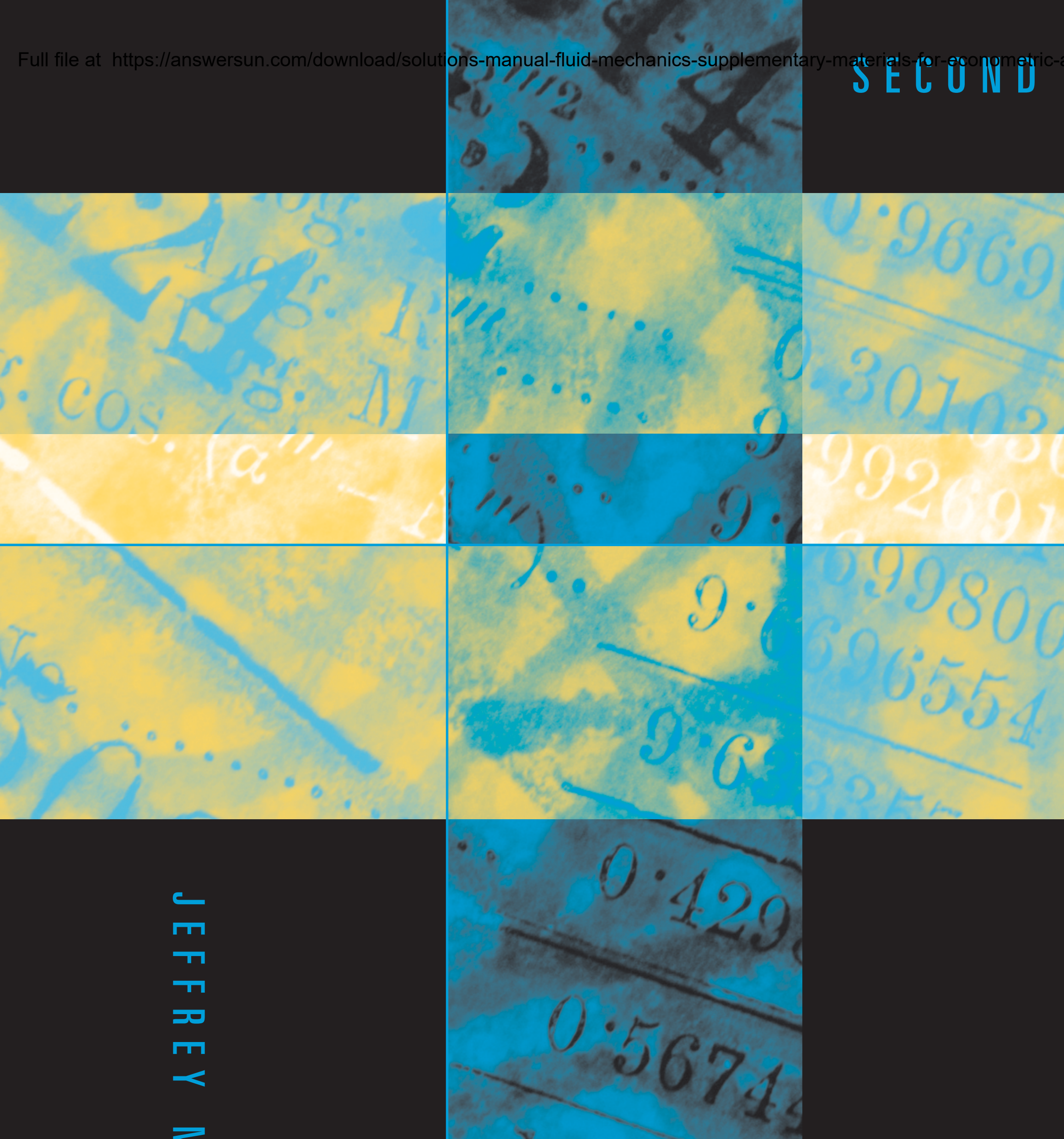
SECOND EDITION



JEFFREY M. WOOLDRIDGE

ECONOMETRIC ANALYSIS OF CROSS SECTION AND PANEL DATA

Econometric Analysis of Cross Section and Panel Data

Econometric Analysis of Cross Section and Panel Data

Second Edition

Jeffrey M. Wooldridge

The MIT Press
Cambridge, Massachusetts
London, England

For information about special quantity discounts, please email special_sales@mitpress.mit.edu

This book was set in Times Roman by Asco Typesetters, Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Wooldridge, Jeffrey M.
Econometric analysis of cross section and panel data / Jeffrey M. Wooldridge.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-23258-6 (hardcover : alk. paper)
1. Econometrics—Asymptotic theory. I. Title.
HB139.W663 2010
330.01′5195—dc22 2010020912

10 9 8 7 6 5 4 3 2 1

Contents

To obtain an estimable equation, replace q in equation (4.19) with equation (4.27) to get

$$y = (\beta_0 + \gamma\theta_0) + \beta_1 x_1 + \cdots + \beta_K x_K + \gamma\theta_1 z + (\gamma r + v). \tag{4.28}$$

Under the assumptions made, the composite error term $u \equiv \gamma r + v$ is uncorrelated with x_j for all j; redundancy of z in equation (4.18) means that z is uncorrelated with v and, by definition, z is uncorrelated with r. It follows immediately from Theorem 4.1 that the OLS regression y on $1, x_1, x_2, \ldots, x_K, z$ produces consistent estimators of $(\beta_0 + \gamma\theta_0), \beta_1, \beta_2, \ldots, \beta_K$, and $\gamma\theta_1$. Thus, we can estimate the partial effect of each of the x_j in equation (4.18) under the proxy variable assumptions.

When z is an **imperfect proxy**, then r in equation (4.27) is correlated with one or more of the x_j. Generally, when we do not impose condition (4.26) and write the linear projection as

$$q = \theta_0 + \rho_1 x_1 + \cdots + \rho_K x_K + \theta_1 z + r,$$

the proxy variable regression gives plim $\hat{\beta}_j = \beta_j + \gamma\rho_j$. Thus, OLS with an imperfect proxy is inconsistent. The hope is that the ρ_j are smaller in magnitude than if z were omitted from the linear projection, and this can usually be argued if z is a reasonable proxy for q; but see the end of this subsection for further discussion.

If including z induces substantial collinearity, it might be better to use OLS without the proxy variable. However, in making these decisions we must recognize that including z reduces the error variance if $\theta_1 \neq 0$: $\text{Var}(\gamma r + v) < \text{Var}(\gamma q + v)$ because $\text{Var}(r) < \text{Var}(q)$, and v is uncorrelated with both r and q. Including a proxy variable can actually reduce asymptotic variances as well as mitigate bias.

Example 4.3 (Using IQ as a Proxy for Ability): We apply the proxy variable method to the data on working men in NLS80.RAW, which was used by Blackburn and Neumark (1992), to estimate the structural model

$$\begin{aligned}\log(wage) = {} & \beta_0 + \beta_1\, exper + \beta_2\, tenure + \beta_3\, married \\ & + \beta_4\, south + \beta_5\, urban + \beta_6\, black + \beta_7\, educ + \gamma\, abil + v,\end{aligned} \tag{4.29}$$

where *exper* is labor market experience, *married* is a dummy variable equal to unity if married, *south* is a dummy variable for the southern region, *urban* is a dummy variable for living in an SMSA, *black* is a race indicator, and *educ* is years of schooling. We assume that *IQ* satisfies the proxy variable assumptions: in the linear projection $abil = \theta_0 + \theta_1 IQ + r$, where r has zero mean and is uncorrelated with IQ, we also assume that r is uncorrelated with experience, tenure, education, and other factors

N very large relative to T, there is no need to downweight correlations between time periods that are far apart, as in the Newey and West (1987) estimator applied to time series problems. Ziliak and Kniesner (1998) do use a Newey-West type procedure in a panel data application with large N. Theoretically, this is not required, and it is not completely general because it assumes that the underlying time series are weakly dependent. (See Wooldridge (1994a) for discussion of weak dependence in time series contexts.) A Newey-West type estimator might improve the finite-sample performance of the GMM estimator.

The asymptotic variance of the optimal GMM estimator is estimated as

$$\left[(\mathbf{X}'\mathbf{Z})\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1}(\mathbf{Z}'\mathbf{X})\right]^{-1}, \tag{8.37}$$

where $\hat{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$; asymptotically, it makes no difference whether the first-stage residuals $\check{\mathbf{u}}_i$ are used in place of $\hat{\mathbf{u}}_i$. The square roots of diagonal elements of this matrix are the asymptotic standard errors of the optimal GMM estimator. This estimator is called a **minimum chi-square estimator**, for reasons that will become clear in Section 8.5.2.

When $\mathbf{Z}_i = \mathbf{X}_i$ and the $\hat{\mathbf{u}}_i$ are the system OLS residuals, expression (8.37) becomes the robust variance matrix estimator for SOLS [see expression (7.28)]. This expression reduces to the robust variance matrix estimator for FGLS when $\mathbf{Z}_i = \hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i$ and the $\hat{\mathbf{u}}_i$ are the FGLS residuals [see equation (7.52)].

8.3.4 The Generalized Method of Moments Three-Stage Least Squares Estimator

The GMM estimator using weighting matrix (8.36) places no restrictions on either the unconditional or conditional (on $\mathbf{Z}_i$) variance matrix of $\mathbf{u}_i$: we can obtain the asymptotically efficient estimator without making additional assumptions. Nevertheless, it is still common, especially in traditional simultaneous equations analysis, to assume that the conditional variance matrix of $\mathbf{u}_i$ given $\mathbf{Z}_i$ is constant. This assumption leads to a system estimator that is a middle ground between system 2SLS and the always-efficient minimum chi-square estimator.

The **GMM three-stage least squares (GMM 3SLS) estimator** (or just 3SLS when the context is clear) is a GMM estimator that uses a particular weighting matrix. To define the 3SLS estimator, let $\check{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\check{\boldsymbol{\beta}}$ be the residuals from an initial estimation, usually system 2SLS. Define the $G \times G$ matrix

$$\hat{\boldsymbol{\Omega}} \equiv N^{-1}\sum_{i=1}^{N}\check{\mathbf{u}}_i\check{\mathbf{u}}_i'. \tag{8.38}$$

natural than the strict exogeneity assumption, which requires conditioning on future values of $\mathbf{x}_{it}$ as well. As we proceed, it is important to remember that equation (11.52) is what we should have in mind when interpreting the estimates of $\boldsymbol{\beta}$. Estimating equations in first differences, such as (11.35), do not have natural interpretations when the explanatory variables are only sequentially exogenous.

As we will explicitly show in the next subsection, models with lagged dependent variables are naturally analyzed under sequential exogeneity. Keane and Runkle (1992) argue that panel data models with heterogeneity for testing rational expectations hypotheses do not satisfy the strict exogeneity requirement. But they do satisfy sequential exogeneity; in fact, the conditioning set in assumption (11.51) can include all variables observed at time $t-1$.

As we saw in Section 7.2, in panel data models without unobserved effects, strict exogeneity is sometimes too strong an assumption, even in static and finite distributed lag models. For example, suppose

$$y_{it} = \mathbf{z}_{it}\gamma + \delta h_{it} + c_i + u_{it}, \tag{11.53}$$

where $\{\mathbf{z}_{it}\}$ is strictly exogenous and $\{h_{it}\}$ is sequentially exogenous:

$$\mathrm{E}(u_{it} \mid \mathbf{z}_i, h_{it}, \ldots, h_{i1}, c_i) = 0. \tag{11.54}$$

Further, h_{it} is influenced by past y_{it}, say

$$h_{it} = \mathbf{z}_{it}\xi + \eta y_{i,t-1} + \psi c_i + r_{it}. \tag{11.55}$$

For example, let y_{it} be per capita condom sales in city i during year t, and let h_{it} be the HIV infection rate for city i in year t. Model (11.53) can be used to test whether condom usage is influenced by the spread of HIV. The unobserved effect c_i contains city-specific unobserved factors that can affect sexual conduct, as well as the incidence of HIV. Equation (11.55) is one way of capturing the fact that the spread of HIV depends on past condom usage. Generally, if $\mathrm{E}(r_{i,t+1}u_{it}) = 0$, it is easy to show that $\mathrm{E}(h_{i,t+1}u_{it}) = \eta\mathrm{E}(y_{it}u_{it}) = \eta\mathrm{E}(u_{it}^2) > 0$ if $\eta > 0$ under equations (11.54) and (11.55). Therefore, strict exogeneity fails unless $\eta = 0$.

Sometimes in panel data applications one sees variables that are thought to be contemporaneously endogenous appear with a lag, rather than contemporaneously. So, for example, we might use $h_{i,t-1}$ in place of h_{it} in equation (11.53) because we think h_{it} and u_{it} are correlated. As an example, suppose y_{it} is percentage of flights cancelled by airline i in year t, and h_{it} is profits in the same year. We might specify $y_{it} = \mathbf{z}_{it}\gamma + \delta h_{i,t-1} + c_i + u_{it}$ for strictly exogenous $\mathbf{z}_{it}$. Of course, at $t+1$, the regressors are $\mathbf{x}_{i,t+1} = (\mathbf{z}_{i,t+1}, h_{it})$, which is correlated with u_{it} if h_{it} is. As we discussed in where, in most cases, $\mathbf{s}(\mathbf{w})$ is the score of an objective function (evaluated at $\boldsymbol{\theta}_o$) and $\mathbf{A}$ is the expected value of the Jacobian of the score, again evaluated at $\boldsymbol{\theta}_o$. (We suppress an "o" subscript here, as the value of the true parameter is irrelevant.) All M-estimators with twice continuously differentiable objective functions (and even some without) have variance matrices of this form, as do GMM estimators. The following lemma is a useful sufficient condition for showing that one estimator is more efficient than another.

LEMMA 14.1 (Relative Efficiency): Let $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ be two $\sqrt{N}$-asymptotically normal estimators of the $P \times 1$ parameter vector $\boldsymbol{\theta}_o$, with asymptotic variances of the form (14.48) (with appropriate subscripts on $\mathbf{A}$, $\mathbf{s}$, and $\mathbf{V}$). If for some $\rho > 0$,

$$\mathrm{E}[\mathbf{s}_1(\mathbf{w})\mathbf{s}_1(\mathbf{w})'] = \rho\mathbf{A}_1, \tag{14.49}$$

$$\mathrm{E}[\mathbf{s}_2(\mathbf{w})\mathbf{s}_1(\mathbf{w})'] = \rho\mathbf{A}_2, \tag{14.50}$$

then $\mathbf{V}_2 - \mathbf{V}_1$ is p.s.d.

The proof of Lemma 14.1 is given in the chapter appendix.

Condition (14.49) is essentially the generalized information matrix equality (GIME) we introduced in Section 12.5.1 for the estimator $\hat{\boldsymbol{\theta}}_1$. Notice that $\mathbf{A}_1$ is necessarily symmetric and positive definite under condition (14.49). Condition (14.50) is new. In most cases, it says that the expected outer product of the scores $\mathbf{s}_2$ and $\mathbf{s}_1$ equals the expected Jacobian of $\mathbf{s}_2$ (evaluated at $\boldsymbol{\theta}_o$). In Section 12.5.1 we claimed that the GIME plays a role in efficiency, and Lemma 14.1 shows that it does so.

Verifying the conditions of Lemma 14.1 is also very convenient for constructing simple forms of the Hausman (1978) statistic in a variety of contexts. Provided that the two estimators are jointly asymptotically normally distributed—something that is almost always true when each is $\sqrt{N}$-asymptotically normal, and that can be verified by stacking the first-order representations of the estimators—assumptions (14.49) and (14.50) imply that the asymptotic covariance between $\sqrt{N}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_o)$ and $\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_o)$ is $\mathbf{A}_2^{-1}\mathrm{E}(\mathbf{s}_2\mathbf{s}_1')\mathbf{A}_1^{-1} = \mathbf{A}_2^{-1}(\rho\mathbf{A}_2)\mathbf{A}_1^{-1} = \rho\mathbf{A}_1^{-1} = \mathrm{Avar}[\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_o)]$. In other words, the asymptotic covariance between the ($\sqrt{N}$-scaled) estimators is equal to the asymptotic variance of the efficient estimator. This equality implies that $\mathrm{Avar}[\sqrt{N}(\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1)] = \mathbf{V}_2 + \mathbf{V}_1 - \mathbf{C} - \mathbf{C}' = \mathbf{V}_2 + \mathbf{V}_1 - 2\mathbf{V}_1 = \mathbf{V}_2 - \mathbf{V}_1$, where $\mathbf{C}$ is the asymptotic covariance. If $\mathbf{V}_2 - \mathbf{V}_1$ is actually positive definite (rather than just p.s.d.), then $[\sqrt{N}(\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1)]'(\hat{\mathbf{V}}_2 - \hat{\mathbf{V}}_1)^{-1}[\sqrt{N}(\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1)] \overset{a}{\sim} \chi^2_P$ under the assumptions of Lemma 14.1, where $\hat{\mathbf{V}}_g$ is a consistent estimator of $\mathbf{V}_g$, $g = 1, 2$. Statistically significant differences between $\hat{\boldsymbol{\theta}}_2$ and $\hat{\boldsymbol{\theta}}_1$ signal some sort of model misspecification. (See Section 6.2.1, where we discussed this form of the Hausman test for comparing are easily obtained using the panel data bootstrap. See Papke and Wooldridge (2008) for more discussion.

Problems

18.1. a. For estimating the mean of a nonnegative random variable y, the Poisson quasi-log likelihood for a random draw is

$$\ell_i(\mu) = y_i \log(\mu) - \mu, \qquad \mu > 0$$

(where terms not depending on μ have been dropped). Letting $\mu_o \equiv \mathrm{E}(y_i)$, we have $\mathrm{E}[\ell_i(\mu)] = \mu_o \log(\mu) - \mu$. Show that this function is uniquely maximized at $\mu = \mu_o$. This simple result is the basis for the consistency of the Poisson QMLE in the general case.

b. The gamma (exponential) quasi-log likelihood is

$$\ell_i(\mu) = -y_i/\mu - \log(\mu), \qquad \mu > 0$$

Show that $\mathrm{E}[\ell_i(\mu)]$ is uniquely maximized at $\mu = \mu_o$.

18.2. Carefully write out the robust variance matrix estimator (18.14) when $m(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta})$.

18.3. Use the data in SMOKE.RAW to answer this question.

a. Use a linear regression model to explain *cigs*, the number of cigarettes smoked per day. Use as explanatory variables log(*cigpric*), log(*income*), *restaurn*, *white*, *educ*, *age*, and age^2. Are the price and income variables significant? Does using heteroskedasticity-robust standard errors change your conclusions?

b. Now estimate a Poisson regression model for *cigs*, with an exponential conditional mean and the same explanatory variables as in part a. Using the usual MLE standard errors, are the price and income variables each significant at the 5 percent level? Interpret their coefficients.

c. Find $\hat{\sigma}$. Is there evidence of overdispersion? Using the GLM standard errors, discuss the significance of log(*cigpric*) and log(*income*).

d. Compare the usual MLE *LR* statistic for joint significance of log(*cigpric*) and log(*income*) with the *QLR* statistic in equation (18.17).

e. Compute the fully robust standard errors, and compare these with the GLM standard errors.

Table 20.1
Salary-Benefits Trade-off for Michigan Teachers

Dependent Variable	$\log(avgsal)$		
	(1)	(2)	(3)
Estimation Method	Pooled OLS	Random Effects	Fixed Effects
Explanatory Variable			
bs	−0.177 (0.122) [0.260]	−0.381 (0.112) [0.150]	−0.495 (0.133) [0.194]
$\log(staff)$	−0.691 (0.018) [0.035]	−0.617 (0.015) [0.036]	−0.622 (0.017) [0.043]
$\log(enroll)$	−0.0292 (0.0085) [0.0257]	−0.0249 (0.0076) [0.0115]	−0.0515 (0.0094) [0.0131]
lunch	−0.00085 (0.00016) [0.00057]	0.00030 (0.00018) [0.00020]	0.00051 (0.00021) [0.00021]
constant	13.724 (0.112) [0.256]	13.367 (0.098) [0.197]	13.618 (0.113) [0.241]
Number of districts	537	537	537
Number of schools	1,848	1,848	1,848

Quantities in parentheses are the nonrobust standard errors; those in brackets are robust to arbitrary within-district correlation as well as heteroskedasticity.
The intercept reported for fixed effects is the average of the estimated district effects.
The fully robust regression based Hausman test, with four degrees-of-freedom in the chi-square distribution, yields $H = 20.70$ and p-value $= 0.0004$.

percentage of students eligible for the federal free or reduced-price lunch program. Using the approximation $\log(1 + x) \approx x$ for "small" x, it can be shown that a dollar-for-dollar trade-off in salary and benefits is the same as $\beta_1 = -1$.

We estimate the equation using three methods: pooled OLS, random effects, and fixed effects. The results are given in Table 20.1. The table contains the nonrobust standard errors for each method—that is, the standard errors computed under the "ideal" set of assumptions for the particular estimator—along with the standard errors that are robust to arbitrary within-district correlation and heteroskedasticity.

The POLS estimates provide little evidence of a trade-off between salary and benefits. The coefficient is negative, but its value, -0.177, is pretty small, and not close to -1 (the hypothesized value for a one-for-one trade-off between salary and benefits). Its fully robust t statistic is less than 0.7 in magnitude. Notice that the robust standard error, which properly accounts for the cluster nature of the data, is more than twice as large as the nonrobust one.