This is to certify that the

dissertation entitled

PANEL DATA WITH CROSS-SECTION VARIATION IN THE SLOPES AS WELL AS THE INTERCEPT: THE EFFECTS OF UNIONS ON WAGES

presented by

CHRISTOPHER MARK CORNWELL

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Economics

Peter Schmidt
Major professor

Date November 13, 1985

PANEL DATA WITH CROSS-SECTION VARIATION IN THE SLOPES AS WELL AS THE INTERCEPT: THE EFFECTS OF UNIONS ON WAGES

by

Christopher Mark Cornwell

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1985

ABSTRACT

Combining time-series and cross-section data is useful in controlling for omitted or unobservable individual specific attributes which may be correlated with the explanatory variables in a regression. A regression function that does not condition on the individual specific effects will not identify the parameters of the model. Econometric models that assume the availability of panel data are usually of the constant slopes and variable intercept form. This study considers a panel data model with cross-sectional variation in some of the slopes as well as the intercept.

An established literature exists on the estimation of the simple model. The choice of estimation procedures depends on the assumptions about the individual effects. We distinguish three sets of assumptions: (1) fixed effects, (2) random effects uncorrelated with the regressors, and (3) random effects correlated with the regressors. The fixed effects model is estimated by analysis of covariance, or within. Generalized least squares is the standard procedure when the effects are random and uncorrelated with the regressors. When the effects are random and correlated with the regressors, the instrumental variables estimator introduced by Hausman and Taylor is appropriate. Each of these estimators is asymptotically well-behaved in the case of many individuals and few time periods. For the general model we derive the analogous within, GLS, and Hausman-Taylor instrumental variables

estimators. Furthermore, we prove that these estimators possess the same properties in the general model that they have in the simple model.

Then, we apply some of our theoretical results to an attempt to measure the impact of unions on wages. Conventional wisdom suggests that cross-section estimates are upwardly biased due to the positive correlation of unobserved individual specific attributes, or "ability", with union status. Most often this bias is addressed through a fixed effects specification of the simple model. However, this approach is criticized for ignoring the sectoral dependence of the individual effects. We consider a special case of our model in an attempt to deal with this criticism. We conclude the conventional wisdom is confirmed in our empirical investigation.

ACKNOWLEGMENTS

The completion of this dissertation leaves me indebted to many people. To a few people the debt I owe is particularly large.

My dissertation chairman, Peter Schmidt, has consistently provided careful guidance and patient instruction. From him I have learned most of what I know about econometrics. The debt I owe to him is certainly beyond my capacity to repay.

Dan Hamermesh has similarly provided often needed guidance and direction. My interest in labor economics is due, in large part, to him.

I also owe much to Harry Holzer. I have benefitted greatly from his comments and advice.

Perhaps the greatest debt is that which I owe to my wife. Without her love, encouragement, and support this research would still be incomplete.

In addition, I am very grateful for the constant support of my parents and the excellent typing provided by Terie Snyder.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Panel or longitudinal data are simply time-series observations on a cross-section. The applied economist may have several years of data on individuals, households, or firms. Typically the number of time observations (T) is small and the number of cross-sectional units (N) is large. Combining time-series and cross-section data is particularly useful in controlling for omitted or unobservable attributes specific to the cross-sectional unit (henceforth taken to be an individual) which are correlated with the explanatory variables. A regression function which does not condition on these individual specific effects will not identify the parameters of the model.

Econometric models that assume the availability of panel data most often take the form

$$(1.1) \qquad Y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it} , \quad i=1,\ldots,N, \; t=1,\ldots,T ,$$

where X_{it} is a vector of explanatory variables and ε_{it} is an iid error. The α_i are individual specific parameters, or effects. So, in (1.1) each individual has a unique intercept.

How we estimate (1.1) depends on our assumptions about the α_i. From the literature we can identify three distinct cases: (1) fixed effects, (2) random effects uncorrelated with the regressors, and (3) random effects correlated with the regressors. Case (1) is the weakest set of assumptions. Here the individual effects are taken to be constant over time (no other assumptions about them are necessary). In (2) the α_i are assumed to be iid random variables that are uncorrelated with all of X. This specification is sometimes referred to as the error components model. The last case drops the independence assumption and allows the individual effects to be correlated with some of X.

This study has as its focus the estimation of an obvious generalization of (1.1) - a panel data model which allows cross-sectional variation in some of the slopes as well as the intercept. Such a model can be written as

$$(1.2) \qquad Y_{it} = X_{it}'\beta + W_{it}'\delta_i + \varepsilon_{it}, \quad i=1,\ldots,N, \quad t=1,\ldots,T \ ,$$

where W_{it} is a vector of explanatory variables associated with coefficients that depend on i. (Alternatively, we could partition X and β and write $Y_{it} = X_{1it}'\beta_1 + X_{2it}'\beta_{2i} + \varepsilon_{it}$). Clearly, if W_{it} = constant then (1.2) reduces to the simple, intercept varying model.

Our investigation proceeds as follows. The next three chapters present theoretical results. Chapter Two considers the fixed effects case; Chapter Three takes up the case of random effects uncorrelated with the regressors; then, the case of random effects uncorrelated with the regressors is covered in Chapter Four. For each set of assumptions, we first review the results on estimation established for the simple model. Then, we extend these results to the general model. In each case we are interested in estimators which have good asymptotic properties as $N\to\infty$ while T is fixed.

In Chapter Two this means deriving an analog to the within (or analysis of covariance) estimator of the simple model. We also show that under normality, the within estimator for the general model is the conditional MLE. The error components model is traditionally estimated by generalized least squares. So, in Chapter Three, we derive the GLS estimator for our model and prove that the properties of GLS in the simple model carry over to the general model. The groundwork for Chapter Four is laid by Hausman and Taylor (1981) (hereafter referred to as H-T). They develop an instrumental variables procedure for the simple model in which the individual effects are random and correlated with some of the regressors. We derive a similar instrumental variables estimator for our model, and following H-T, detail conditions under which it differs from the fixed effects estimator.

In Chapter Five, we apply some of our theoretical results in an empirical exercise where we attempt to measure the impact of unions on earnings. Using data from the years 1978-1981 of the Michigan Panel Study of Income Dynamics (PSID), we estimate: (1) the simple cross-sectional earnings equations for the four years of our sample; (2) the usual panel data model in which only the intercept varies across individuals; and (3) a special case of our general panel data model. Conventional wisdom states that the cross-section estimates are upwardly biased due to the positive correlation of unobserved individual specific attributes - collectively referred to as "ability" - with union status. Most often this bias is addressed through a fixed effects specification of the simple model. We note some criticisms of this approach and examine a special case of the fixed effects version of our model that attempts to deal with the criticisms. For the sake of comparison, we also estimate (2) and (3) under both sets of random effects assumptions. In general, we are able to confirm the conventional wisdom that cross-section estimates of the union wage effect are upwardly biased.

In Chapter Six, we present a summary of our results and offer some final remarks on panel data models in which some slopes as well as the intercept vary cross-sectionally.

CHAPTER TWO

FIXED EFFECTS

2.1 Introduction

In this Chapter, we consider the estimation of (1-1) and (1-2) under the weakest set of assumptions; i.e., fixed effects. Since the number of individual specific parameters increase with sample size, we focus our analysis on the estimation of β. In particular, we seek estimators of β that are consistent in the common panel case of large N and small T.

First, we review the estimation of the simple model. We derive the "within" estimator of covariance analysis, which possesses the above consistency property. This is equivalent to maximum likelihood. Chamberlain (1980) demonstrates the incidental parameters problem can also be circumvented through a conditional likelihood approach. He derives the conditional MLE of β in (1.1) which is also equivalent to the within estimator.

Secondly, we extend the results of the standard model to the more general model which allows cross-sectional variation in some of the slopes as well as the intercept. There exists a substantial literature on the case in which all coefficients vary across i (see, for example, Judge et al (1985, section 13.5). In this case, the model can be considered as N seemingly unrelated regressions; and if not all the coefficients vary across i, then cross-equation restrictions are implied. Since we are concerned with an asymptotic theory in which $N\to\infty$ and T is fixed, this treatment is unsatisfactory. Mundlak (1978) has investigated this case, and notes (given standard assumptions about the errors) the cross-sectionally constant regression coefficients can be estimated by a version of least squares.

So, we follow Chamberlain, and derive the conditional MLE of β in (1.1). This is shown to be equivalent to the obvious least squares estimator, which is a comforting result. Our conclusions are summarized in section four.

2.2 The Standard Model

Recall the usual representation of a linear regression model with panel data. This is described in (1.1) as

$$Y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it} \qquad i=1,\ldots,N;\; t=1,\ldots,T.$$

where ε_{it} is assumed to be iid $N(0,\sigma^2)$. The α_i are incidental parameters and β is a κ-dimensional vector of cross-sectionally constant coefficients.

As outlined in the previous section, we seek an estimator of β which is consistent for the usual panel case of large N and small T. It is well known that the "within" estimator of covariance analysis possesses this consistency property. Let us review this estimation procedure.

As a matter of notation, define

$$(2.2.1) \qquad Y_i = \begin{vmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{vmatrix} , \quad X_i = \begin{vmatrix} X_{i1}' \\ X_{i2}' \\ \vdots \\ X_{iT}' \end{vmatrix} , \quad \varepsilon_i = \begin{vmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{vmatrix} .$$

This allows us to write

$$(2.2.2) \qquad Y_i = X_i\beta + \alpha_i + \varepsilon_i .$$

Then, we may consider all NT observations as

$$(2.2.3) \qquad Y = X\beta + D\alpha_* + \varepsilon ,$$

where

$$(2.2.4) \qquad Y = \begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{vmatrix} , \quad X = \begin{vmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{vmatrix} , \quad \varepsilon = \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{vmatrix}$$

and

$$(2.2.5) \qquad D = I_N \otimes e_T = \begin{vmatrix} e_T & 0 & \cdots & 0 \\ 0 & e_T & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & e_T \end{vmatrix} , \qquad \alpha_* = \begin{vmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{vmatrix} ,$$

where e_T is a T-dimensional vector of ones. Notice D is simply a matrix of individual indicator (dummy) variables.

The within estimator of β is derived as follows. Let

$$(2.2.6) \qquad M_D = I_{NT} - D(D'D)^{-1}D' .$$

Then transform the data by premultiplying (2.2.3) by M_D, thereby obtaining

$$M_D Y = M_D X\beta + M_D D\alpha_* + M_D \varepsilon ,$$

which reduces to

$$(2.2.7) \qquad M_D Y = M_D X\beta + M_D \varepsilon ,$$

since $M_D D=0$. This transformation changes a vector of observations into deviations from individual means.[1] Least squares applied to (2.2.7) yields the within estimator of β, defined as

$$(2.2.8) \qquad \hat{\beta}_W = (X'M_D X)^{-1}X'M_D Y .$$

Under the condition that X_{it} varies over time, $\hat{\beta}_W$ is consistent as $N\to\infty$ for fixed T.

It should be clear that the within estimation does not depend on normality of the errors. When we invoke normality, we see that maximum-likelihood is, in fact, analysis of covariance. Chamberlain (1980) derives the conditional MLE of β, which he shows is also equivalent to $\hat{\beta}_W$. The conditional likelihood approach employs a set of sufficient statistics for the α_i, removing any incidental parameters problem. The consistency of $\hat{\beta}_W$ is confirmed by the coincidence of the conditional and joint MLE's.

2.3 A Generalization

As described in (1.2), a straightforward generalization of the standard panel data model is

$$Y_{it} = X_{it}'\beta + W_{it}'\delta_i + \varepsilon_{it} \qquad i=j,\ldots,N;\ t=1,\ldots,T .$$

The W_{it} and δ_i are L-dimensional vectors of explanatory variables and coefficients, respectively. The remaining variables and parameters are defined as in the simple model.

This distinguishing feature of this model is that we allow for cross-sectional variation in some of the slopes as well as the intercept. Obviously, if W_{it} is a constant, (1.2) reduces to the simple model. Again, we seek a consistent (as $N\to\infty$ for fixed T) estimator of β.

Let $W_i = (W_{i1}, W_{i2},\ldots, W_{it})'$. This allows us to write

$$(2.3.1) \qquad Y_i = X_i\beta + W_i\delta_i + \varepsilon_i \ .$$

Then, considering all NT observations, we obtain

$$(2.3.2) \qquad Y = X\beta + Q\,\delta_* + \varepsilon \ ,$$

where

$$(2.3.3) \qquad Q = \begin{vmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_N \end{vmatrix} \ , \quad \delta_* = \begin{vmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{vmatrix} \ .$$

This general model can be estimated by least squares. By analogy to the within transformation, we premultiply (2.3.2) by the idempotent matrix M_Q, which is defined as

$$(2.3.4) \qquad \begin{aligned} M_Q &= I_{NT} - Q(Q'Q)^{-1}Q' \\ &= \begin{vmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_N \end{vmatrix} \end{aligned}$$

with

$$(2.3.5) \qquad M_i = I_T - W_i(W_i'W_i)^{-1}W_i' \ .$$

Then, we may apply least squares to the transformed model,

$$(2.3.6)\qquad M_Q Y = M_Q X\beta + M_Q \varepsilon \ ,$$

which yields the following estimator of β:

$$(2.3.7)\qquad \tilde{\beta}_W = (X'M_Q X)^{-1} X'M_Q Y = (\sum_i X_i'M_i X_i)^{-1} \sum_i X_i'M_i Y_i \ .$$

The estimator $\tilde{\beta}_W$ is consistent for fixed T if $(X'M_Q X)^{-1} \to 0$ as $N \to \infty$. Essentially, this is a condition which requires sufficient temporal variation in X_i not explained by W_i. When W_i is only a constant term, we have the familiar condition that X_{it} must vary over time.

As in the standard model, the above is straightforward and does not depend on normality. We now invoke normality to prove that $\tilde{\beta}_W$ is also the conditional MLE of β.

Following Chamberlain, consider the δ_i as incidental parameters for which we need to find sufficient statistics. The likelihood of Y, conditional on the sufficient statistics, will not depend on the δ_i. Maximizing this conditional likelihood should provide a consistent estimator of β.

To prove this, we first show that $W_i'Y_i$ is sufficient for δ_i. Consider the (T+L)X1 vector

$$(2.3.8)\qquad y = \begin{vmatrix} y_1 \\ \\ y_2 \end{vmatrix} \equiv \begin{vmatrix} Y_i \\ \\ W_i'Y_i \end{vmatrix} \ .$$

The vector y has a (singular) multivariate normal distribution with mean μ and covariance matrix Σ, which with the above partitioning gives

$$
\begin{aligned}
\mu_1 &= X_i\beta + W_i\delta_i \\
\mu_2 &= W_i'X_i\beta + W_i'W_i\delta_i \\
\Sigma_{11} &= \sigma^2 I_T \\
\Sigma_{22} &= \sigma^2 W_i'W_i \\
\Sigma_{12} &= \sigma^2 W_i \\
\Sigma_{21} &= \sigma^2 W_i'
\end{aligned} \tag{2.3.9}
$$

Standard results on normal distributions imply the distribution of Y_i conditional on $W_i'Y_i$ is (singular) normal with mean

$$
\begin{aligned}
E(Y_i \; W_i'Y_i) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2) \\
&= W_i(W_i'W_i)^{-1}W_i'Y_i - M_iX_i\beta \;,
\end{aligned} \tag{2.3.10}
$$

and covariance matrix

$$
\Omega_i = \mathrm{COV}(Y_i \; W_i'Y_i) = \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \sigma^2 M_i \;. \tag{2.3.11}
$$

Neither (2.3.10) nor (2.3.11) depends on δ_i. Hence, $W_i'Y_i$ is indeed sufficient for δ_i.

To construct the conditional likelihood function, we need an explicit formula for the conditional density. Here we employ the standard result [see, e.g., Rao (1973, p. 528)] that if $y \sim N(\mu,\Sigma)$ with dim(Y)=P and rank (Σ) = k<P, then

$$(2.3.12) \quad f(y) = (2\pi)^{-k/2}\xi^{-1/2}\exp\left[-1/2(y-\mu)'\Sigma^{+}(y-\mu)\right] ,$$

where Σ^{+} is any generalized inverse of Σ, and ξ is the product of the k positive eigenvalues of Σ. Since M_i is idempotent (with rank T-L),

$$(2.3.13) \quad \Omega_i^{+} = \frac{1}{\sigma^2} M_i$$

and

$$(2.3.14) \quad \xi = (\sigma^2)^{(T-L)} .$$

Given the conditional mean of the distribution,

$$(2.3.15) \quad Y_i - E(Y_i \; W_i'Y_i) = M_i(Y_i - X_i\beta) .$$

Therefore, we obtain the following expression for the conditional density:

$$(2.3.16) \quad f(Y_i/W_i'Y_i) = (2\pi)^{-(T-L)/2}\sigma^{-(T-L)}\exp\left[-\frac{1}{2\sigma^2}(Y_i-X_i\beta)'M_i(Y_i-X_i\beta)\right] .$$

Since observations are assumed independent across i, we may multiply over i to obtain the conditional likelihood function,

$$(2.3.17) \quad \mathcal{L} = (2\pi)^{-N(T-L)/2}\sigma^{-N(T-L)}\exp\left[-\frac{1}{2\sigma^2}\sum_i (Y_i-X_i\beta)'M_i(Y_i-X_i\beta)\right].$$

This likelihood function is indeed maximized by $\tilde{\beta}_W$ as given in (2.3.7).

2. Conclusion

We have considered an extension of the usual fixed effects panel data model in which some of the explanatory variables may have coefficients which vary across i.

The results we obtain are essentially the same as those obtained in the standard, simpler case. Our model may be estimated by least squares, and the resulting estimates (of the cross-sectionally constant coefficients) are consistent given a reasonable condition on the variability of the regressors. Under normality this estimator is in fact the conditional MLE. However, we cannot claim asymptotic efficiency. Unlike the direct MLE, the conditional MLE will not, in general, be efficient in the sense that its asymptotic variance equals the Cramer-Rao lower bound (see Andersen (1970)). On the other hand, we do not know of any estimators with superior asymptotic behavior.

CHAPTER THREE

RANDOM EFFECTS

UNCORRELATED WITH THE REGRESSORS

3.1 Introduction

Having demonstrated the consistency of the within estimator in the usual (fixed T) panel case, we must now acknowledge two drawbacks of the fixed effects specification. First, for small T, the within estimator is not fully efficient since it ignores variation between individuals. Secondly, time-invariant explanatory variables are orthogonal to the within transformation and therefore cannot be incorporated into a fixed effects model. This is potentially a serious problem, since in many applications, attention is focused on the coefficients of such variables (e.g. on the coefficients of race or education in an earnings equation).

As a remedy to the problems of fixed-effects models, a random effects specification is sometimes proposed. Random effects models take the individual effects to be iid random variables independent of the explanatory variables and the disturbance. Estimation of the simple model, also referred to as the error components model, is well documented. The basic results are reviewed in the next section.

Then, in section three, we extend the results of the error components model to our more general model where some of the slope coefficients are allowed to vary cross-sectionally. Our model, under the assumption of random effects, is essentially a Swamy random coefficient model where some of the coefficients do not vary across i. Section four summarizes our results.

3.2 Error Components

Assume the α_i in (1.2) are iid $N(0,\sigma_\alpha^2)$ variables. Furthermore, let α_i be uncorrelated with the columns of (X,ε). Under these assumptions, the usual panel data model has an error components structure, where

$$(3.2.1) \qquad v_{it} = \alpha_i + \varepsilon_{it} ,$$

and therefore

$$(3.2.2) \qquad Y_{it} = X_{it}'\beta + v_{it} .$$

Then, considering all NT observations,

$$(3.2.3) \qquad Y = X\beta + v .$$

with $v = D\alpha_* + \varepsilon$.

Traditionally, models like (3.2.3) are estimated by generalized least squares. The GLS estimator of β is defined as

$$(3.2.4) \quad \hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y ,$$

where

$$(3.2.5) \quad \Omega = COV(v) = \sigma^2 I_{NT} + \sigma_\alpha^{\ 2} DD'$$

and

$$(3.2.6) \quad \Omega^{-1} = \frac{1}{\sigma}(I_{NT}-\theta_1^{\ 2}DD') , \quad \theta_1^{\ 2} = \frac{\sigma_\alpha^{\ 2}}{\sigma^2+T\sigma_\alpha^2} .$$

Now, like $\hat{\beta}_W$, $\hat{\beta}_{GLS}$ is consistent as $N\to\infty$ for fixed T. But, $\hat{\beta}_{GLS}$ is asymptotically more efficient than $\hat{\beta}_W$. This efficiency gain is a result of the exploitation by GLS of both within and "between" (across individuals) variation. Within estimation only uses variation within i. However, this efficiency gain disappears and $\hat{\beta}_{GLS} \to \hat{\beta}_W$ as $T\to\infty$.[1]

An additional advantage of the random effects specification is that time-invariant explanatory variables can be incorporated into the model (Recall the within transformation annihilates time-invariant regressors). So, in the classical panel case (large N, small T), the error components model may be preferred.

One caveat is in order. The consistency of $\hat{\beta}_{GLS}$ depends crucially on the assumption that the α_i are uncorrelated with the columns of X. This is often an unreasonable assumption; and, if violated, $\hat{\beta}_{GLS}$ is no longer consistent.[2] Ironically, it is inconsistency in the presence of such correlation that led to the original fixed effects model.

3.3 A Generalization

Generalizing the error components model to include cross-sectionally varying slope coefficients is straightforward. Let

$$(3.3.1) \qquad \delta_i = \delta_0 + u_i \ .$$

Assume the u_i are iid $N(0,\Delta)$ random variables, with $\Delta \equiv \mathrm{COV}(u_i)$. As in the previous case, take the u_i to be uncorrelated with the explanatory variables and ε. Given (3.3.1), the full model may be defined as

$$(3.3.2) \qquad Y_{it} = X_{it}'\beta + W_{it}'\delta_0 + v_{it} \ ,$$

where

$$(3.3.3) \qquad v_{it} = W_{it}'u_i + \varepsilon_{it} \ .$$

More conveniently, we have

$$(3.3.4) \qquad Y = X\beta + W\delta_0 + v$$

and

$$(3.3.5) \qquad v = Qu_* + \varepsilon \ ,$$

where

$$W = \begin{vmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{vmatrix} \quad \text{and} \quad u_* = \begin{vmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{vmatrix} .$$

Applying GLS to (3.3.4) we obtain the following estimator of β:

$$(3.3.6) \quad \tilde{\beta}_{GLS} = (X'\Omega^{-1/2} M_{W*}\Omega^{-1/2}X)^{-1} M_{W*}\Omega^{-1/2}Y ,$$

where

$$(3.3.7) \quad M_{W*} = I - W*(W*'W*)^{-1}W*' = I - \Omega^{-1/2}W[(\Omega^{-1/2}W)'\Omega^{-1/2}W]^{-1}W'\Omega^{-1/2},$$

and

$$(3.3.8) \quad \Omega = COV(v) = \sigma^2 I_{NT} + Q\Lambda Q' ,$$

with $\Lambda = I_N \otimes \Delta$. Note that

$$(3.3.9) \quad \Omega^{-1} = \frac{1}{\sigma^2} M_Q + Q(Q'Q)^{-1}\Gamma^{-1}(Q'Q)^{-1}Q' ,$$

where $\Gamma = \sigma^2(Q'Q)^{-1} + \Lambda$. (For the derivation of Ω^{-1}, see the appendix which concludes this chapter).

Now, as in the usual error components model, $\tilde{\beta}_{GLS}$ is consistent as $N\to\infty$ for fixed T. Then, do the efficiency results of GLS carry through to this more general model? Indeed, we would be surprised if they did not. It is straightforward to prove that $\tilde{\beta}_{GLS}$ is efficient relative to $\tilde{\beta}_W$ for fixed T and that $\tilde{\beta}_{GLS} \rightarrow \tilde{\beta}_W$ as $T \rightarrow \infty$.

If $\tilde{\beta}_{GLS}$ is efficient relative to $\tilde{\beta}_W$, then $[COV(\tilde{\beta}_W) - COV(\tilde{\beta}_{GLS})]$ must be a positive semidefinite (PSD) matrix. It is well known that for any two nonsingular matrices A and B, (A-B) is PSD if and only if $(B^{-1} - A^{-1})$ is PSD. Therefore, our problem is to show that

$$(3.3.10) \quad X'\Omega^{-1/2} M_{W*} \Omega^{-1/2} X - \frac{1}{\sigma^2} X' M_Q X$$

is PSD.

Rewriting (3.3.10), we obtain

$$(3.3.11) \quad \begin{aligned} & X'\Omega^{-1/2}[I - \Omega^{-1/2} W(W'\Omega^{-1}W)^{-1} W'\Omega^{-1/2}]\, \Omega^{-1/2} X - \frac{1}{\sigma^2} X' M_Q X \\ & = X'\Omega^{-1}X - X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}X - \frac{1}{\sigma^2} X' M_Q X \,. \end{aligned}$$

Now, given (3.3.9),

$$(3.3.12) \quad X'\Omega^{-1}X = \frac{1}{\sigma^2} X' M_Q X + X'Q(Q'Q)^{-1}\Gamma^{-1}(Q'Q)^{-1}Q'X$$

and

$$(3.3.13) \quad \begin{aligned} & X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}X \\ & = X'Q(Q'Q)^{-1}\Gamma^{-1}E_{NL}(E_{NL}'\Gamma^{-1}E_{NL})^{-1}E_{NL}'\Gamma^{-1}(Q'Q)^{-1}Q'X \,, \end{aligned}$$

where $E_{NL} = e_N \otimes I_L$, e_N a N-dimensional vector of one's. So we need only show that

$$(3.3.14)\qquad X'Q(Q'Q)^{-1}\Gamma^{-1}(Q'Q)^{-1}Q'X - X'Q(Q'Q)^{-1}\Gamma^{-1}E_{NL}(E_{NL}'\Gamma^{-1}E_{NL})^{-1}E_{NL}'\Gamma^{-1}(Q'Q)^{-1}Q'X$$

is PSD.

For simplicity define

$$(3.3.15)\qquad \begin{aligned} S &= X'Q(Q'Q)^{-1} \\ R &= \Gamma^{-1} \\ \bar{S} &= SR\ E_{NL}(E_{NL}RE_{NL})^{-1}E_{NL}'. \end{aligned}$$

Note that R is PD. Then, consider the PSD matrix,

$$(3.3.16)\qquad (S-\bar{S})\ R\ (S-\bar{S})'\ .$$

Expanding (3.3.16), we obtain

$$\begin{aligned} SRS' + \bar{S}R\bar{S}' - SR\bar{S}'-\bar{S}RS' \\ = SRS + SRE_{NL}(E_{NL}'RE_{NL})^{-1}E_{NL}'RE_{NL}(E_{NL}'RE_{NL})E_{NL}'(SR)' \\ - SRE_{NL}(E_{NL}'RE_{NL})'E_{NL}'(SR)' \\ - SRE_{NL}(E_{NL}'RE_{NL})^{-1}E_{NL}(SR)'\ , \end{aligned}$$

which is equivalent to (3.3.14). Hence (3.3.10) is PSD, and $\tilde{\beta}_{GLS}$ is efficient relative to $\tilde{\beta}_{GLS}$ for fixed T.

However, as $T\to\infty$ this efficiency gain disappears. That is to say, as $T\to\infty$,

(3.3.17) $$\frac{1}{T}\,[X'\Omega^{-1/2}M_{W*}\Omega^{-1/2}X - \frac{1}{\sigma^2}X'M_QX]\to 0 ,$$

or

(3.3.18) $$\begin{aligned}&\frac{1}{T}\,[X'Q(Q'Q)^{-1}\Gamma^{-1}(Q'Q)^{-1}Q'X\\ &\qquad - X'Q(Q'Q)^{-1}\Gamma^{-1}E_{NL}(E_{NL}'\Gamma^{-1}E_{NL})^{-1}E_{NL}'\Gamma^{-1}(Q'Q)^{-1}Q'X]\to 0 .\end{aligned}$$

To see this, assume $\frac{1}{T}Q'Q$, $\frac{1}{T}X'Q$, and $\frac{1}{T}Q'X$ have finite nonsingular limits.

In particular, let

(3.3.19) $$\frac{1}{T}Q'Q\to U ,\ \frac{1}{T}Q'X\to V, \text{ and } \frac{1}{T}X'Q\to V'$$

Note that $(Q'Q)^{-1}\to 0$. This implies $\Gamma = \sigma^2(Q'Q)^{-1}+\Lambda\to\Lambda$ and $\Gamma^{-1}\to\Lambda^{-1}$. Then,

(3.3.20) $$\frac{1}{T}X'Q\,(\frac{1}{T}Q'Q)^{-1}\,\Gamma^{-1}(\frac{1}{T}Q'Q)^{-1}\,\frac{1}{T}Q'X\to V'U^{-1}\Lambda^{-1}U^{-1}V$$

and

$$\frac{1}{T} X'Q(\frac{1}{T} Q'Q)^{-1}\Gamma^{-1}E_{NL}(E_{NL}'\Gamma^{-1}E_{NL})^{-1}E_{NL}'\Gamma^{-1}(\frac{1}{T} Q'Q)^{-1} \frac{1}{T} Q'X$$

$$\rightarrow V'U^{-1}\Lambda^{-1}E_{WL}(N\Lambda^{-1})^{-1}E_{NL}'\Lambda^{-1}U^{-1}V..$$

Since $\frac{1}{T}\Lambda^{-1}\rightarrow 0$, (3.3.18) is confirmed, implying $\tilde{\beta}_{GLS}$ is equivalent to $\tilde{\beta}_W$ for large T.

Now, the caveat concerning the assumption of independence of the individual effects with the regressors also carries through to this model. It remains that GLS has little appeal if purchased from unreasonable assumptions. In the next Chapter, we consider a random effects model in which the independence assumption is dropped.

3.4 Summary

In this Chapter, we have considered the estimation of our general model in a random effects context. Taking the δ_i to be iid random variables independent of the regressors and the disturbance, we sought to extend the results of the usual error components model to this more general model.

As in the fixed effects case, the results we obtain for the general model are essentially the same as those established for the simpler model. Estimation is by GLS, and the GLS estimator of β is consistent for fixed T. Moreover, for fixed T, $\tilde{\beta}_{GLS}$ is efficient relative to $\tilde{\beta}_W$ (however, $\tilde{\beta}_{GLS}\rightarrow\tilde{\beta}_W$ as $T\rightarrow\infty$). And, the random effects specification does allow the inclusion of time-invariant explanatory variables.

However, as in the standard error components model, the unreasonableness of the independence assumption reduces the appeal of the random effects specification.

CHAPTER THREE

APPENDIX

CHAPTER THREE

APPENDIX

We make use of the following fact:

<u>Fact:</u> Provided the relevant inverses exist,

$$(A+BDB')^{-1} = A^{-1}BEB'A^{-1}BEB'A^{-1}+A^{-1}BE(E+D)^{-1}EB'A^{-1},$$

where $E = (B'A^{-1}B)^{-1}$.

The covariance matrix of V is given by

$$\Omega = \sigma^2 I_{NT} + Q\Lambda Q' . \tag{A.1}$$

In applying the above fact,

let $A = \sigma^2 I$, $B=Q$, $D=\Lambda$, and $E = (B'A^{-1}B)^{-1} = \sigma^2 Q'Q$. Then,

$$\begin{aligned} \Omega^{-1} &= \frac{1}{\sigma^2} I - \frac{1}{\sigma} Q(Q'Q)^{-1}Q' + Q(Q'Q)^{-1} [\sigma^2(Q'Q)^{-1}+\Lambda](Q'Q)^{-1}Q' \\ &= \frac{1}{\sigma^2} M_Q + Q(Q'Q)^{-1}\Gamma^{-1}(Q'Q)^{-1}Q' , \end{aligned} \tag{A.2}$$

where $\Gamma = \sigma^2(Q'Q)^{-1}+\Lambda$.

CHAPTER FOUR

RANDOM EFFECTS CORRELATED WITH THE REGRESSORS

The conventional random effects specification allows us to include time-invariant explanatory variables which cannot be incorporated into a fixed effects model. In the usual fixed T case, GLS estimation of this specification is more efficient than within estimation of the fixed effects model. However, these improvements usually come at the expense of an untenable assumption: that the individual effects are uncorrelated with all the regressors.

In this Chapter we drop this assumption. We investigate random effects panel data models in which the individual effects are assumed correlated with some of the explanatory variables. As in the previous two chapters, we focus on the fixed T case and begin by discussing the results, established by Hausman and Taylor (1981), for a model with only an intercept that varies across i. Briefly, they use prior information to construct exogeniety restrictions that are then employed to derive a consistent and asymptotically efficient instrumental variables estimator. They also derive the conditions under which it differs from the fixed effects estimator.

Next, we generalize the H-T analysis to include slopes that vary across i. Although the derivation of our estimator is more complicated, we show that the results of H-T carry through to the general model. In the last section we offer a brief summary.

4.2 The Hausman-Taylor Analysis

In addressing the problems associated with both the fixed effects and error components specifications, H-T consider the following model:

$$(4.2.1) \qquad Y_{it} = X_{it}'\beta + Z_i'\gamma + \alpha_i + \varepsilon_{it} ,$$

where Z_i is a J-dimensional vector of time-invariant explanatory variables (notice it is not indexed by t) and γ is a conformably dimensioned parameter vector. The α_i, as in the error components model, are assumed to be iid $N(0,\sigma_\alpha^2)$ random variables. However, unlike the usual random effects specification, H-T take the α_i to be correlated with some of the columns of X and Z.

According to H-T, consistent and asymptotically efficient estimation of all of the parameters in (4.2.1) hinges on our ability to distinguish columns of X and Z which are not correlated with the α_i.

To examine this, let us adopt a more convenient form of the model. Let

$$(4.2.2) \qquad Z = Z_* \otimes e_T , \qquad Z_* = \left| \begin{matrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{matrix} \right| .$$

So, we may write

$$(4.2.3) \qquad Y = X\beta + Z\gamma + V ,$$

where $V = D\alpha_* + \varepsilon$ (as before). Then, suppose we have prior information on which of the columns of X and Z are correlated with the α_i. Let

$$(4.2.4) \qquad X = [X_1, X_2] , \quad Z = [Z_1, Z_2] ,$$

where X_1 is $NTXk_1$, X_2 is $NTXk_2$, Z_1 is $NTXj_1$, and Z_2 is $NTXj_2$ (and $k_1+k_2=K$, $j_1+j_2=J$). For fixed T, assume

$$(4.2.5) \qquad \begin{array}{ll} \frac{1}{N} X_1' D\alpha_* \to 0, & \frac{1}{N} Z_1' D\alpha_* \to 0 \\ \frac{1}{N} X_2' D\alpha_* \to h_X \neq 0 & \frac{1}{N} Z_2' D\alpha_* \to h_Z \neq 0 \end{array} .$$

Now, it should be noted that although the condition $E(\alpha_i | X_{it}, Z_i) = 0$ fails, consistent, though inefficient estimates of β and γ may still be obtained from the within regression.[1] First, we estimate β by within, obtaining $\hat{\beta}_W$ defined in (2.2.8). Secondly, we compute the within residuals, $(Y-X\hat{\beta}_W)$. From the within residuals we estimate the individual means, defined as

$$(4.2.6) \qquad \hat{d} = P_D(Y-X\hat{\beta}_W) = Z\gamma + D\alpha_* + P_D(\varepsilon+\text{error}).$$

where $P_D = D(D'D)^{-1}D'$ and "error" denotes estimation error from the within regression. Treating $P_D(\varepsilon+\text{error})$ as an unobservable zero mean disturbance, we attempt to estimate γ from (4.2.6). We know OLS and GLS are inconsistent for γ since the α_i are not independent of Z_2. However, if the columns of X_1 (which are uncorrelated with $D\alpha_*$) provide sufficient instruments for the columns of Z_2 (which are correlated with $D\alpha_*$), consistent estimation of γ from (4.2.6) is possible. A necessary condition for this is that the model must include at least as many time-varying exogenous variables as time-invariant endogenous variables; i.e., it must be that $k_1 > j_2$.

If this condition is fulfilled, instrumental variables applied to (4.2.6), using as instruments

(412.7) $B = [X_1, Z_1]$,

yields the following estimator for γ (denoted $\hat{\gamma}_W$):

(4.2.8) $\hat{\gamma}_W = (Z'P_B Z)^{-1} Z'P_B \hat{d}$,

where $P_B = B(B'B)^{-1}B'$, the projection onto the column space of B. This estimator is consistent for fixed T, but not fully efficient since it is calculated from the within-residuals. (Recall $\hat{\beta}_W$ is not fully efficient since it ignores between variation.)

Now, consistent and asymptotically efficient estimates of β and γ can be derived if these parameters can be identified using prior information like that given in (4.2.5). Even without (4.2.5) all of the elements of β are identifiable as is clear from the within regression (i.e., $X'M_D X$ is nonsingular). However, without this information, no elements of γ are identifiable. But, given (4.2.5) we have the set of instruments.[2]

$$(4.2.9)\quad A = [M_D, X_1, Z_1]$$

and the corresponding projection P_A, suggesting the following proposition:

Proposition 1(H-T): A necessary and sufficient condition for the identification of (β,γ) in (4.2.3) is that

$$\begin{vmatrix} X_1' \\ Z_1' \end{vmatrix} \quad P_A\,[X_1, Z_1]$$

be nonsingular.

And, associated with this rank condition is the order condition to which we referred earlier:

Proposition 2 (H-T): A necessary condition for the identification of (β,γ) in (4.2.3) is that $k_1 > j_2$.

Suppose the parameters of (4.2.3) are identified by the information in (4.2.5).[3] Let

$$(4.2.10)\quad \Omega^{-1/2} = M_D + \theta P_D = I_{NT} - (1-\theta)P_D\ ,$$

where $\theta = (1 - T\theta_1{}^2)^{1/2} = [\sigma^2/(\sigma^2 + T\sigma_\alpha^2)]^{1/2}$.[4] Then, perform instrumental variables on the transformed equation,

$$(4.2.11)\quad \Omega^{-1/2}Y = \Omega^{-1/2}X\beta + \Omega^{-1/2}Z\gamma + \Omega^{-1/2}V\ ,$$

using the set of instruments given by A in (4.2.9). This procedure yields consistent and asymptotically efficient estimates of β and γ. Equivalently, and more computationally convenient,[5] we may apply OLS to

$$(4.2.12)\quad P_A\Omega^{-1/2}Y = P_A\Omega^{-1/2}X\beta + P_A\Omega^{-1/2}Z\gamma + P_A\Omega^{-1/2}V ,$$

where P_A is the projection onto the column space of A. We denote the estimators of (β,γ) obtained from (4.2.11) or (4.2.12) as $(\hat{\beta}^*,\hat{\gamma}^*)$.

Now, in evaluating the information given in (4.2.5), three cases are possible: under-identification, exact-identification, and over-identification. First, in the under-identified case $(k_1<j_2)$, $\hat{\beta}^*=\hat{\beta}_W$ and $\hat{\gamma}^*$ does not exist. Secondly, in the case of exact-identification $(k_1=j_2)$, $\hat{\beta}^*=\hat{\beta}_W$ and $\hat{\gamma}^*=\hat{\gamma}_W$, where $\hat{\gamma}_W$ is defined in (4.2.8). Finally, if the model is over-identified $(k_1>j_2)$, $(\hat{\beta}^*,\hat{\gamma}^*)\neq(\hat{\beta}_W,\hat{\gamma}_W)$ and $(\hat{\beta}^*,\hat{\gamma}^*)$ is more efficient than $(\hat{\beta}_W,\hat{\gamma}_W)$.[6]

4.3 H-T: An Extension

This section extends the H-T analysis to our more general model. In doing so, we are able to incorporate time-invariant explanatory variables and allow for endogeneity of some regressors based on correlation with the individual effects. The version of our model we consider here is

$$(4.3.1)\qquad Y_{it}=X_{it}'\beta + Z_{it}'\gamma + W_{it}'\delta_i + \varepsilon_{it} ,$$

where Z_{it} is a JX1 vector of explanatory variables. We treat the Z_{it} as time-invariant in the sense that any time variation in the Z_{it} must be fully explained by the time variation in the W_{it}.[7] Again, let $\delta_i=\delta_0+u_i$, where u_i are taken to be iid $N(0,\Delta)$ random variables. However, we now consider the case where the u_i are correlated with some of the regressors.

Consider all NT observations and write our model as

$$(4.3.2) \qquad Y = X\beta + Z\gamma + W\delta_o + v ,$$

where

$$(4.3.3) \qquad Z = \begin{vmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{vmatrix} , \quad Z_i = \begin{vmatrix} Z_{i1}' \\ Z_{i2}' \\ \vdots \\ Z_{iT}' \end{vmatrix}$$

and $V = QU_* + \varepsilon$. Suppose, for fixed T,

$$(4.3.4) \qquad \frac{1}{N} X_1' QU_* \to 0 \qquad \frac{1}{N} Z_1'QU_* \to 0 \qquad \frac{1}{N} W'QU_* \to 0$$

$$\frac{1}{N} X_2' QU_* \to g_X \neq 0 \qquad \frac{1}{N} Z_2' QU_* \to g_Z \neq 0 ,$$

where X and Z are partitioned in the same way described by (4.2.4). Note that we have assumed the effects to be uncorrelated with some of the columns of X and Z, and all of the columns of W.

Using prior information like (4.3.4), H-T derived two different estimation procedures for their model. One combined the prior information with the fixed effects regression to obtain consistent though inefficient estimates of the parameters. The other applied identifying restrictions to an $\Omega^{-1/2}$ transformation of the equation to obtain consistent and asymptotically efficient parameter estimates. Next, we construct analogous procedures for our model.

First, consider estimation based on the fixed effects regression. The translation of the initial steps of the H-T procedure to our model is direct and straightforward. We simply estimate β by applying OLS to (2.3.6), thereby obtaining $\tilde{\beta}_W$ in (2.3.7), and then form the least squares residuals, $(Y-X\tilde{\beta}_W)$. At this point the translation becomes more subtle. The next step in H-T is to calculate $\hat{d}$ in (4.2.6) by premultiplying the residuals by P_D. This would suggest the $\hat{d}$ analog in our model should be constructed by premultiplying our residuals by $P_Q = Q(Q'Q)^{-1}Q'$. But this is not optimal.

The correct procedure in our case is to calculate[8]

$$(4.3.5) \qquad \begin{aligned} \tilde{d} &= \Omega^{-1/2}(Y-X\tilde{\beta}_W) = \Omega^{-1/2}Z\gamma + \Omega^{-1/2}W\delta_0 \\ &\quad + \Omega^{-1/2}(QU_* + \epsilon + \text{error}) . \end{aligned}$$

where

$$(4.3.6) \qquad \Omega^{-1/2} = \frac{1}{\sigma} M_Q + F$$

and

$$(4.3.7) \quad F = Q(Q'Q)^{-1/2}[\sigma^2 I_{NL}+(Q'Q)^{1/2}\Lambda(Q'Q)^{1/2}]^{-1/2}(Q'Q)^{-1/2}Q'.$$

Then, if $k_1>j_2$,[9] we perform instrumental variables on (4.3.5), using the set of instruments[10]

$$(4.3.8) \quad B_* = \Omega^{-1/2}B = \Omega^{-1/2}(X_1,Z_1,W) .$$

This yields

$$(4.3.9) \quad \begin{vmatrix} \tilde{\gamma}_W \\ \\ \tilde{\delta}_{OW} \end{vmatrix} = [(Z_1W)'\Omega^{-1/2}P_{B_*}\Omega^{-1/2}(Z,W)]^{-1}(Z,W)^{-1}\Omega^{-1/2}P_{B_*}\tilde{d} ,$$

with $P_{B_*} = B_*(B_*'B_*)^{-1}B_*'$.

To understand this, recall the definition of the H-T $\Omega^{-1/2}$ in (4.2.10). Suppose we substitute (4.2.10) for P_D in the calculation of $\hat{d}$. The, instrumental variables using $B_* = \Omega^{-1/2}(X_1,Z_1)$ yields

$$(4.3.10) \quad \hat{\gamma}_W = (Z'\Omega^{-1/2}P_{B_*}\Omega^{-1/2}Z)^{-1}Z'\Omega^{-1/2}P_{B_*}\Omega^{-1/2}(Y-X\hat{\beta}_W) ,$$

which is generally different from $\hat{\gamma}_W$ in (4.2.8).[11] However, when $k_1=j_2$ (the exact identification case), both (4.2.8) and (4.3.10) are equivalent to[12]

$$(4.3.11) \quad \hat{\gamma}_W = (B'Z)^{-1} B'(Y-X\hat{\beta}_W) .$$

So, in the case in which this procedure is appropriate (in the sense it is equivalent to "efficient" estimation when $k_1=j_2$), it makes no difference in the H-T model whether we use $P_D(Y-X\hat{\beta}_W)$ or $\Omega^{-1/2}(Y-X\hat{\beta}_W)$ for $\hat{d}$, or whether we employ (X_1,Z_1) or $\Omega^{-1/2}(X_1,Z_1)$ as instruments. In our model, it is never the case that substitution of P_Q for $\Omega^{-1/2}$ and B for B_* results in estimates equivalent to the correct $\tilde{\gamma}_W$ and $\tilde{\delta}_{0W}$ given in (4.3.9).

In sum, consistent but inefficient estimation, based on the fixed effects regression, of all the parameters in (4.3.2) is possible if the columns of X_1 provide sufficient instruments for the columns of Z_2. As before, the inefficiency is grounded in the use of the fixed effects residuals.

Now we turn to efficient estimation of the model. Specifically, we seek identifying restrictions from which instruments may be formed to estimate our model consistently and asymptotically efficiently. Following H-T, consider the set of instruments

$$(4.3.12) \quad A = [M_Q, X_1, Z_1, W],$$

and the projection onto the column space of A, P_A. Then, given Propositions 1 and 2, rank and order conditions for identification are easily derived. The order condition, which is mentioned above, is the same as in H-T; namely that $k_1 > j_2$ is a necessary but not sufficient condition for the identification of β, γ, and δ_0 in (4.3.2). The rank condition is almost the same as in H-T: a necessary and sufficient condition for the identification of all the parameters in (4.3.2) is that $G'P_AG$ be nonsingular, where $G = (X,Z,W)$.

Suppose the rank condition is fulfilled by the information in (4.3.4).[13] Then, similarly to (4.2.11), we transform (4.3.2) by our $\Omega^{-1/2}$ and perform instrumental variables using the set of instruments

$$A_* = \Omega^{-1/2}(M_Q, X_1, Z_1, W) . \tag{4.3.13}$$

Equivalently, we may apply OLS to

$$\begin{aligned} P_{A_*}\Omega^{-1/2}Y &= P_{A_*}\Omega^{-1/2}X\beta + P_{A_*}\Omega^{-1/2}Z\gamma \\ &\quad + P_{A_*}\Omega^{-1/2}W\delta_0 + P_{A_*}\Omega^{-1/2}V , \end{aligned} \tag{4.3.14}$$

where P_{A_*} is the projection onto the column space of A_*. This yields

$$\begin{vmatrix} \tilde{\beta}^* \\ \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix} = (G'\Omega^{-1/2}P_{A_*}\Omega^{-1/2}G)^{-1}G'\Omega^{-1/2}P_{A_*}\Omega^{-1/2}Y . \tag{4.3.15}$$

These estimates are consistent and asymptotically efficient.

Returning to the information given in (4.3.4), we (again) distinguish three separate cases. Appendix A formally derives the characteristics of $\tilde{\beta}^*$, $\tilde{\gamma}^*$, and $\tilde{\delta}_0^*$ when the model is under-identified, exactly-identified, or over-identified. Although the derivations differ,[14] the characteristics of these estimators when $k \lesseqgtr j_2$ are essentially the same as in the H-T model. To summarize, if $k_1 < j_2$, $\tilde{\beta}^* = \tilde{\beta}_W$ and $(\tilde{\gamma}^*, \tilde{\delta}_0^*)$ does not exist. If $k = j_2$, $\tilde{\beta}^* = \tilde{\beta}_W$ and $(\tilde{\gamma}^*, \tilde{\delta}_0^*) = (\tilde{\gamma}_W, \tilde{\delta}_{0W})$, when $(\tilde{\gamma}_W, \tilde{\delta}_{0W})$ is defined in (4.3.9). And, if $k_1 > j_2$, $(\tilde{\beta}^*, \tilde{\gamma}^*, \tilde{\delta}_0^*) \neq (\tilde{\beta}_W, \tilde{\gamma}_W, \tilde{\delta}_{0W})$ with the former being more efficient than the latter.

Finally, there is the following computational note. While it is possible to calculate $\Omega^{-1/2}$ (or F) and estimate the transformed model (see Appendix B), it is not necessary. Instead, we may directly calculate

$$\begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix} \text{ and } \begin{vmatrix} \tilde{\beta}^* \\ \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix}$$

, using Ω^{-1} (or F^2). Recall Ω^{-1} is defined in (3.3.9) as $[\frac{1}{\sigma^2} M_Q + Q(Q'Q)^{-1}\Gamma^{-1}(Q'Q)^{-1}Q']$. However, this may be expressed as

$$(4.3.16) \quad \Omega^{-1} = \frac{1}{\sigma^2} M_Q + F^2 \,.$$

So, in the consistent but inefficient procedure,

$$(4.3.17) \quad \Omega^{-1/2} P_{B_*} \Omega^{-1/2} = \Omega^{-1} B(B'\Omega^{-1}B)^{-1} B'\Omega^{-1} \,,$$

Simplifying (4.3.9) to

$$(4.3.18) \quad \begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix} = [(Z,W)'F^2 B(B'\Omega^{-1}B)^{-1}B'F^2(Z,W)]^{-1} \\ (Z,W)'F^2 B(B'\Omega^{-1}B)^{-1}B'F^2\tilde{d} \,.$$

And since $P_{A_*} = M_Q$ is a projection onto $P_Q\Omega^{-1/2}B$, the efficient estimator in (4.3.15) can be computed as

$$(4.3.19)\quad \begin{vmatrix} \tilde{\beta}^* \\ \tilde{\gamma}^* \\ \tilde{\delta}_o^* \end{vmatrix} = \{G'[\frac{1}{\sigma^2} M_Q + F^2B(B'F^2B)^{-1}B'F^2]G\}^{-1}G'[\frac{1}{\sigma^2} M_Q + F^2B(B'F^2B)^{-1}B'F^2]Y$$

4.4 Summary

We have considered a random effects specification of our model in which the unreasonable independence assumption of Chapter Three is dropped and time-invariant explanatory variables are added.

Following Hausman and Taylor (1981), we derived a consistent and asymptotically efficient estimator of our model using identifying restrictions constructed from prior information about which explanatory variables are uncorrelated with the individual effects. We also derive conditions under which this estimator differs from the within estimator discussed in section three of Chapter Two.

This represents a significant improvement over the fixed effects specification since we now can estimate coefficients of time-invariant variables and gain efficiency without requiring that all the regressors be exogenous.

CHAPTER FOUR

APPENDICES

CHAPTER FOUR

APPENDIX A

To derive the characteristics of $\tilde{\beta}^*$, $\tilde{\gamma}^*$, and $\tilde{\delta}_0{}^*$ when $k_1 \lessgtr j_2$, we will make use of the following lemma, due to Trevor Breusch (personal communication):

<u>Lemma</u>: Let H and C be nxm and nxp matrices respectively, such that H and H'C both have rank m. Then H and $H(H'H)^{-1}H'C$ have the same column space.

<u>Case I</u> (under-identification): If $k_1 < j_2$, $\tilde{\beta}^* = \tilde{\beta}_W$ and $\begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_o{}^* \end{vmatrix}$ does not exist.

First, consider β. The "efficient" estimator of β, $\tilde{\beta}^*$, is obtained (separately) by OLS of $P_{A_*}\Omega^{-1/2}Y$ on the part of $P_{A_*}\Omega^{-1/2}X$ orthogonal to $P_{A_*}\Omega^{-1/2}(Z,W)$. Now, since $A_* = \Omega^{-1/2}(M_Q, X_1, Z_1, W) = (M_Q, B_*)$, $P_{A_*} = M_Q$ + projection onto the part of B_* orthogonal to M_Q; i.e.,

$$P_{A_*} = M_Q + P_Q B_* (B_*' P_Q B_*)^{-1} B_*' P_Q \tag{A.1}$$

$$= M_Q + P_Q \Omega^{-1/2} B (B' \Omega^{-1/2} P_Q \Omega^{-1/2} B)^{-1} B' \Omega^{-1/2} P_Q$$

But, $P_{A_*} = M_Q + FB(B'F^2B)^{-1}B'F$

and

(A.3) $$P_{A_*}\Omega^{-1/2} = \frac{1}{\sigma} M_Q + FB(B'F^2B)^{-1}B'F^2 .$$

Therefore,

(A.4) $$P_{A_*}\Omega^{-1/2}X = \frac{1}{\sigma} M_Q X + FB(B'F^2B)^{-1}B'F^2X$$

(A.5) $$P_{A_*}\Omega^{-1/2}(Z,W) = FB(B'F^2B)^{-1}B'F^2(Z,W)$$

(A.6) $$P_{A_*}\Omega^{-1/2}Y = \frac{1}{\sigma} M_Q Y + FB(B'F^2B)^{-1}B'F^2Y .$$

Given the above lemma (with H = FB and C = F(Z,W)), when $k_1 < j_2$, the rank of B determines the rank of $P_{A_*}\Omega^{-1/2}(Z,W)$. Thus, both $P_{A_*}\Omega^{-1/2}(Z,W)$ and FB share the same column space and null space. Hence, the part of $P_{A_*}\Omega^{-1/2}X$ orthogonal to $P_{A_*}\Omega^{-1/2}(Z,W)$ must also be orthogonal to FB. This part of $P_{A_*}\Omega^{-1/2}(Z,W)$ must also be orthogonal to FB. This part of $P_A\Omega^{-1/2}X$ is $\frac{1}{\sigma} M_Q X$. So, when $k_1 < j_2$,

(A.7) $$\begin{aligned} \tilde{\beta}^* &= [(\tfrac{1}{\sigma} M_Q X)' \tfrac{1}{\sigma} M_Q X]^{-1}(\tfrac{1}{\sigma} M_Q X)'[\tfrac{1}{\sigma} M_Q + FB(B'F^2B)^{-1}B'F^2]Y \\ &= (X'M_QX)^{-1}X'M_QY \\ &= \tilde{\beta}_W . \end{aligned}$$

Now, consider γ and δ_0. Since the column space of $P_{A_*}\Omega^{-1/2}(Z,W)$ equals the column space of FB, when $k_1<j_2$, $P_{A_*}\Omega^{-1/2}(Z,W)$ is not of full column rank. So, a (J+L) -dimensional vector ξ such that

$P_{A_*}\Omega^{-1/2}(Z,W)\xi = 0$ and $\begin{vmatrix} \gamma \\ \delta_0 \end{vmatrix}$ cannot be distinguished from

$\begin{vmatrix} \gamma \\ \delta_0 \end{vmatrix} + \xi$. Therefore, $\begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_o^* \end{vmatrix}$ does not exist in the under-identified case.

Case II (exact-identification): If $k_1 = j_2$, $\tilde{\beta}^* = \tilde{\beta}_W$ and

$$\begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix} = \begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix}.$$

Again, rank $[F(Z,W)]$ = rank (FB). So, following the argument in case I, $\tilde{\beta}^* = \tilde{\beta}_W$ when $k_1 = j_2$.

Since $\tilde{\beta}^* = \tilde{\beta}_W$ in this case,

$$\begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix} = \text{OLS of } P_{A_*}\Omega^{-1/2}\tilde{d} \text{ on } P_{A_*}\Omega^{-1/2}(Z,W)$$

(A.8)

$$\begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix} = \text{OLS of } P_{B_*}\Omega^{-1/2}\tilde{d} \text{ on } P_{B_*}\Omega^{-1/2}(Z,W)$$

Now,

(A.9) $$P_{A_*}\Omega^{-1/2}\tilde{d} = \frac{1}{\sigma} M_Q\tilde{d} + FB(B'F^2B)^{-1}B'F^2\tilde{d}$$

and

$$\text{(A.10)} \qquad P_{B_*}\Omega^{-1/2}\tilde{d} = \Omega^{-1/2}B(B'\Omega^{-1}B)^{-1}B'\Omega^{-1}\tilde{d} = \Omega^{-1/2}B(B'\Omega^{-1}B)^{-1}B'F^2\tilde{d},$$

since $B'M_Q\tilde{d} = 0$. Then note

$$\text{(A.11)} \qquad P_{B_*}\Omega^{-1/2}(Z,W) = \Omega^{-1/2}B(B'\Omega^{-1}B)^{-1}B'F^2(Z,W) \ .$$

So, the regressions of (A.8) yield

$$\text{(A.12)} \qquad \begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_0{}^* \end{vmatrix} = [(Z,W)'F^2B(B'F^2B)^{-1}B'F^2(Z,W)]^{-1}(Z,W)'F^2B(B'F^2B)^{-1}B'F^2\tilde{d}$$

$$\text{(A.13)} \qquad \begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix} = [(Z,W)'F^2B(B'\Omega^{-1}B)^{-1}B'F^2(Z,W)'F^2B(B'\Omega^{-1}B)B'F^2\tilde{d} \ ,$$

which are not generally equivalent since $B'F^2B \neq B'\Omega^{-1}B$ (specifically, $X'F^2X \neq X'\Omega^{-1}X$). However, when $k_1 = j_2$, $B'F^2(Z,W)$ is nonsingular and therefore

$$\text{(A.14)} \qquad \begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_0{}^* \end{vmatrix} = \begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix} = [B'F^2(Z,W)]^{-1}B'F^2\tilde{d} \ .$$

Case III (over-identification): If $k_1 > j_2$, $\begin{vmatrix} \tilde{\beta}^* \\ \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix} \neq \begin{vmatrix} \tilde{\beta}_W \\ \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix}$

and the former are more efficient.

If $k_1 > j_2$, rank (FB) > rank F(W,Z). Then, the column space of $P_{A_*}\Omega^{-1/2}(Z,W)$. Intuitively, this means that there are parts of $P_{A_*}\Omega^{-1/2}X$ orthogonal to $P_{A_*}\Omega^{-1/2}(Z,W)$ even though they are not orthogonal to FB. Hence $\tilde{\beta}^* \neq \tilde{\beta}_W$.

Since $\tilde{\beta}^* \neq \tilde{\beta}_W$ in this case, $(Y - X\tilde{\beta}^*) \neq \hat{d} = (Y - X\tilde{\beta}_W)$. Additionally, there is the general nonequivalence of $B'\Omega^{-1}B$ and $B'F^2B$ (which is not mentioned by H-T). So, for two reasons, $\begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix} \neq \begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix}$. And, because $\begin{vmatrix} \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{vmatrix}$ is asymptotically efficient, $\begin{vmatrix} \tilde{\gamma}_W \\ \tilde{\delta}_{0W} \end{vmatrix}$ is not in this case.

CHAPTER FOUR

APPENDIX B

We now consider the computation of the consistent and asymptotically efficient estimates defined in (4.3.15). First, we need a consistent $\hat{\Omega}^{-1/2}$. More precisely, we need consistent (as $N\to\infty$) estimation of σ^2 and Δ, since

$$\Omega^{-1/2} = \frac{1}{\sigma^2} M_Q + F = \frac{1}{\sigma^2} M_Q \tag{B.1}$$

$$+Q(Q'Q)^{-1/2} [\sigma^2 I_{NL} + (Q'Q)^{1/2}(I_N \otimes \Delta)(Q'Q^{1/2}]^{-1/2}(Q'Q)^{-1/2}Q'.$$

As in the H-T model, $\hat{\sigma}^2$ is derived from the within residuals. Let $\tilde{Y} = M_Q X$, $\tilde{\varepsilon} = M_Q \varepsilon$, and $M_{\tilde{X}} = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$. Then, the SSE from the within regression in (2.3.6) may be written as

$$\tilde{Y}'M_{\tilde{X}}\tilde{Y} = \tilde{\varepsilon}'\tilde{\varepsilon} - \tilde{\varepsilon}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\varepsilon} \;, \tag{B.2}$$

and therefore

$$\hat{\sigma}^2 = \frac{1}{N(T-2)} \tilde{Y}'M_{\tilde{X}}\tilde{Y} = \frac{1}{N(T-2)} Y'M_Q(Y - X\tilde{\beta}_W). \tag{B.3}$$

This estimator is consistent:

$$\text{(B.4)} \qquad \text{plim } \hat{\sigma}^2 = \text{plim } \frac{1}{N(T-2)} \tilde{Y}' M_{\tilde{X}} \tilde{Y} = \text{plim } \tilde{\varepsilon}'\tilde{\varepsilon} = \sigma^2,$$

since $\tilde{\varepsilon}'\tilde{\varepsilon} = \varepsilon'\varepsilon - \varepsilon' Q(Q'Q)^{-1} Q'\varepsilon$.

A consistent estimator of Δ may be constructed as follows. Perform instrumental variables on

$$\text{(B.5)} \qquad (Y - X\tilde{\beta}_W) = Z\gamma + W\delta_0 + QU_* + \varepsilon \ ,$$

using $B = (X_1, Z_1, W)$ as instruments. From the IV residuals, we can form

$$\text{(B.6)} \qquad \tilde{\Omega} = ee' = \begin{vmatrix} e_1 e_1' & & \\ & \ddots & \\ & & e_N e_N \end{vmatrix} ,$$

where e_i denotes a TX1 vector of IV residuals.. Obviously, $\tilde{\Omega}$ is not a consistent estimator of Ω. But,

$$\text{(B.7)} \qquad Q\tilde{\Omega}Q = \sigma^2 Q'Q + Q'Q(I_N \otimes \Delta)Q'Q \ ,$$

and

$$\text{(B.8)} \qquad (Q'Q)^{-1} Q'\tilde{\Omega}Q(Q'Q)^{-1} - \sigma^2 (Q'Q)^{-1} = I_N \otimes \Delta.$$

This suggests

(B.9) $$\hat{\Delta} = \frac{1}{N}\sum_i [(W_i'W_i)^{-1}W_i'e_ie_i'W_i(W_i'W_i) - \hat{\sigma}^2(W_i'W_i)^{-1}]$$

Since the IV estimates of β, γ, and δ_0 are consistent,

(B.10)
$$\text{plim}\,\hat{\Delta} = \text{plim}\,\frac{1}{N}\sum_i [(W_i'N_i)^{-1}W_i(W_iu_i+\varepsilon_i)(W_iu_i+\varepsilon_i)'W_i(W_i'W_i)] - \frac{1}{N}\sum_i \sigma^2(W_i'W_i)^{-1}$$

$$= \text{plim}\,\frac{1}{N}\sum_i u_iu_i'$$

$$= \Delta\ .$$

Hence, $\hat{\Delta}$ is consistent.

Given σ^2 and $\hat{\Delta}$, $\Omega^{-1/2}$ can be easily calculated once the troublesome terms of F are decomposed into matrices of their eigenvalues and eigenvectors. For example, $(Q'Q)^{-1/2} = PD^{-1/2}P'$, where P is an orthogonal matrix of the eigenvectors of Q'Q, and D is a diagonal matrix of the eigenvalues of Q'Q.

To implement the efficient estimation procedure defined by (4.3.14), note that (4.3.2) and (4.3.4) may be expressed as a system of one structural equation and two reduced form equations:

(B.11) $$\Omega^{-1/2}Y = \Omega^{-1/2}X\beta + \Omega^{-1/2}Z\gamma + \Omega^{-1/2}W\delta_0 + \Omega^{-1/2}V$$

(B.12) $$\Omega^{-1/2}X_2 = \Omega^{-1/2}X_1\Pi_{11}+\Omega^{-1/2}Z_1\Pi_{12}+\Omega^{-1/2}W\Pi_{13}+\Omega^{-1/2}M_Q\Pi_{14}+\eta_1$$

(B.13) $$\Omega^{-1/2}Z_2 = \Omega^{-1/2}X_1\Pi_{21}+\Omega^{-1/2}Z_1\Pi_{22}+\Omega^{-1/2}W\Pi_{23}+\Omega^{-1/2}M_Q\Pi_{24}+\eta_2.$$

The fitted values of $\Omega^{-1/2}X_2$ and $\Omega^{-1/2}Z_2$ are calculated from their reduced forms. However, estimation of (B.12) and (B.13) is made cumbersome due to the presence of M_Q. Alternatively, we may "parse out" M_Q in both reduced form equations by premultiplying by P_Q. This allows us to calculate $\Omega^{-1/2}Z_2$ from a regression of FZ_2 on FB. To obtain $\Omega^{-1/2}X_2$, we must replace M_QX_2 "parsed out" by P_Q. So, we form $\Omega^{-1/2}X_2 = M_QX_2 + P_Q\Omega^{-1/2}X_2 = M_QX_2 + \hat{FX}_2$, where $\hat{FX}_2$ are the fitted values obtained from a regression of FX_2 on FB.

Finally, $\Omega^{-1/2}X_2$ and $\Omega^{-1/2}Z_2$ can be combined with X_1 and Z_1, in a least squares regression to obtain consistent and asymptotically efficient estimates of β, γ, and δ_0.

CHAPTER FIVE

ESTIMATION OF UNION WAGE DIFFERENTIALS

The previous three chapters have presented some theoretical results on the estimation of panel data models with cross-sectional variation in both slopes and intercept. In this chapter we will apply these results to the problem of measuring the impact of unions on wages.[1] For the purpose of exposition let us begin with a brief methodological review.

Prior to the existence of large longitudinal data sets, attempts to measure the union wage differential were characterized by the estimation of various forms of a standard cross-sectional earnings equation. To consider a specific form, let

$$(5.1.1) \qquad Y_i = X_i'\beta + U_i\delta + \varepsilon_i \ , \qquad i = 1,\ldots,N \ ,$$

where Y_i is the natural logarithm of the wage, X_i is a K-dimensional vector of explanatory variables (e.g., education, experience, sex), U_i is a binary variable indicating whether individual i's job is covered by a union contract, and ε_i is an iid disturbance. Ordinary least squares estimation of (5.1.1) provides an estimate of δ which is interpreted as the union relative wage differential.

One of the usual objections to estimates based on (5.1.1) is that, given high union wages, firms will hire workers of commensurate quality, resulting in

$$(5.1.2) \quad E(\varepsilon_i U_i) > 0 \ .$$

In other words, there exists an omitted variables problem, since in the simple cross-section we are unable to control for those individual wage determining attributes - collectively referred to as "ability" - which are correlated with union status. Consequently, ε_i is positively correlated with U_i and the least squares estimate of δ is biased upward.

With the availability of panel data, these individual attributes, or effects, can be taken into account. Typically this has been done within a fixed-effects framework,[2] where

$$(5.1.3) \quad Y_{it} = X_{it}'\beta + U_{it}\delta + \alpha_i + \varepsilon_{it}, \qquad i=1,\ldots,N, \ t=1,\ldots,T \ .$$

and α_i is the individual effect, assumed constant over time. The inclusion of the α_i's can be viewed as a form of differencing. Indeed, this is the essence of the within transformation, to difference away the correlation between the error and union status. Then, applying least squares to the within-transformed equation yields an unbiased and consistent estimate of the union wage differential.

The notion of omitted variable bias is supported by those studies in which models like (5.1.3) are estimated. Invariably these analyses report lower union wage effects than do cross-section studies.[3]

Now, models like (5.1.3) are also subject to criticism. First, there is the issue of measurement error with respect to the union status variable. This problem is not addressed in this paper, but Freeman (1984) and Chowdhury and Nickell (1985) correctly point out that errors in the reporting of union status are accentuated in longitudinal studies since the estimation of panel data models usually depends on a small number of union status changes.

A second criticism, which we do consider, is offered by Stewart (1983). He notes that the standard fixed effects model (varying intercept only) ignores the possibility that the individual effects may be sector dependent. In recognizing that the processes which determine wages are different in the union and nonunion sectors, Stewart constructs a model which allows the union wage differential to vary across people. As we will see below, his model is just a special case of the fixed effects version of our model presented in section 3 of Chapter Two.

We may express Stewart's model as

$$(5.1.4) \qquad Y_{it} = X_{it}'\beta + U_{it}\gamma + \phi_{it}\alpha_i + \varepsilon_{it} ,$$

where

$$(5.1.5) \qquad \phi_{it} = \left| \begin{array}{l} 1, \text{ if } U_{it} = 0 \\ \\ \lambda, \text{ if } U_{it} = 1 \end{array} \right| .$$

Clearly, if $\lambda = 1$ (5.1.4) reduces to the standard fixed effects model. If $\lambda \neq 1$, then the individual effects are sector dependent. Stewart suggests $\lambda<1$ since union contracts, in their standardization of wage rates, "may alter the link between pay and productivity." That is to say, the individual effects may have greater impact in the nonunion sector.[4]

The Stewart model is seen as a special case of our own when we write

$$(5.1.6) \qquad Y_{it} = X_{it}'\beta + U_{it}\delta_i + \alpha_i + \varepsilon_{it} ,$$

where $\delta_i = \gamma + (\lambda-1)\alpha_i$ is the union wage differential for individual i. For simplicity, in our empirical application of Stewart's model, we consider an unrestricted version of (5.1.6), one in which δ_i is not constrained to be a linear function of the individual effect. This allows us to investigate, with little computational difficulty, fixed and random effects estimation of Stewart's model.[5]

In the next two sections, we present our empirical results on the union wage effect. We compare estimates from the three structural frameworks described in (5.1.1), (5.1.3), and (5.1.6). In addition, we contrast the fixed and random effects estimates of the simple and general panel data models. Our conclusions are presented in section four.

5.2 Data

Our data set consists of 1706 heads of household taken from the Michigan Panel Study of Income Dynamics (PSID). We consider the four years 1978-1981, and restrict the sample to include only those individuals who report receiving a nonzero wage (each year), are not in the military, and were between the ages of 18 and 65 in 1978. The variables used are those commonly found in an earnings equation. They are described in some detail below. Table 1 gives the means and standard deviations of these variables for each of the four years and for the pooled sample.

The dependent variable is the natural logarithm of the wage (LW). The set of regressors includes the following continuous explanatory variables. Education (ED) is measured by years (grades) of schooling completed. Notice this variable is time-invariant. Experience (X) is recorded as the number of years that an individual reports he/she has worked full time. Now, in the PSID, the values of experience in 1979-1981 are the same as the value of experience recorded in 1978. Therefore, we construct experience responses for the last three years. We do this by simply adding one year of experience to the 1978 value for each of the following three years. The tenure (TEN) variable is not the preferred "tenure with the employer." For 1978-1980, the PSID only has observations for "tenure on the job." So, this is what we use, measured in months. As is standard procedure, we also include a quadratic in experience (X^2) and tenure (TEN^2).

Other individual characteristics are defined by a set of binary variables. If an individual lives in a standard metropolitan statistical area (i.e., a city with population > 50,000), city size (SMSA) is given the value 1. Similarly, we use a redneck (REG) dummy, set equal to 1, if an individual lives in the South. Marital status (MARR) is recorded with a value of 1 if married. The gender variable (SEX) is equal to 1 if an individual is male. An individual's race

TABLE 1

Means (and Standard Deviations)

Variables	1978	1979	1980	1981	POOLED
LW	6.3706 (.4843)	6.4779 (.4825)	6.5773 (.4707)	6.6616 (.4688)	6.5218 (.4902)
ED	12.1219 (2.8715)	12.1237 (2.8687)	12.1237 (2.8687)	12.1237 (2.8687)	12.1232 (2.8687)
X	13.9461 (11.0018)	14.9461 (11.0018)	15.9461 (11.0018)	16.9461 (11.0018)	15.4461 (11.0561)
X^2	315.4619 (412.2647)	344.3540 (433.4538)	375.2462 (454.7202)	408.1383 (476.0537)	360.8001 (446.0018)
TEN	70.0305 (77.6660)	77.7907 (81.0625)	88.4877 (90.0519)	119.2339 (103.6219)	88.8882 (90.6037)
TEN^2	10932.7456 (23348.9671)	12618.6829 (25332.7136)	15936.4297 (30691.7007)	24947.4209 (40331.7051)	16108.9448 (31108.8640)
SMSA	.6940 (.4610)	.6858 (.4643)	.6899 (.4627)	.6858 (.4643)	.6889 (.4630)
REG	.4302 (.4953)	.4314 (.4954)	.4308 (.4853)	.4326 (.4956)	.4313 (.4953)
MARR	.7145 (.4518)	.7216 (.4484)	.7315 (.4433)	.7292 (.4445)	.7242 (.4469)
SEX	.8224 (.3823)	.8224 (.3823)	.8224 (.3823)	.8224 (.3823)	.8224 (.3822)
RACE	.6676 (.4712	.6676 (.4712)	.6676 (.4712)	.6676 (.4712)	.6676 (.4712)
CB	.3458 (.4893)	.3646 (.4825)	.3581 (.4707)	.3628 (.4688)	.3579 (.4902)

N = 1706

(RACE) is defined as 1 if he/she is white (the nonwhite category only consists of blacks). Union coverage (i.e., whether a person's job is covered by a union contract), rather than union membership, is chosen as the union status variable. It (CB) is set to 1 if a person's job is unionized. Finally, we include a series of one-digit occupation dummies to control for the effect of occupation on wages.

In the next section we combine the data with the structural models described in 5.1 to examine the impact of unions on wages.

5.3 Estimation and Results

Here we present the results of estimation. Our primary concern is to gain a better understanding of the union wage effect. Of related interest is a comparison of the results obtained from the simple and general panel data models (in particular, the within and H-T estimators). We proceed by considering, in turn, the models described by (5.1.1), (5.1.3), and (5.1.6).

As mentioned in section one of this Chapter, the first attempts to measure the union wage differential were simple cross-section studies where ordinary least squares was applied to an equation like (5.1.1). We replicate this procedure on each of the four years of our sample. These cross-section results are given in Table 2. In general, our estimates are very similar to those obtained in other cross-section studies. Two exceptions are the coefficients of tenure and tenure-squared. In no year are these coefficients estimated with much precision. This is due, at least in part, to the fact that our tenure variable is defined as "tenure on the job" rather than the preferred

TABLE 2

Cross-Section Estimates
Dependent Variable: LOG Wage

Explanatory Variables	1978	1979	1980	1981
ED	.0472 (.0039)	.0481 (.0038)	.0505 (.0037)	.0547 (.0037)
X	.0238 (.0028)	.0235 (.0029)	.0214 (.0029)	.0167 (.0032)
X^2	-.00044 (.00007)	-.00047 (.00007)	-.00038 (.00007)	-.00032 (.00007)
TEN	.00372 (.00360)	.00336 (.00336)	-.00264 (.00300)	.00840 (.00288)
TEN^2	-.00004 (.00013)	.0000006 (.0001253)	.00014 (.00010)	-.00012 (.00009)
SMSA	.1173 (.0187)	.1164 (.0183)	.1170 (.0177)	.1193 (.0179)
REG	-.0481 (.0185)	-.0547 (.0183)	-.0520 (.0177)	-.0509 (.0181)
MARR	.0700 (.0263)	.0827 (.0264)	.1102 (.0264)	.1325 (.0272)
SEX	.2585 (.0334)	.2299 (.0331)	.2035 (.0327)	.1865 (.0331)
RACE	.1071 (.0202)	.1389 (.0201)	.1204 (.0196)	.1305 (.0199)
CB	.1758 (.0188)	.1573 (.0182)	.1915 (.0179)	.1809 (.0182)
$\bar{R}^2$	.550	.551	.556	.535

N = 1706 . Standard errors in parentheses "tenure with the employer." The coefficient on union status ($\hat{\delta}$) ranges from .157 to .191, which implies a union wage differential of 17 to 21 percent.[6] Like our parameter estimates in general, this range of estimated union wage differentials is in agreement with earlier cross-section results. Perhaps also not surprising is the decline and rise in the union wage effect over the four year period. This may be explained by the incompleteness of union cost-of-living adjustments (COLA's) during the inflationary period of the late 1970's, and the countercyclical nature of the union wage differential (demonstrated by the effects of the 1980 recession).

Since we strongly suspect the selectivity of union workers causes an upward bias in the cross-section estimates of the union wage effect, we turn to the panel data model of (5.1.3), where each individual has a unique intercept (α_i, the intercept effect). For the sake of comparison, we first estimate this model by OLS. These results are presented in the first column of Table 3. Notice they are vary much like those obtained by the simple cross-section regressions. This is because OLS ignores the longitudinal nature of the data. In other words, OLS assumes no correlation between the explanatory variables and the individual effects. So, the $\hat{\delta}$ calculated by OLS is upwardly biased for the same reason as the union status coefficients given in Table 2.

The claim of this bias in the cross-section and OLS/panel estimates of δ is supported by the fixed effects results given in the third column of Table 3. Performing OLS on a within transformation of the data (i.e., deviations from individual means) yields an estimate of the union

TABLE 3

Panel Estimates: Intercept Varying
Dependent Variable: LOG Wage

Explanatory Variables	OLS	GLS	Within	HT
ED	.0514 (.0020)	.0728 (.0152)		.1573 (.0677)
X	.0233 (.00151)	.0960 (.0017)	.1164 (.0027)	.1155 (.0027)
X^2	-.00045 (.00004)	-.0021 (.0001)	-.00063 (.00007)	-.00062 (.00007)
TEN	.00648 (.00156)	.00360 (.00003)	.00096 (.00003)	.00180 (.00003)
TEN^2	-.0000003 (.0000004)	-.0000011 (.0000002)	-.0000003 (.0000002)	-.0000005 (.0000002)
SMSA	.1132 (.0094)	-.0108 (.0156)	-.0150 (.0157)	-.0151 (.0157)
REG	-.0519 (.0094)	-.0427 (.0285)	-.0416 (.0298)	-.0308 (.0291)
MARR	.1003 (.0138)	.0058 (.0132)	-.0027 (.0131)	-.0045 (.0131)
SEX	.2298 (.0171)	.3393 (.1065)		.2912 (.1154)
RACE	.1247 (.0104)	.1712 (.0917)		-.0450 (.1761)
CB	.1803 (.0095)	.0975 (.0096)	.0945 (.0096)	.0941 (.0096)
$\bar{R}^2$	.536	.397	.405	.403

N = 1706. Standard errors in parentheses.

status coefficient of .094, a 65 percent decrease from the cross-section estimate.[7] With the exceptions of experience and experience-squared, the impacts of the other explanatory variables are reduced by allowing for the individual effects. The within estimates are unbiased and consistent even if the effects are correlated with the regressors. However, they are not fully efficient and the within transformation removes all time-invariant variables (ED, SEX, RACE) from the model. As we stated at the outset of Chapter Three, this is a potentially serious problem if one is interested in, for example, the return to schooling.

An alternative specification of (5.1.3) is the error components model, which was reviewed in section two of Chapter Three. In this case, the α_i are taken to be iid random variables uncorrelated with the regressors. If this assumption is true, generalized least squares based on consistent estimates of the variance components will produce consistent and asymptotically efficient estimates of all the parameters, including the coefficients of the time-invariant variables; and GLS is simply computed. It is equivalent to OLS on a $(1-\theta)$ differencing of the data, where $\theta = [\sigma^2/(\sigma^2+T\sigma^2_\alpha)]^{1/2}$ (see Note 4 of Chapter Four). Consistent estimates of σ^2 and σ^2_α can be derived from the within residuals (see H-T, p. 1384). In our case,

$$(5.3.1) \quad \hat{\sigma}^2 = .025, \quad \hat{\sigma}^2_\alpha = 3.418, \quad \hat{\theta} = .043 .$$

The results of GLS estimation are listed in the second column of Table 3. Now we have argued that there exists an omitted variables problem in the cross-section. Since GLS also assumes no correlation between the effects and the explanatory variables, the error-components specification is vulnerable to the same criticism leveled at the cross-section estimates. The GLS estimate of δ (.098) is higher than the within estimate (.094). However, it is surprisingly low given the argument of omitted variables bias. In any case, if the effects are correlated with union status, the GLS estimate is biased and inconsistent. The coefficient on education, which we also expect to be correlated with the effects, is closer to the OLS estimate. In sum, unless we are prepared to reject the story of selectivity of union workers, the fixed effects specification should be preferred over the error components model for measuring the union wage differential.

However, we need not rely solely on the fixed effects version of (5.1.3). Instead, we may let the α_i be random variables correlated with the regressors. Then, following the H-T coefficient procedure described in Chapter Four, we are able to include time-invariant explanatory variables, and obtain consistent and asymptotically efficient estimates, thereby meeting the objections to the OLS, within, and GLS estimators.

Implementation of this procedure requires that we be able to distinguish those regressors that are correlated with the effects from those that are not. Since the effects presumably control for ability, obvious choices as endogenous variables are union status and education. In addition, applications of this technique by H-T and Chowdhury and Nickell lead us to include experience and tenure (and their quadratic terms) as endogenous variables.[8] Recalling the H-T order condition for identification of the model, we know that we need at least as many time varying explanatory variables (SMSA, REG, MARR) uncorrelated with the effects as we have time invariant explanatory variables correlated with the effects (ED). Clearly this condition is fulfilled here.[9] In fact, the model is over-identified. Hence, the H-T estimator should yield an efficiency improvement over the within estimator for this model.[10]

Beyond determining which explanatory variables are endogenous, the only major difficulty is computational. However, the computational difficulty can be reduced. Following H-T (Appendix B), the fitted values of the endogenous variables (both time-varying and time-invariant) can be calculated from their reduced forms in a manner that reduces the size of the estimation problem from sample size NT regressions to sample size N regressions. The predicted values of the time-invariant endogenous variables are obtained from a regression of these variables on the time-invariant exogenous variables and the individual means of the time varying exogenous variables. The calculation of the predicted values of the time-varying endogenous variables is almost as simple. They are derived from a regression of the individual means of the time-varying endogenous variables on the time-invariant exogenous variables and the individual means of the time-varying exogenous variables. To these predicted individual means we must add the true deviations from means since the correct prediction of the time-varying endogenous variables is calculated with the set of instruments that includes the projection used to transform the data into deviations from means. Then, fitted values are combined with the variables that are uncorrelated with the effects to obtain consistent and asymptotically efficient estimates of β and δ.

The results from the H-T efficient estimation of (5.1.3) are presented in the last column of Table 3. First, note that the estimated coefficient on union status is essentially the same as that calculated by within. This should not be surprising since correlation between the effects and union status is assumed by this specification. Education is also taken as endogenous and the effect of this is a marked increase in its coefficient, to .157, over both the OLS (.051) and GLS (.073) estimates. This rise in the returns to schooling is not in accordance with a story of positive correlation between education and ability leading to an upwardly biased OLS (or GLS) estimate. But, H-T point out that when the amount of education is endogenous, there may be a negative correlation between ability and the amount of education chosen.[11] Their application of this procedure to an investigation of the returns to schooling reveals a similar rise over OLS and GLS estimates.

The coefficients of the other time-varying explanatory variables are reasonably close to the within estimates. The sex and race parameters, which cannot be estimated by within, are both considerably different from either the OLS or GLS estimates. In particular, notice the effect of race has essentially vanished.

Now, the H-T procedure produces parameter estimates which support the omitted-variables bias argument, and are consistent and fully efficient. However, within the framework of (5.1.3) we are still vulnerable to the criticism of Stewart. Thus, we next consider the unrestricted version of (5.1.6), where the union wage differential is allowed to vary cross-sectionally. We examine (as we did for the simple model), the fixed-effects and two random-effects specifications of this special case of our general panel data model.

First, we take δ_i and α_i in (5.1.6) to be fixed over time. The fixed effects version of the general model is estimated by performing OLS on the within (deviations from means) transformation of the simple model (see chapter two). These results are presented in the second column of Table 4. Since T=4, the individual δ_i's cannot be estimated consistently. However, their mean, $\frac{1}{N}\sum_i \hat{\delta}_i$ equals .093, which is only slightly less than the usual fixed effects estimate. In general, the coefficients of all the explanatory variables are not too different from those calculated by within on the simple model, even though the effects are now composed of both δ_i and α_i. Finally, notice the time-invariant variables are (again) eliminated by the transformation.

We have consistently argued that any estimation method which assumes no correlation between the regressors and the effects is inappropriate for measuring the union wage effect. However, for the sake of comparison, let δ_i and α_i be iid random variables uncorrelated with the explanatory variables. With this specification we estimate by GLS, but we first need consistent estimates of the variance of ε_{it} and the covariance matrix of (α_i,δ_i). Using the estimates defined in Appendix B to Chapter Four, we obtain

$$\hat{\sigma}^2 = .032$$

(5.3.2)
$$\hat{\Delta} = \begin{vmatrix} .774 & .003 \\ .003 & .043 \end{vmatrix} .$$

Given (5.3.2) we can calculate a consistent estimator of $\Omega^{-1/2}$ (again, see chapter four, appendix B). Then, GLS is obtained by performing an $\hat{\Omega}^{-1/2}$ transformation on the data and running OLS.

The resulting estimates, which are only consistent if the uncorrelatedness assumption is true, are given in the first column of Table 4. As in the simple model, the GLS estimate of the union wage effect (.095) remains higher than, though close to, the fixed effects estimate. The estimates of the other parameters, with the exception of the coefficients of the experience and tenure variables, are very different from those calculated by GLS on the simple model. One reason for this must be the inclusion of the δ_i as part of the random effects, which are assumed to be uncorrelated with the regressors. These results should not be too upsetting, however, since they are based on an unrealistic assumption and are therefore biased and inconsistent.

Finally, we address the drawbacks of within and GLS. Now we take the α_i and δ_i's to be random variables correlated with the explanatory variables. To this version of (5.1.6) we apply our extension of the H-T analysis. The variables taken to be endogenous in the simple model are also assumed to be correlated with the effects here. The only exception is union status, which is now exogenous.[12] Like the simple model, our model is over-identified, and therefore the H-T procedure should yield consistent and asymptotically efficient estimates of the general model.

These estimates are presented in the last column of Table 4. After distinguishing those variables that are uncorrelated with the effects, as in the simple model, the main difficulty is computational. In an analogous fashion to the procedure outlined for the simple model, we calculate the fitted values of the endogenous variables from their reduced forms. For the general model, however, efficient estimation requires the reduced forms be transformed by $\hat{\Omega}^{-1/2}$ (constructed from $\hat{\sigma}^2$ and $\hat{\Delta}$ given in (5.3.2)). (General reduced form expressions for those variables correlated with the effects are

TABLE 4

Estimates: Slopes and Panel Intercept Varying
Dependent Variable: LOG Wage

Explanatory Variables	GLS	Within	HT
ED	.3364 (.0045)		.3743 (.0227)
X	.1236 (.0031)	.1141 (.0026)	.1677 (.0123)
X^2	−.00142 (.00008)	−.00061 (.00007)	−.00095 (.00033)
TEN	.00204 (.00120)	.00072 (.00096)	−.00168 (.00480)
TEN^2	−.00004 (.00004)	−.00003 (.00003)	−.00001 (.00016)
SMSA	.0276 (.0175)	−.0144 (.0153)	−.0624 (.0328)
REG	.1272 (.0297)	−.0220 (.0290)	−.0677 (.0556)
MARR	−.0212 (.0151)	−.0091 (.0126)	−.0208 (.0237)
SEX	1.1048 (.0573)		.3498 (.1658)
RACE	−.0610 (.0522)		−.4510 (.1135)
CB	.0955 (.0074)	.0930	.0854 (.0128)
$\bar{R}^2$	.911	.403	.791

N = 1706. Standard errors in parentheses.

defined in chapter four, appendix B).

Direct application of OLS to the ($\hat{\Omega}^{-1/2}$ transformed) reduced forms is cumbersome since the set of instruments includes not only the exogenous variables, but also the projection used to transform the fixed-effects model. This projection can be "parsed out" in both reduced forms before performing OLS. However, to obtain the correct predicted values of the time-varying endogenous variables, the "within" transformed exogenous variables must be added back to the fitted values calculated from the least-squares regression with the "within" projection "parsed out". (This is formally described in appendix B to chapter four). The predicted values of the endogenous variables are then combined with the ($\hat{\Omega}^{-1/2}$ transformed) exogenous variables in an OLS regression to obtain consistent and asymptotically efficient estimates of the parameters in the general model. The union status coefficient is estimated to be .085, implying a union wage differential of 8.9 percent. This is very close to the mean of the $\hat{\delta}_i$'s derived from the within regression. As in the simple model, the H-T estimate of the return to schooling is much higher than that calculated by the original OLS regression. This lends further support to the story offered by H-T of a negative correlation between ability and education when the amount of schooling is made endogenous. The coefficients on the other time-varying explanatory variables are closer to the within estimates of the general model than those obtained by GLS. However, one particularly peculiar result is the estimate of the race parameter. In the simple model the race effect essentially disappears. In their return to schooling exercise, H-T report a similarly reduced effect of race on earnings from their efficient procedure; but, a negative race effect of such magnitude is completely unexpected. Why the inclusion of δ_i as part of the individual effect would lead to this result is unclear.

5.4 Conclusions

Table 5 contains a tabular summary of the union wage differentials obtained from the different models and estimation techniques we have considered. Because of the selectivity of union workers it is likely that ability is correlated with union status. If union status is observed without error, then we conclude that there is no real justification for measuring the union wage differential from the simple cross-section, or from a panel estimation method which does not allow for correlation between the regressors and the individual effects. The simplest means of estimating the union wage effect is within estimation of the usual fixed effects model. These estimates are unbiased and consistent. Similar, and likewise consistent results may be obtained by estimating a fixed effects version of our general model (e.g., Stewart's specification). In this way we can let the union wage differential vary cross-sectionally (or allow for sectoral dependence of the effects). Finally, if we are also concerned about the coefficients of the time invariant explanatory variables, we can apply H-T to the simple (intercept varying) model and obtain consistent and asymptotically efficient estimates of all the parameters in the model (provided the parameters are identified). In the general model, the H-T analysis provides somewhat peculiar results for the coefficients of the time-invariant variables. This may be due to misspecification, since union status is not endogenous in the general model. In sum, all of the

TABLE 5

Percent Union Wage Effects

	Cross-Section	Panel	OLS	WITHIN	GLS	HT
1978:	19.22	Intercept Varying:	19.76	9/91	10.24	9.87
1979:	17.03	Slopes and				
1980:	21.11	Intercept Varying:		9.75	10.02	8.92
1981:	19.83					

estimation methods which allow for correlation between the individual effects and the explanatory variables yield union wage differentials that are much smaller than those obtained in simple cross-section or OLS/panel estimation.

CHAPTER SIX

SUMMARY AND CONCLUSIONS

A regression function which does not control for omitted or unobservable variables that are correlated with the explanatory variables will not identify the parameters of the model. Conventional estimation of such a regression function will produce biased and inconsistent results. However, the availability of panel data allows us to control for these omitted or unobserved characteristics through the inclusion of individual specific parameters or effects.

The focus of this study is on the estimation of panel data models in which there is cross-sectional variation in some of the slopes as well as the intercept. A well established literature exists on the estimation of the simpler case in which only the intercept varies cross-sectionally. The results for the simple model are a function of the assumptions about the individual effects. We identify three different cases: (1) fixed effects, (2) random effects uncorrelated with the regressors, and (3) random effects correlated with the regressors.

For each set of assumptions we review the appropriate method of estimation for the simple model and then extend these results to our more general model. In both the review and the extension, our primary interest is in estimation techniques that behave well when we have a small number of time observations on a large cross-section, a typical case with longitudinal data.

First, we consider the weakest set of assumptions: fixed effects. In this case, the individual effects, while differing across people, are assumed to be constant for each person. Traditionally, the simple model is estimated by OLS on the within transformed data. The within estimator is consistent for fixed T. We develop the analogous transformation for the general model and show that OLS on our transformed model is also consistent (given a reasonable assumption about the variability of the regressors) in the case of fixed T. In addition, we prove that under normality, the within estimator of the general model is also the conditional MLE.

Two drawbacks of the fixed effects specification are noted. The first, and perhaps less serious, is that (for both models) the within estimator is not fully efficient when T is small. Secondly, time-invariant explanatory variables are orthogonal to, and therefore eliminated by, the within transformation.

One solution to these two problems is to adopt a random-effects specification where one assumes the individual effects are iid random variables uncorrelated with the regressors. In the model with only a variable intercept, estimation is by GLS. The GLS estimator is consistent and efficient relative to within when T is small. We derive the GLS estimator for our model and show that the results from the simple model carry over to the general model. However, the consistency of GLS in both models hinges on the assumption that the effects be uncorrelated with the explanatory variables. This assumption is not justified in most empirical applications.

Finally, we consider the case of random effects which are correlated with the regressors. Under this set of assumptions, Hausman and Taylor (1981) derive an instrumental variables procedure for the simple model that allows the inclusion of time invariant variables, and yields consistent and asymptotically efficient estimates. Their IV estimator is unique in that it uses the included exogenous time-varying variables as instruments for the endogenous time-invariant variables. Specifically, their procedure requires that we have at least as many time-varying explanatory variables uncorrelated with the effects as we have time invariant explanatory variables correlated with the effects. This is essentially an order condition for the identification of the parameters in the model. We apply the H-T analysis to the case in which slopes and intercepts are allowed to vary. An analogous order condition for the identification of our model is obtained, and a consistent and asymptotically efficient IV estimator is derived. Then, following H-T, we detail conditions under which the efficient IV estimator of our model differs from within.

After our theoretical examination of panel data models in which slopes and intercepts are allowed to vary, as an empirical exercise we consider the estimation of unions' impact on earnings. The issue of the union wage effect offers an appropriate empirical question for the application of a special case of our model.

The first attempts to measure the union wage effect were conducted within the framework of a simple cross-sectional earnings equation containing a union status dummy. These studies ignored the selectivity of union workers. Given high union wages, firms tend to select more able workers, producing a positive correlation between union status and the disturbance. Put differently, there is an omitted variables problem since "ability" is ignored in the cross-section regression. The result is a measure of the union wage differential which is upwardly biased.

With the availability of panel data, the typical response to the biased cross-section results has been through the standard fixed effects panel data model. Here individual specific intercepts (effects) are included to control for ability. This can be viewed as a form of differencing, which is the essence of the within transformation. Estimation by within yields an unbiased and consistent estimate of the union wage differential.

However, the usual fixed effects model is not without its critics. Stewart (1983), noting that the processes which determine wages are different in the union and nonunion sectors, constructs a model that allows for the individual effects to be sector dependent. This is equivalent to letting the union wage differential vary across the individuals in the sample. Thus Stewart's model is just a special case of the fixed effects version of our general panel-data model. In his model, the union wage differential is constrained to be a linear function of the individual effects. Estimation of his model is by nonlinear least squares. We consider an unrestricted version of Stewart's model which is estimated by within. In either case, the individual union wage effects cannot be estimated consistently as long as T is fixed. However, we can calculate the average union wage differential for the sample.

Using data from the years 1978-1981 of the PSID we estimate: (1) the simple cross-sectional earnings equation, (2) the usual fixed effects model, and (3) the unconstrained Stewart model. Since within estimation of (2) and (3) eliminates time invariant variables (e.g., education), and because the resulting estimates are not fully efficient in the case of small T (here, T=4), we also estimate the random-effects (correlated and uncorrelated with regressors) specifications of each.

From (1) we obtain estimates of the union wage differential in the range of 17 to 21 percent for the years considered. These results are in agreement with other cross-section studies. Estimation of the standard fixed-effects model yields a much smaller union wage differential of 9.9 percent. Averaging the individual union wage effects obtained from within estimation of (3) produces a measure of the union wage differential of about 9.7 percent. Similar results are obtained when we take the effects to be random and correlated with the regressors. In general, we find that in every case where we allow correlation between the regressors and the individual effects, the story of upward bias in the cross-section is confirmed with substantially reduced estimates of the impact of unions on wages.

We conclude from this empirical exercise that unless there is interest in the coefficients of the time invariant explanatory variables, the fixed effects specification of either the simple or general model provides a satisfactory framework for measuring the union wage effect. And, since within estimates of each are always consistent, they can serve as a basis of comparison for results from more restrictive models.

Our final remarks concern what remains to be done. Aside from other applications of the theoretical results presented here, there are further extensions of our model to be considered. One is to allow some of the variables associated with the cross-sectionally varying coefficients to be correlated with the individual effects. Another extension might explore the limits of conditional likelihood analysis in a system of simultaneous equations with panel data, where some of the slope coefficients vary across individuals. Such considerations are topics for future research.

FOOTNOTES

CHAPTER TWO

NOTES

1 To see this, notice M_D may also be expressed as

$$M_D = I_{NT} - [I_N \otimes \frac{1}{T} e_T e_T'] \ .$$

So, prelmultiplication of the data by M_D yields

$$M_D Y = \begin{bmatrix} Y_{11} - \bar{Y}_1 \\ Y_{12} - \bar{Y}_1 \\ \vdots \\ Y_{1T} - \bar{Y}_1 \\ \hline Y_{21} - \bar{Y}_2 \\ Y_{22} - \bar{Y}_2 \\ \vdots \\ Y_{2T} - \bar{Y}_2 \\ \hline \vdots \\ \hline Y_{N1} - \bar{Y}_N \\ Y_{N2} - \bar{Y}_N \\ \vdots \\ Y_{NT} - \bar{Y}_N \end{bmatrix} , \quad \bar{Y}_i = \frac{1}{T} \sum_t Y_{it}$$

and $M_D X$ similarly.

CHAPTER THREE

NOTES

[1] Recall the within estimator,

$$\hat{\beta}_W = (X'M_DX)^{-1}X'M_DY \; .$$

where $M_D = I - D(D'D)^{-1}D'$. Let $P_D = D(D'D)^{-1}D'$. The "between" estimator is defined as

$$\hat{\beta}_B = (X'P_DX)^{-1}X'P_DY.$$

The projection P_D transforms the data into individual means. So, $\hat{\beta}_B$ is derived from a regression of

$$\begin{vmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_N \end{vmatrix} \quad \text{on} \quad \begin{vmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_N \end{vmatrix} \; ,$$

and therefore uses variation across individuals.

Now, $\hat{\beta}_{GLS}$ may be expressed as a matrix weighted average of $\hat{\beta}_W$ and $\hat{\beta}_B$:

$$\hat{\beta}_{GLS} = (X'(M_D X + \theta^2 X' P_D X)^{-1} (X' M_D + \theta^2 X' P_D) Y.$$

where $\theta^2 = 1 - T\theta_1^2 = \frac{\sigma^2}{\sigma^2 + T\sigma^2_\alpha}$. For fixed T, the use of between variation by GLS results in an efficiency gain over within. But, $\theta^2 \to 0$ as $T \to \infty$ implying $\hat{\beta}_{GLS} = \hat{\beta}_W$ for large T.

[2] Except where the individual effects are correlated with all of the columns of X. In this case GLS = within (see Mundlak (1978)).

CHAPTER FOUR

NOTES

[1]Specification tests are outlined on pp. 1382-3 of H-T.

[2]Any vector orthogonal to a time-invariant vector can be used as an instrument. Since the time-invariant α_i are the only components of the disturbance which are correlated with an explanatory variable, M_D may be included as an instrument. As H-T note, the time-invariance of the α_i provides N(T-1) linearly dependent instruments for (4.2.3). (See H-T p. 1384-5).

[3]These identifying restrictions can be tested. See H-T, pp. 1388-9 for details.

[4]The matrix $\Omega^{-1/2}$ transform Ω into a scalar matrix; i.e. $\Omega^{-1/2}\Omega\Omega^{-1/2} = \sigma^2 I_{NT}$. When used to transform (4.2.3), $\Omega^{-1/2}$ yields a simple $(1-\Theta)$ differencing of the data,

$$Y_{it} - (1-\Theta)\bar{Y}_i = [X_{it}-(1-\Theta)\bar{X}_i]\beta + \theta Z_i\gamma + \Theta\alpha_i + (\varepsilon_t-(1-\theta)\bar{\varepsilon}_i].$$

OLS applied to this transformation is GLS (see note 2). This is computationally convenient.

[5]This procedure combines the computational convenience of the $\Omega^{-1/2}$ transformation with the simplifications provided by P_A. First, P_A applied to the exogenous variables yields the variables themselves. Secondly, the projection of the endogenous variables onto the column space of A can be derived by using only individual means (see H-T appendix B).

[6]The proofs of these results are given in Appendix A of H-T.

[7]Formally we assume $M_Q Z = 0$ (i.e., $M_i Z_i = 0, \forall_i$).

[8]Derivation of $\Omega^{-1/2}$ is a straightforward application of Wansbeek and Kapteyn (1982).

[9]Or, equivalently if $k_1 + j_1 + L > J + L$.

[10]Since $(\Omega^{-1/2}X_1, \Omega^{-1/2}Z_1, \Omega^{-1/2}W)$ is more highly correlated than $(X_1, Z_1\ W)$ with $\Omega^{-1/2}(X, Z, W)$.

[11]Since $\Omega^{-1/2}Z = \Theta Z$ and $\Omega^{-1/2}(Y-X\hat{\beta}_W) = \Theta(Y-X\hat{\beta}_W)$,

$$\hat{\gamma}_W = [Z;B(B'\Omega^{-1}B)^{-1}B'Z]^{-1}\ Z'B\ (B'\Omega^{-1}B)^{-1}B'(Y-X\hat{\beta}_W).$$

This is generally different from $\hat{\gamma}_W$ in (4.2.8); i.e., (4.2.8) uses

$$B'B = \begin{vmatrix} Z_1'X_1 & X_1'Z_1 \\ Z_1'X_1 & Z_1'Z_1 \end{vmatrix}$$

while (4.3.10) uses

$$B'\Omega^{-1}B \quad = \quad \begin{vmatrix} X_1'I(M_D+\Theta^2P_D)X_1 \\ \Theta^2Z_1'X_1 \end{vmatrix} \quad \begin{vmatrix} \Theta^2X_1'Z_1 \\ \Theta^2Z_1'Z_1 \end{vmatrix} \quad .$$

Incidently, $B'B \neq B'\Omega^{-1}B$ is another reason (one H-T do not mention) for $\hat{\gamma}_W \neq \hat{\gamma}^*$ in the over-identified case.

[12]Since B'Z is nonsingular when $k_1 = j_2$.

[13]Again, the identifying restrictions may be tested. See note 6.

[14]Compare with H-T Appendix A.

CHAPTER FIVE

NOTES

[1]An extensive literature exists on the estimation of union wage differentials. Two excellent surveys, discussing methodological issues as well as empirical findings, are Freeman and Niedoff (1981) and Lewis (1983).

[2]A random effects/GLS approach to this problem would not make sense for obvious reasons. However, at least one study, Chowdhury and Nickell (1985), has applied the H-T analysis to the question of unions' effect on wages.

[3]See Lewis (1983) for a critique of these studies.

[4]See Freeman (1980).

[5]The restricted version of Stewart's model can be estimated consistently by nonlinear least squares (i.e., searching over the values of λ).

[6]Percentage differentials are calculated by $(\exp(\hat{\delta})-1)100$.

[7]Percentage changes are calculated by differences in natural logarithm.

[8]The occupation dummies are also treated as endogenous.

[9]The rank condition is also fulfilled.

[10]Recall the H-T efficient estimator is equivalent to within in the exactlky identified case.

[11]See also Griliches (1977) and Griliches, Hall, and Hausman (1978).

[12]Union status is now a part of W in our general model (see (4.3.4)). While this may be intuitively unsatifying, allowing parts of W to be correlated with the efffects makes estimation of the model overly complicated.

REFERENCES

Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," Review of Economic Studies, 47, 225-238.

Chowdhury, G. and Nickell, S. (1985), "Hourly Earnings in the United States: Another Look at Unionization, Schooling, Sickness, and Unemployment Using PSID Data," Journal of Labor Economics, 3, 38-69.

Freeman, B. (1984), "Longitudinal Analysis of the Effects of Trade Unions," Journal of Labor Economics, 2, 1-2.

Griliches, Z., Hall, B., and Hausman, J. (1978), "Missing Data and Self-Selection in Large Panels," Annales de L'Insee, 30-31, 137-176.

Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica, 95, 1-22.

Hausman, J.A. and Taylor, W. (1981), "Panel Data and Unobservable Individual Effects," Econometrica, 14, 1377-1399.

Judge, G.G. (1985), The Theory and Practice of Econometrics, New York: Wiley.

Lewis, H.G. (1983), Union Relative Wage Effects: A Survey, Chapters 5 and 7, presented at the NBER Conference on the Economics of Trade Unions.

Mundlak, Y. (1978), "Models with Variable Coefficients: Integration and Extension," Annales de L'Insee, 30-31, 483-504.

Mundlak, Y. (1978), "On the Pooling of Time-Series and Cross-Section Data," Econometrica, 46, 69-86.

Rao, C.R. (1973), Linear Statistical Inference and Its Applications, New York: Wiley.

Stewart, M.B. (1983), "The Estimation of Union-Wage Differentials from Panel Data: The Problem of Not-So-Fixed Effects," Mimeo, Princeton University Industrial Relations Section.

Wansbeek, T. and Kapteyn, A. (1982), "A Class of Decompositions of the Variance-Covariance Matrix of a General Error Components Model," Econometrica, 50, 713-724.