UNIVERSITY OF SOUTHAMPTON

FACULTY OF HUMANITIES, ARTS & SOCIAL SCIENCES

School of Social Sciences

Unobservable Factors and Panel Data Sets

by

Eleonora Patacchini

Thesis for the degree of Doctor of Philosophy

January 2004

UNIVERSITY OF SOUTHAMPTON
ABSTRACT

FACULTY OF HUMANITIES, ARTS & SOCIAL SCIENCES

SCHOOL OF SOCIAL SCIENCES

Doctor of Philosophy

UNOBSERVABLE FACTORS AND PANEL DATA SETS

by Eleonora Patacchini

This thesis addresses statistical issues related to linear panel data models with the joint occurrence of unobserved heterogeneity and measurement errors-in-variables. Specifically, it is concerned with hypotheses testing and estimation techniques in a static and in a dynamic framework respectively.

Chapter 1 presents a methodological revision of the use of the Hausman test (Hausman, 1978) for correlated effects with panel data. The consequences of deviations from the basic assumptions underlying the construction of the Hausman statistic are investigated. In particular, the distribution of the Hausman statistic in cases of misspecification of the variance-covariance matrix of the errors is examined. It is shown that the size distortion may be serious. An alternative robust formulation of the test with panel data, based on the use of an auxiliary regression, is proposed. This test, which we call the Hausman Robust or HR-test, gives correct significance levels in common cases of misspecification of the variance-covariance matrix of the errors and has a power comparable to the standard Hausman test when no evidence of misspecification is present. It can be easily implemented using a standard econometric package, e.g. Stata.

In Chapter 2 this robust version of the Hausman test (suitably tailored) is used to compare different pairs of panel data estimators in a particular sequence. The resulting two-step testing procedure is intended to distinguish between an endogeneity problem caused by correlation between regressors and individual effects, and an endogeneity problem due to measurement errors. The statistical performance of the sequential test is assessed using simulated data. This methodology is then applied to an empirical job-search matching model to investigate the effects of measurement errors and unobserved heterogeneity that, as is well-known, contaminate two of the variables extensively used in labour market research, namely the stock of unemployed and the stock of vacancies. The economic implications of the inference results using the proposed methodology are compared with those produced by a possible traditional analysis.

Chapter 3 presents consistent estimators (which differ in terms of efficiency) for an autoregressive (stationary) model of panel data that superimposes the errors-in-variables problem and the unobserved heterogeneity issue on a dynamic framework. Moreover, the measurement errors are not 'classical' (i.e., uncorrelated with everything else in the model included their own past values) but are assumed to have a more complicated structure. The analysis of an example demonstrates the empirical relevance of this modelling. Furthermore, because the cross sectional units in the panel data set considered have a spatial connotation (UK counties), spatial features are also incorporated in the econometric analysis. The resulting empirical model is a spatio-temporal panel data model with unobserved heterogeneity and systematic measurement errors.

Contents

List of Figures

Preface

A panel data set contains repeated observations over the same units (individuals, households, firms...), collected over a number of periods. Since the pioneering papers by Kuh (1959), Hoch (1962) and Balestra and Nerlove (1966), the pooling of cross sections and time series data has become an increasingly popular way of quantifying economic relationships. Each dimension provides information lacking in the other, so a combination of both allows one to specify and estimate more complicated and more realistic models than a single cross section or a single time series would do. Panel data sets are mainly more informative data sets: they allow us to control for unobserved heterogeneity among individuals and furthermore they give more variability, less collinearity among the variables, more degrees of freedom and more efficiency than traditional time series or cross sectional studies (Baltagi, 1995). As a result, important advantages of panel data models compared to time series or cross sectional data models are that they allow estimation methods robust to certain classes of omitted variables and that they allow the identification of certain parameters of interest in presence of endogenous regressors or measurement errors, without the need to make restrictive assumptions. Furthermore, they can be used to study and compare the dynamics of different individuals.

The first aim of this thesis is to provide appropriate panel data techniques for dealing with different sources of unobservable factors. The failure to control for various types of unobservable factors may lead to unreliable estimation results.

> "There are at least three types of unobservable in econometric models: (*i*) fixed parameters to be estimated from the data, (*ii*) variables which affect the observable variables but which are not themselves directly observed either because the observed magnitudes are subject to measurement error or because these variables do not correspond directly to anything that is likely to be measured, (*iii*) disturbances, either as errors of measurement or as errors in equations. There is no clear dividing

line between these three categories."[1]

That data collection is subject to a variety of errors needs no reiteration or much documentation. In general, the data collection and thus the responsibility for the quality of the collected material, is still largely delegated to institutions outside the control of the analyzing team. It is also fair to note that part of the problem arises from our requirements about what we would like to observe and the complexity of the phenomena which we are trying to measure. Nevertheless, the role of unobservable factors has been relatively neglected in econometric practice.[2] The standard errors-in-variables models have not been applied widely mainly because to be identified they require extraneous information. This conclusion of lack of identification in errors-in-variables models, however, relates to uni-dimensional data, i.e. pure cross sections or pure time series. In presence of two-dimensional data, e.g. panel data, it may be possible to handle the errors-in-variables identification problem and estimate consistently the coefficients without extraneous information.

Many studies using panel data sets face the simultaneous problem of unobserved heterogeneity and measurement errors. The literature dealing specifically with panel data models capturing both features is not large and not yet well developed (see Matyas and Sevestre, 1996, Ch. 10 for a review). This thesis addresses statistical issues related to linear panel data models with the joint occurrence of unobserved heterogeneity and measurement errors-in-variables. Specifically, it is concerned with hypothesis testing and estimation techniques in a static and in a dynamic framework respectively.

The Thesis is organized in three chapters. Chapter 1 and 2 consider static models.

Chapter 1 (joint with R. O'Brien) presents a methodological revision of the use of the Hausman test for correlated effects with panel data. The application of the test, a common practice in applied work, does not always give reliable results. The assumptions underlying the construction of the Hausman statistic (Hausman, 1978) are too strong in

[1] Griliches (1979).

[2] For an extensive discussion, see Aigner et al. (1984).

many empirical cases.

The main contributions are the following ones.

- The consequences of deviations from the basic assumptions underlying the construction of the Hausman statistic are investigated. Within the discussion, it is shown that the assumptions in Lemma 2.1. in Hausman (1978) are sufficient but not necessary. In particular, it is demonstrated that the attainment of the absolute Fisher lower bound can be replaced by the attainment of a relative minimum variance bound.

- The distribution of the Hausman statistic in cases of misspecification of the variance-covariance matrix of the errors is examined. The size distortion is numerically assessed. It is found that the test will reject more often than allowed by its nominal size. Furthermore, in common cases of misspecification, it is shown analytically for the asymptotic case that the size distortion is sensitive to the ratio between the intra-groups and inter-groups variation of the covariates.

- An alternative robust formulation of the test in a panel data context is constructed. It is based on the use of the auxiliary regression proposed by Arellano (1993). The power is assessed using a simulation experiment. This test, which we call the *HR-test*, gives correct significance levels in common cases of misspecification of the variance-covariance matrix of the errors and has a power comparable to the standard Hausman test when no evidence of misspecification is present. It can be easily implemented using a standard econometric package.

The contribution of R. O'Brien is in the derivations contained in Section 1.4 and in Appendix 1.1.

Chapter 2 emphasizes the misleading inference results which one can obtain by testing for unobserved heterogeneity without conditioning on the existence or non existence of measurement errors and illustrates how panel data sets can be used to detect and treat properly different kinds of unobservable factors. A concrete case of study is presented.

We seek to investigate the importance of unobserved heterogeneity and measurement errors that contaminate two of the variables extensively used in labor market research, namely the stock of unemployed and the stock of vacancies.

The main contributions are the following ones.

- Robust formulations of the Hausman test (*HR-tests*) for the comparison of different pairs of panel data estimators are implemented. Chapter 1 shows how to construct a panel data artificial regression in order to get a robust test for correlated effects, i.e. for the presence of (group) individual-specific unobservable effects constant over time correlated with the regressors. It is based on the comparison of two particular estimators, i.e. the *Within Groups* and the *Between Groups*, as the *Within Groups* transformation removes unobservable factors of this sort. In Chapter 2 similar artificial regressions are constructed to compare other pairs of panel data estimators for different purposes.

- A sequential testing procedure is presented. It consists on using different *HR-tests* in a particular sequence. The aim is to distinguish between an endogeneity problem caused by correlation between regressors and individual effects, and an endogeneity problem due to measurement errors. This approach is founded on the idea that essential specification errors can be recognized by comparing estimators which behave differently if the assumptions of the model are satisfied or if some of them are not. The statistical performance of the sequential test is assessed using simulated data. Considerations on the significance level and power of the test are presented.

- The methodology is applied to an empirical job search matching model. We compare different panel data estimators of the coefficients of the stocks of unemployed and vacancies. The choice of appropriate instruments is discussed. The use of this procedure suggests what is the most reliable model specification to analyze the data at hand. The economic implications of the inference results using the proposed methodology are compared with those produced by a possible traditional analysis. It is shown to what extent conclusions lacking accuracy in the choice of the model specification may be misleading.

Chapter 3 focuses on the estimation of linear dynamic models when measurement errors and unobserved heterogeneity are jointly taken into consideration. Errors-in-variables models require the use of instrumental variables techniques in order to obtain consistent estimators and dynamic models complicate the estimation procedure because only predetermined instruments may be available. Furthermore, certain transformations typically used to purge the model from unobserved heterogeneity, such as first differences or deviation from time-means, lead to inconsistent estimators when instruments are predetermined (see, for instance, Arellano and Bond, 1991). Consistent estimators for (stationary) autoregressive panel data models with white noise errors (assuming exact measurement of the variables) are presented, among others, by Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998). Consistent estimators for static panel data models with measurement errors-in-variables are discussed, among others, by Biorn (2000). This chapter combines these two strands of the panel data literature.

The main contributions are the following.

- Consistent estimators (which differ in terms of efficiency) for an autoregressive model of panel data that superimposes the errors-in-variables problem and the unobserved heterogeneity issue on a dynamic framework are constructed. Moreover, the measurement errors are not "classical" (i.e., uncorrelated with everything else in the model included their own past values) but have a more complicated structure which invalidates the estimation techniques in the existing related literature. The empirical relevance of such a modelling is clear if we think about measurement errors not only as observation errors in the narrow sense but also as discrepancies between theoretical variable definitions and their observable counterparts in a wider sense. It seems sensible to assume that the difference between a typical variable of interest and its empirical counterpart almost never has a simple random structure.

- The proposed estimation techniques are applied in a concrete case of study. Because the cross sectional units in the panel data set considered have a spatial connotation (UK counties), spatial features are incorporated in the econometric analysis. The resulting empirical model is a spatio-temporal panel data model with unobserved heterogeneity and systematic measurement errors in variables. It is shown to what extent the illustrated estimation methodology for dynamic panel data models can be tailored and applied in order to obtain reliable results in the specific context analyzed.

Acknowledgments

My greatest thanks go to my supervisors Raymond O'Brien and Patricia Rice for their constant advice, support, and especially their endless patience and warm care.

I wish also to thank all the other members of the Department of Economics, in particular Anurag Banerjee, Grant Hillier, Robin Mason, Jan Podivinsky, Jackie Wahba, Yves Zenou, for their help, suggestions, encouragement and friendship. I take this opportunity to thank also Gill Crumplin, Claire Caffrey, Fatima Fauzel, and the other secretaries in the Department for being always ready to help with a smile.

I am grateful to all my friends that have contributed to make my last three years a wonderful experience, and in particular to Grazia Rapisarda, Emanuela Lotti, my office mate Mokhrazinim Mokhatar and my flat mates in Bencraft Court Halls of Residence.

Special thanks to my parents. They have relentlessly supported me.

Partial financial support from a postgraduate scholarship of the Department of Economics and from the University of Rome "La Sapienza" is gratefully acknowledged.

Last, but not least, I express my gratitude to Federico Martellosio for being close to me in every possible way.

Chapter 1

Testing the Exogeneity Assumption in Panel Data with "Non Classical" Disturbances

This chapter is concerned with the use of the Durbin-Wu-Hausman test for correlated effects with panel data. The assumptions underlying the construction of the statistic are too strong in many empirical cases. The consequences of deviations from the basic assumptions are investigated. The size distortion is assessed. In the case of measurement errors, the Hausman test is shown to be a test of the difference in asymptotic biases of *Between* and *Within Groups* estimators. However, its 'size' is sensitive to the relative magnitude of the intra-groups and inter-groups variations of the covariates, and can be so large as to preclude the use of the statistic. We show to what extent some assumptions can be relaxed in a panel data context and we discuss an alternative robust formulation of the test. Power considerations are presented.

1.1 Introduction

The Hausman test is the standard procedure used in empirical panel data analysis in order to discriminate between the fixed effects and random effects model. The general set up can be described as follows.[1]

Suppose that we have two estimators for a certain parameter θ of dimension $K \times 1$. One of them , $\widehat{\vartheta}_r$, is robust, i.e. consistent under both the null hypothesis H_0 and the alternative H_1, the other, $\widehat{\vartheta}_e$, is efficient and consistent under H_0 but inconsistent under H_1. The difference between the two is then used as the basis for testing. It can be shown (Hausman, 1978) that, under appropriate assumptions, under H_0 the statistic h based on $\left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right)$ has a limiting chi-squared distribution:

$$h = \left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right)' \left[\widehat{Var}\left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right)\right]^{-1} \left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right) \overset{a}{\sim} \chi^2_K.$$

If this statistic lies in the upper tail of the chi-squared distribution we reject H_0. If the variance matrix is consistently estimated, the test will have power against any alternative under which $\widehat{\vartheta}_r$ is robust and $\widehat{\vartheta}_e$ is not. Holly (1982) discusses the power in the context of maximum likelihood.

In a panel data context the test is typically used as a test for correlated effects.

Consider the model

$$y_{it} = x_{it}'\beta + \eta_i + v_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T \tag{1.1}$$

where x_{it} is a $K \times 1$ vector of stochastic regressors, $\eta_i \sim iid\left(0, \sigma^2_\eta\right)$, $v_{it} \sim iid\left(0, \sigma^2\right)$ uncorrelated with x_{it} and $Cov\left(\eta_i, v_{it}\right) = 0$.

The null hypothesis assumes lack of correlation between the individual effect η_i and

[1]This approach is also used by Durbin (1954) and Wu (1973). For this reason tests based on the comparison of two sets of parameter estimates are also called Durbin-Wu-Hausman tests, or DWH. For simplicity of exposition we will refer to the Hausman (1978) set up.

the explanatory variables x_{it},

$$H_0 : Cov(x_{it}, \eta_i) = 0.$$

The *Within Groups* estimator, $\widehat{\beta}_{wg}$, is robust regardless of the correlation between η_i and x_i. The *Balestra-Nerlove* estimator, $\widehat{\beta}_{BN}$, is efficient under H_0 but inconsistent under H_1,

$$H_1 : Cov(x_{it}, \eta_i) \neq 0.$$

The Hausman statistic in this case takes the form

$$h_1 = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)' \left[\widehat{Var}\left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)\right]^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right) \overset{a}{\sim} \chi^2_k. \tag{1.2}$$

However, using the results in Hausman (1978), the statistic used in practice is

$$h_2 = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)' \left(\widehat{V}_{wg} - \widehat{V}_{BN}\right)^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right), \tag{1.3}$$

where $V_{wg} = Var\left(\widehat{\beta}_{wg}\right)$ and $V_{BN} = Var\left(\widehat{\beta}_{BN}\right)$. It is based on the result that the variance of the difference between an estimator and an efficient estimator is equal to the differences of the variances:

$$Var\left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right) = V_{wg} - V_{BN}. \tag{1.4}$$

In the time series-cross section model considered in Hausman (1978) this equality holds because $\widehat{\beta}_{BN}$ is an efficient estimator in the sense that it attains the Cramér-Rao Lower Bound for fixed λ (defined below), and $Cov\left(\widehat{\beta}_{wg}, \widehat{\beta}_{BN}\right) = Var\left(\widehat{\beta}_{BN}\right)$. This implies

$$\begin{aligned} Var\left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right) &= Var\left(\widehat{\beta}_{wg}\right) + Var\left(\widehat{\beta}_{BN}\right) - 2Cov\left(\widehat{\beta}_{wg}, \widehat{\beta}_{BN}\right) \\ &= Var\left(\widehat{\beta}_{wg}\right) + Var\left(\widehat{\beta}_{BN}\right) - 2Var\left(\widehat{\beta}_{BN}\right) \\ &= Var\left(\widehat{\beta}_{wg}\right) - Var\left(\widehat{\beta}_{BN}\right) = V_{wg} - V_{BN}. \end{aligned}$$

However, in applied studies, this may not always be the case and one should be careful in using h_2 automatically. If equality (1.4) does not hold, h_2 does not follow an asymptotic chi-squared distribution, even under H_0.

This chapter considers the effects on the Hausman statistic used in applied panel data studies, h_2, of deviations from the conditions in Lemma 2.1 in Hausman (1978), which guarantees that equality (1.4) holds. This lemma is stated as follows.

Lemma 1 Consider two estimators $\widehat{\beta}_0$, $\widehat{\beta}_1$ which are both consistent and asymptotically normally distributed with $\widehat{\beta}_0$ attaining the asymptotic Cramér-Rao bound so that $\sqrt{T}\left(\widehat{\beta}_0 - \beta\right) \overset{a}{\sim} N\left(0, V_0\right)$ and $\sqrt{T}\left(\widehat{\beta}_1 - \beta\right) \overset{a}{\sim} N\left(0, V_1\right)$ where V_0 is the inverse of Fisher's information matrix. Consider $\widehat{q} = \widehat{\beta}_1 - \widehat{\beta}_0$. Then the limiting distributions of $\sqrt{T}\left(\widehat{\beta}_0 - \beta\right)$ and $\sqrt{T}\widehat{q}$ have zero covariance, $Cov\left(\widehat{\beta}_0, \widehat{q}\right) = 0$, a null matrix.

The plan of the chapter is as follows. Regarding the attainment of the Cramér-Rao Lower Bound, in Section 1.2 we show that if we want to compare different estimators within a specific set, the assumption of full efficiency is not necessary. A relative lower bound for the variance can play the role. The variance of the difference between two estimators belonging to such a set is still equal to the difference of the variances if one of the two is the minimum variance estimator in the specific set considered. The algebraic derivation of this result is provided in the panel data framework. The Lemmas contained in Appendix 1.1 prove that this holds both in the exact and in the limiting case. Given that the *Balestra-Nerlove* estimator can be obtained as a matrix weighted average of the *Between Groups*, $\widehat{\beta}_{bg}$, and the *Within Groups* estimators (see, e.g. Maddala, 1971), we consider the set of estimators which is defined by a matrix weighted average of two unbiased (or consistent in the limiting case) estimators.

However, even the attainment of a minimum variance bound may be a strong assumption in empirical studies. This circumstance is related to assumptions about the error term. A failure of the assumption of spherical disturbances is quite common in practice. Section 1.3 presents a robust formulation of the Hausman test for correlated effects, which is based on the construction of an auxiliary regression. We explain and discuss to what extent the use of artificial regressions may allow us to construct tests based on the difference between two estimators in a panel data model without making strong assumptions about the disturbances. The motivation underlying the implementation of the robust test is that the size distortion of the standard Hausman test, h_2, in cases of misspecification of the variance-covariance matrix of the disturbances may be serious. This is investigated in Section 1.4.

Section 1.5 compares the power of the standard Hausman test and the robust formulation presented in Section 1.3 using a Monte Carlo experiment. Section 1.6 concludes.

1.2 The Failure of the Assumption of Full Efficiency

Consider model (1.1). Defining the disturbance term

$$\varepsilon_{it} = \eta_i + v_{it},$$

the variance-covariance matrix of the errors is

$$\underset{(NT\times NT)}{\Sigma} = I_N \otimes \Omega$$

where

$$\Omega = \begin{pmatrix} \sigma_\eta^2 + \sigma^2 & \dots & \sigma_\eta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \dots & \sigma_\eta^2 + \sigma^2 \end{pmatrix} = \sigma^2 I_T + \sigma_\eta^2 \, \iota\iota' \tag{1.5}$$

and ι is a column vector of T ones. The unobserved heterogeneity implies correlation over time for single units, but there is no correlation across units.

Hausman and Taylor (1981) propose three different specification tests for the null hypothesis of uncorrelated effects: one based on the difference between the *Within Groups* and the *Balestra-Nerlove* estimator, another on the difference between the *Balestra-*

Nerlove and the *Between Groups* and a third on the difference between the *Within Groups* and the *Between Groups.* They show that the chi-squared statistics for the three tests are identical. We now analyze the Hausman statistic constructed on the difference between the *Within Groups* and the *Balestra-Nerlove* estimator, commonly used in empirical work.

Hereafter, we define as fully efficient an estimator that reaches the Cramér-Rao Lower Bound and as minimum variance the one that has the minimum variance within a specific class. Let

$$\lambda = \frac{\sigma^2}{\sigma^2 + T\sigma_\eta^2}.$$

If we assume normality in model (1.1), it is well-known that the *Balestra-Nerlove* estimator, i.e. the generalized least squares estimator, is fully efficient if the variance-ratio parameter λ is known, and asymptotically fully efficient if λ is consistently estimated. (A distributional assumption is required in order to obtain the Cramér-Rao Bound.) Therefore the hypotheses underlying the construction of the Hausman statistic are satisfied and the results of the test are reliable. However, we will demonstrate that even without assuming normality of the ε_{it} the results of the standard Hausman test are reliable, the key assumption being (1.5). We will use the panel data framework as an example. In what follows we take λ as known. The same result holds asymptotically if a consistent estimator $\widehat{\lambda}$ is available. It is implied by the Hausman-Taylor result that we can construct the same test using different pairs of estimators, as will be clarified below.

Consider model (1.1). We write the *Balestra-Nerlove* estimator (Balestra and Nerlove, 1966) as a function of the variables in levels

$$\widehat{\beta}_{BN} = \left(X^{'}QX + \lambda X^{'}MX\right)^{-1}\left(X^{'}Q + \lambda X^{'}M\right)Y \tag{1.6}$$

where

$$Q \;=\; I_N \otimes Q^{+},$$

$$
\begin{aligned}
Q^+ &= I_T - \frac{1}{T}ii', \\
M &= I_N \otimes M^+, \\
M^+ &= \frac{1}{T}ii' = I_T - Q^+,
\end{aligned}
$$

$$
X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}, \; Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \; X_i = \begin{bmatrix} x'_{i1} \\ x'_{i2} \\ \vdots \\ x'_{iT} \end{bmatrix}, \; y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}.
$$

Q^+ is the matrix that transforms the data to deviations from the individual time means, M^+ is the matrix that transforms the data to averages. Rearranging

$$
\widehat{\beta}_{BN} = \left[X' \left[\lambda I_{NT} + (1-\lambda)\, Q\right] X\right]^{-1} X' \left[\lambda I_{NT} + (1-\lambda)\, Q\right] Y. \tag{1.7}
$$

The variance is

$$
\begin{aligned}
Var(\widehat{\beta}_{BN}) &= \left\{\left[X' \left[\lambda I_{NT} + (1-\lambda)\, Q\right] X\right]^{-1} X' \left[\lambda I_{NT} + (1-\lambda)\, Q\right]\right\} Var(Y) \\
&\quad \times \left\{\left[\lambda I_{NT} + (1-\lambda)\, Q\right] X \left[X' \left[\lambda I_{NT} + (1-\lambda)\, Q\right] X\right]^{-1}\right\}.
\end{aligned} \tag{1.8}
$$

Using a simplified version of the Sherman-Morrison-Woodbury formula (Golub and Van Loan, 1983, p.50) one can show (Appendix 2.1) that, under assumption (1.5), the variance of y_i can be written as

$$
\begin{aligned}
Var(y_i) &= \sigma^2 \left[I_T - \frac{\sigma_\eta^2}{\sigma^2 + T\sigma_\eta^2}\iota\iota'\right]^{-1} = \sigma^2 \left[I_T - \frac{1}{T}(1-\lambda)\iota\iota'\right]^{-1} \\
&= \sigma^2 \left[\left(I_T - \frac{1}{T}\iota\iota'\right) + \lambda \frac{1}{T}\iota\iota'\right]^{-1}.
\end{aligned}
$$

This can also be obtained by ignoring time effects, and thus setting $\omega = 0$, in Nerlove

(1971). Using the matrices involved in formula (1.6), we can rewrite this expression as

$$
\begin{aligned}
Var(y_i) &= \sigma^2 \left[\left(I_T - M^+\right) + \lambda M^+\right]^{-1} \\
&= \sigma^2 \left[Q^+ + \lambda I_T - \lambda Q^+\right]^{-1} \\
&= \sigma^2 \left[\lambda I_T + (1-\lambda) Q^+\right]^{-1}.
\end{aligned} \tag{1.9}
$$

Thus

$$
Var(Y) = I_N \otimes Var(y_i) = \sigma^2 \left[\lambda I_{NT} + (1-\lambda) Q\right]^{-1}.
$$

Substituting (1.9) in (1.8), we obtain

$$
\begin{aligned}
& Var(\widehat{\beta}_{BN}) \\
&= \sigma^2 \left[X' \left[\lambda I_{NT} + (1-\lambda) Q\right] X\right]^{-1} X' \left[\lambda I_{NT} + (1-\lambda) Q\right] \left[\lambda I_{NT} + (1-\lambda) Q\right]^{-1} \\
&\quad \times \left[\lambda I_{NT} + (1-\lambda) Q\right] X \left[X' \left[\lambda I_{NT} + (1-\lambda) Q\right] X\right]^{-1} \\
&= \sigma^2 \left[X' \left[\lambda I_{NT} + (1-\lambda) Q\right] X\right]^{-1}.
\end{aligned} \tag{1.10}
$$

Similarly, using the Q matrix defined in formula (1.6), we can write also the *Within Groups* estimator as a function of the initial variables in levels

$$
\widehat{\beta}_{wg} = \left[X' Q X\right]^{-1} X' Q Y. \tag{1.11}
$$

The variance is

$$
Var(\widehat{\beta}_{wg}) = \left[X' Q X\right]^{-1} X' Q (Var Y) Q' X \left[X' Q X\right]^{-1}. \tag{1.12}
$$

If we transform the data into deviations, the variance of y_i can be written as

$$
Var(Q^+ y_i) = Q^+ Var(y_i) Q^{+'} = \sigma^2 Q^+ \left[I_T + \theta \iota \iota'\right] Q^+ = \sigma^2 Q^+ Q^+ = \sigma^2 Q^+ \tag{1.13}
$$

where $\theta = \sigma_\eta^2/\sigma^2$ and $Q^+\iota = 0$, a vector of zeros. Thus

$$Var(QY) = \sigma^2 I_N \otimes Q^+ = \sigma^2 Q.$$

Substituting (1.13) in (1.12), we obtain[2]

$$\begin{aligned} Var(\widehat{\beta}_{wg}) &= \sigma^2 \left[X'QX\right]^{-1} X'QQQ'X \left[X'QX\right]^{-1} \\ &= \sigma^2 \left[X'QX\right]^{-1}. \end{aligned} \tag{1.14}$$

Hence, from (1.10) and (1.14)

$$Var(\widehat{\beta}_{wg}) - Var(\widehat{\beta}_{BN}) = \sigma^2 \left\{ \left[X'QX\right]^{-1} - \left[X' \left[\lambda I_{NT} + (1-\lambda) Q\right] X\right]^{-1} \right\}. \tag{1.15}$$

Next, we show that such expression is exactly equal to the variance of the difference between the two estimators.

$$Var(\widehat{\beta}_{BN} - \widehat{\beta}_{wg}) = Var(\widehat{\beta}_{BN}) - Cov(\widehat{\beta}_{BN}, \widehat{\beta}_{wg}) - Cov(\widehat{\beta}_{wg}, \widehat{\beta}_{BN}) + Var(\widehat{\beta}_{wg}).$$

From (1.7) and (1.11)

$$\begin{aligned} Cov(\widehat{\beta}_{BN}, \widehat{\beta}_{wg}) &= \sigma^2 \left[X' \left[\lambda I_{NT} + (1-\lambda) Q\right] X\right]^{-1} X' \left[\lambda I_{NT} + (1-\lambda) Q\right] \\ &\quad \times \left[\lambda I_{NT} + (1-\lambda) Q\right]^{-1} QX \left[X'QX\right]^{-1} \\ &= \sigma^2 \left[X' \left[\lambda I_{NT} + (1-\lambda) Q\right] X\right]^{-1} = Var(\widehat{\beta}_{BN}). \end{aligned}$$

This is symmetric, and thus equal to $Cov(\widehat{\beta}_{wg}, \widehat{\beta}_{BN})$. Therefore, we obtain

$$\begin{aligned} Var(\widehat{\beta}_{BN} - \widehat{\beta}_{wg}) &= Var(\widehat{\beta}_{BN}) - Var(\widehat{\beta}_{BN}) - Var(\widehat{\beta}_{BN}) + Var(\widehat{\beta}_{wg}) \\ &= Var(\widehat{\beta}_{wg}) - Var(\widehat{\beta}_{BN}) \end{aligned}$$

[2] Recall that Q is an idempotent matrix.

as required. We have proved that equality (1.4) holds for λ known or otherwise fixed.

As we said, the case of estimated λ can be treated by using the Hausman-Taylor result that an algebraically identical test statistic can be constructed using the difference between $\widehat{\beta}_{wg}$ and the *Between Groups* estimator $\widehat{\beta}_{bg}$. We obtain

$$(\widehat{\beta}_{wg} - \widehat{\beta}_{bg})' \left[Var(\widehat{\beta}_{wg}) + Var(\widehat{\beta}_{bg}) \right]^{-1} (\widehat{\beta}_{wg} - \widehat{\beta}_{bg})$$

as the estimators have zero covariance. In this form, we can see that estimating σ^2 and λ (or σ_η^2) affects only the variance matrix of the test statistic. We thus obtain the same test statistic whatever λ is, and (1.3) maintains the assumed distribution. It does not follow from these arguments that the equality (1.4) can be made exact for estimated λ. If $\widehat{\beta}_{BN}$ and $\widehat{\beta}_{wg}$ were independent of $\widehat{\lambda}$, the result would follow, but this requires normality of the disturbances. Viewing $\widehat{\beta}_{BN}$ as a feasible *GLS* estimator, Kakwani (1967) implies it is unbiased. However, conditional on $\widehat{\lambda}$ it may or may not be unbiased. Further, the variances obtained are for λ fixed, not conditional on $\widehat{\lambda}$. So attempts to obtain unconditional variances from conditional variances and variances of conditional expectations do not seem fruitful. So it would appear that the exact result (1.4) may require normality of the ε_{it} or λ fixed. Equality (1.4) implies that for fixed and known λ, and known σ^2, under normality h_2 would have an exact chi-squared distribution. If λ is estimated, and/or the ε_{it} are not normal, h_2 is asymptotically chi-squared as long as x_{it} are sufficiently well-behaved to ensure that $\widehat{\beta}_{BN}$ and $\widehat{\beta}_{wg}$ are asymptotically normal, and σ^2 and σ_η^2 (or equivalently λ) are appropriately estimated. This is less restrictive than the assumptions required for the identification of the Cramér-Rao bound. We obtain the result (1.4) without assuming normality because we compare two linear unbiased estimators, one of them achieving the minimum variance for a linear estimator. Lemma 4 in Appendix 1.1 shows that the variance result depends only on minimum variance properties, not on normality or achievement of a particular (Cramér-Rao) bound. However, in order to get a panel data generalized version of Lemma 1 (Lemma 2.1 in Hausman, 1978), it is necessary to prove a similar result in the limiting case. This aim is achieved in Lemma 10 in Appendix 1.1. The minimum variance property required is within a set of the form

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}$$

where t_1 and t_2 are estimators of the parameter vector θ. For completeness, Lemma 9 establishes that sets of this form will contain minimum variance members.

We can summarize as follows. If we want to use the Hausman statistic to compare two different estimators, e.g. one linear and one non linear, the assumption of normality may be crucial because it allows us to find an absolute lower bound for the variance of the estimators. However, if we want to compare different estimators within a set of the form of $\mathcal{T}$ neither the assumption of normality nor the attainment of the Cramér-Rao Lower Bound, even in the limiting case, is crucial. A lower bound for the variance can play the required role. The variance of the difference between two estimators belonging to the same set is still equal to the difference of the variances if one of the two is the minimum variance estimator in the specific set. Lemma 10 in Appendix 1.1 allows us to rely on the results provided by a traditional Hausman test in a more general set-up.

It is worth noting that we are not removing the assumption of asymptotic normality of the estimators in Lemma 1, which is needed to obtain the chi-squared distribution of the Hausman statistic. Our generalization applies for estimators that are asymptotically normally distributed but that do not reach the Cramér-Rao Bound.

We prove the result for a specific set of estimators but this does not rule out the possibility of extending the result to wider contexts. For instance, the GMM estimator is asymptotically normally distributed and attains the asymptotic Cramér-Rao Lower Bound only in some cases. Nevertheless, if we compare an arbitrary GMM estimator, e.g. using the identity matrix, and the one which uses the optimal weighting matrix (Hansen, 1982), Lemma 10 implies that the difference between these two estimators can be used as basis for an Hausman test.

1.3 The Failure of the Assumption of Spherical Disturbances

In the previous section, we relaxed the assumption of full efficiency in Lemma 1. However, even the assumption that one of the two estimators has the minimum variance or that both are consistent under the null hypothesis can be still too strong in many empirical cases. In the panel data framework above considered (model (1.1)), the crucial assumption for (1.4) to hold is (1.5). In other words, the form of the covariance matrix has to be assumed. In cases of misspecification, i.e. if $Var(y_i) = \Omega^* \neq \Omega$, equality (1.4) does not hold any longer.

As Hausman clearly states at the very beginning of his article (Hausman, 1978), the specification test he presents assumes that the disturbances have a spherical covariance matrix. He considers the standard regression framework

$$y = X\beta + \varepsilon,$$

where

$$E(\varepsilon/X) = 0, \tag{1.16}$$

and

$$Var(\varepsilon/X) = \sigma^2 I. \tag{1.17}$$

In most of the articles that followed, assumption (1.17) is never relaxed. The emphasis of this part of literature is placed in testing the orthogonality assumption, i.e. $E(\varepsilon/X) = 0$. In the panel data framework ((model (1.1)) a test of the assumption (1.16) tests for correlated effects. Also in this context the assumption (1.17) is usually maintained.

The reason is straightforward if we consider the comparison between the *Within Groups* estimator and the *Balestra-Nerlove* estimator as a comparison between an *OLS* and a *GLS* estimator. One basic assumption in the construction of the Hausman statistic (Lemma 1) is that one of the two estimators has to reach the asymptotic Cramér-Rao

Lower Bound or, using the generalization provided in Lemma 10 in Appendix 1.1, that at least has to be the minimum variance estimator in a specific class. In the panel data framework the *Balestra-Nerlove*, that is the generalized least square estimator, is the BLUE estimator if the *GLS* transformation produces spherical disturbances. This is the case if the correlation in the covariance matrix of the initial errors is due only to the presence of unobserved individual effects, i.e. if the initial disturbances are spherical.

To make it clear, we analyze in detail the construction of the *Balestra-Nerlove* estimator. In practice, the *Balestra-Nerlove* estimator can be calculated running an *OLS* regression on a transformed model. Assuming model (1.1), which implies the disturbances variance covariance matrix (1.5), the transformation of the y_i and the x_i is the following

$$\Omega^{-\frac{1}{2}} y_i = \begin{bmatrix} y_{i1} - \theta \overline{y}_{i.} \\ y_{i2} - \theta \overline{y}_{i.} \\ \vdots \\ y_{iT} - \theta \overline{y}_{i.} \end{bmatrix}$$

where $\overline{y}_{i.}$ is the individual i time mean,

$$\Omega^{-\frac{1}{2}} = I - \frac{\theta}{T} ii', \quad \theta = 1 - \frac{\sigma}{\left(\sigma^2 + T\sigma_\eta^2\right)^{\frac{1}{2}}}$$

and likewise for the rows of x_i.

Under assumption (1.5), which implies initial spherical disturbances, this is a *GLS* transformation that produces a model with spherical disturbances. Hence running *OLS* on such a model we obtain the BLUE estimator. However, if assumption (1.5) does not hold, the *GLS* transformation does not guarantee that the new disturbances are spherical. In this case the *GLS* estimator, namely the *Balestra-Nerlove*, is still consistent but it may not be the minimum variance estimator. The consequence is that we can no longer be sure that the equality (1.4) still holds. In these circumstance the results of the test h_2 (statistic (1.3)) may not be reliable. However, if the two estimators remain consistent under H_0 the comparison can still be conducted, but the methodology needs to be adjusted in an appropriate way.

In what follows, we present a robust version of the Hausman test for panel data. It is based on the use of an artificial regression. It allows us to compare different estimators without assuming normality or ranking them in terms of efficiency. Specifically, such methodology does not use the hypothesis that the variance of the difference of the two estimators is equal to the difference of the variances. It estimates directly the variance of the difference of the two estimators. It simply uses the statistic (1.2) instead of (1.3). Moreover, it provides an estimate of this variance that is consistent and robust to heteroskedasticity and/or serial correlation of arbitrary form in the within groups covariance matrix of the disturbances. This estimator is obtained using White's formulae (White, 1984). It will be made clear to what extent the application of White's heteroskedasticity consistent estimators of covariance matrices in a panel data framework may also allow for the presence of autocorrelation within groups.

Baltagi (1996) and Ahn and Lo (1996) propose different artificial regressions to test for the presence of correlated effects. However, the assumption of initial spherical disturbances is never relaxed. As shown by Baltagi (1997, 1998), under the assumption of spherical disturbances, the three approaches, i.e. the Hausman specification test, Baltagi (1996) and Ahn and Lo (1996), yield exactly the same test statistic. However, as first noted by Arellano (1993) in the same panel data framework, an auxiliary regression can also be used to obtain a generalized test for correlated effects which is robust to heteroskedasticity and correlation of arbitrary forms in the within groups disturbances. Davidson and MacKinnon (1993) list at least five different uses of artificial regressions including the calculation of estimated covariances matrices. We will use this device to estimate directly the variance between the two estimators without using equality (1.4). Furthermore, the application of White's formulae (White, 1984) in the panel data case will lead to heteroskedasticity and autocorrelation consistent estimators of such variance. Therefore, we can use an artificial regression to construct a test for the comparison of different pairs of estimators which is robust to deviations from the assumption of spherical disturbances. From now on we will call this technique the *HR-test*, for Hausman-Robust test.

Next we present the auxiliary regression that was proposed by Arellano (1993) to test for random versus fixed effects in a static panel data model.

Consider the panel data model[3]

$$\underset{(T\times 1)}{y_i} = \underset{(T\times K)}{X_i}\ \beta + \underset{(T\times 1)}{\mu_i}, \qquad i = 1, ..., N. \tag{1.18}$$

This system of T equations in levels can be transformed into $(T-1)$ equations in deviations and one in averages. We obtain respectively

$$\begin{cases} y_i^* = X_i^* \beta + \mu_i^* \longrightarrow (T-1) \text{ equations} \\ \overline{y_i} = \overline{X_i}\beta + \overline{\mu_i} \longrightarrow \quad 1 \ \text{ equation.} \end{cases}$$

Estimating by *OLS* the $N(T-1)$ equations in deviations from individual time-means we obtain the *Within Groups* estimator, i.e. $\widehat{\beta}_{wg}$. Estimating by *OLS* the N average equations we obtain the *Between Groups* estimator, i.e. $\widehat{\beta}_{bg}$.

Let

$$\beta_{wg} = plim\left(\widehat{\beta}_{wg}\right)$$

and

$$\beta_{bg} = plim\left(\widehat{\beta}_{bg}\right).$$

Rewrite the system as

$$\begin{cases} y_i^* = X_i^* \beta_{wg} + \mu_i^* - X_i^* \beta_{bg} + X_i^* \beta_{bg} \\ \overline{y_i} = \overline{X_i}\beta_{bg} + \overline{\mu_i}. \end{cases} \tag{1.19}$$

[3]For simplicity of exposition, we exclude the case when any time-invariant covariates are included.

Rearranging, we obtain

$$\begin{cases} y_i^* = X_i^* \left(\beta_{wg} - \beta_{bg}\right) + X_i^* \beta_{bg} + \mu_i^* \\ \overline{y_i} = \overline{X_i} \beta_{bg} + \overline{\mu_i}. \end{cases}$$

Call

$$Y_i^+ = \begin{pmatrix} y_i^* \\ \overline{y_i} \end{pmatrix}, W_i^+ = \begin{pmatrix} X_i^* & X_i^* \\ 0 & \overline{X_i} \end{pmatrix},$$

$$\beta^+ = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_{wg} - \beta_{bg} \\ \beta_{bg} \end{pmatrix}, \mu_i^+ = \begin{pmatrix} \mu_i^* \\ \overline{\mu_i} \end{pmatrix}.$$

The augmented auxiliary model is

$$Y_i^+ = W_i^+ \beta^+ + \mu_i^+, \qquad i = 1, ..., N. \tag{1.20}$$

If we estimate β^+ in (1.20) by *OLS*, we obtain directly the variance of the difference of the two estimators in the upper left part of the variance-covariance matrix of β^+. Under the assumption of spherical disturbances a Wald test on appropriate coefficients in the auxiliary regressions is equivalent to the standard Hausman test (Arellano, 1993). Instead, in this chapter, by estimating the variance-covariance matrix of β^+ using the White's formulae we obtain a formulation of the Hausman test robust to deviations from the assumption of spherical disturbances. Appendix 4.1 provides an analytical derivation of this result. The following Lemma is proved.

Lemma 2 Given model (1.20),

$$\widehat{\beta}_1 = \widehat{\beta}_{wg} - \widehat{\beta}_{bg}, \tag{1.21}$$

$$Var(\widehat{\beta}_1) = Var\left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right), \tag{1.22}$$

An appropriate estimator $\widehat{Var}(\widehat{\beta}_1)$ *consistently estimates* $Var(\widehat{\beta}_1)$. (1.23)

It is shown that, in order to get a consistent estimate of the variance, the first set of equations in system (1.19) has to be scaled.[4]

In what follows, we will clarify to what extent an application of White's formulae for estimators of covariances matrices (White, 1984) in a panel data context provides a consistent estimator which is robust to heteroskedasticity and arbitrary correlation in the covariance matrix of the random disturbances. It may also control for the presence of fixed effects. This latter possibility may be accommodated if we make further assumptions, i.e. cross-sectional heteroskedasticity which takes on a finite number of different values.

Consider a simple panel data framework without individual effects

$$\begin{aligned} y_{i1} &= \beta x_{i1} + \varepsilon_{i1} \\ y_{i2} &= \beta x_{i2} + \varepsilon_{i2}, \\ &\vdots \\ y_{iT} &= \beta x_{iT} + \varepsilon_{iT}, \qquad i = 1, ..., N, \end{aligned}$$

where

$$E(\epsilon_i \epsilon_i^{'}) = \begin{pmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 I_T = \Sigma, \qquad \epsilon_i = \underset{(T\times 1)}{\begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}}.$$

Assume that in the complete model

$$\underset{(NT\times NT)}{\Omega} = I \otimes \Sigma = \begin{pmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma \end{pmatrix}. \tag{1.24}$$

[4]The use of an artificial regression does require $Var(y_i)$ to be constant over $i = 1, .., N$, and diagonal, so only heteroscedasticity over time is protected against. More general cases require separate estimation of $Var(\widehat{\beta}_{wg})$ and $Var(\widehat{\beta}_{bg})$.

Define

$$X_i = \underset{(T\times 1)}{\begin{pmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{pmatrix}}, \qquad y_i = \underset{(T\times 1)}{\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}}$$

and rewrite the model as

$$\underset{(T\times 1)}{y_i} = \underset{(T\times 1)}{X_i\beta} + \underset{(T\times 1)}{\epsilon_i}, \qquad i = 1, ..., N. \tag{1.25}$$

This formulation allows us to consider panel data in the framework defined in White (1984). If we assume no cross-sectional correlation and $N \longrightarrow \infty$, all the hypotheses underlying the derivation of White's results are satisfied. Hence, Proposition 7.2 in White (1984, p. 165) applies.

$$\widehat{\Sigma} = N^{-1} \sum_{i=1}^{N} \widehat{\epsilon}_i \widehat{\epsilon}_i' \xrightarrow{p} \Sigma \tag{1.26}$$

and

$$\widehat{\Omega} = I \otimes \widehat{\Sigma} \xrightarrow{p} \Omega.$$

However, while with uni-dimensional data sets we obtain heteroskedasticity consistent estimators because ϵ_i is a scalar, in the two dimensional case ϵ_i is a vector and we obtain a consistent estimator of the whole matrix Σ. Hence, by applying the result (1.26) in the panel data case we obtain a consistent estimator of the variance covariance matrix of the disturbances that also allows for the presence of dynamic effects within groups.

Therefore, the estimators of the variance of the *OLS* estimators of β in the panel data model (1.25) can be obtained by

$$\widehat{Var(\beta)} = \left[\sum_{i=1}^{N} \left(X_i' X_i\right)\right]^{-1} \sum_{i=1}^{N} X_i' \widehat{\Omega} X_i \left[\sum_{i=1}^{N} \left(X_i' X_i\right)\right]^{-1}. \tag{1.27}$$

As stated by Arellano (1993), they are heteroskedasticity and autocorrelation consistent.

Such estimators are the ones used in the implementation of the *HR-test.* This case is referred in White (1984) as *contemporaneous covariance estimation.*

However, White (1984) also implements consistent estimators in another case that takes explicitly into consideration a grouping structure of the data. Consider again the panel data model (1.25). Replace assumption (1.24) by

$$\underset{(NT\times NT)}{\Omega} = \begin{pmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_N \end{pmatrix}.$$

In this context, in a slightly different notation from that used by White (1984, p.172-173), suitable for the panel data framework, we can obtain consistent estimators of the covariance matrix Ω using

$$\widehat{\Omega} = diag(\widehat{\Sigma}_1, \widehat{\Sigma}_{2,} \dots \widehat{\Sigma}_N)$$

where

$$\widehat{\Sigma}_i = T^{-1}\widehat{\epsilon}_i\widehat{\epsilon}_i{}'.$$

In other words, a consistent estimator (when $N \to \infty$) for the covariance matrix of group i is constructed by averaging the group residuals over only the observations in group i. In the balanced panel data case, their number is constant between groups and equal to T. This estimator is not only robust to autocorrelation of arbitrary form within groups but it also allows for the possibility that individual error covariance matrices may differ according to observable characteristics (such as region, union, race, etc....).

1.4 The Size of the Test

In this section we investigate the size distortion which occurs in the use of the standard Hausman test (statistic h_2, formulation (1.3)) when the basic assumptions (Lemma 1) are not satisfied.

Consider the panel data model (1.1) presented in Section 1.1. The Hausman test investigates the presence of specification errors of the form $Cov(x_{it}, \eta_i) \neq 0$. The robust version proposed in Section 1.3 tests such orthogonality assumption between explanatory variables and disturbances in presence of other forms of misspecification. In particular we are interested in a possible misspecification in the variance-covariance matrix of the disturbances arising, for instance, from the presence of measurement errors in variables. This case may be the rule rather than the exception in applied studies.

We want to test the hypothesis

$$H_o : Cov(x_{it}, \eta_i) = 0 \tag{1.28}$$

against the alternative

$$H_1 : Cov(x_{it}, \eta_i) \neq 0,$$

when

$$Var(\varepsilon_i | x_{it}) \neq \Omega, \tag{1.29}$$

Ω defined in (1.5).

Hausman (1978) shows that under H_o the test statistic

$$h = \widehat{q}'\widehat{V}(\widehat{q})^{-1}\widehat{q} \sim \chi^2_k \tag{1.30}$$

where $V(\widehat{q})$ is the asymptotic variance of q, and k is the length of q. The same test statistic is obtained if we consider the vector $\widehat{q}$ equal to

$$\begin{aligned} \widehat{q}_1 &= (\widehat{\beta}_{wg} - \widehat{\beta}_{BN}), \\ \text{or } \widehat{q}_2 &= (\widehat{\beta}_{bg} - \widehat{\beta}_{BN}), \\ \text{or } \widehat{q}_3 &= (\widehat{\beta}_{wg} - \widehat{\beta}_{bg}). \end{aligned}$$

As Hausman and Taylor (1981) pointed out they are all nonsingular transformations of one another. The estimate of the variance covariance matrix used in the three cases is

$$\begin{aligned}
\widehat{V}(\widehat{q}_1) &= \widehat{V}(\widehat{\beta}_{wg}) - \widehat{V}(\widehat{\beta}_{BN}), \\
\text{or } \widehat{V}(\widehat{q}_2) &= \widehat{V}\widehat{\beta}_{bg}) - \widehat{V}(\widehat{\beta}_{BN}), \\
\text{or } \widehat{V}(\widehat{q}_3) &= \widehat{V}(\widehat{\beta}_{wg}) + \widehat{V}(\widehat{\beta}_{bg}).
\end{aligned}$$

If we are in presence of misspecification of the form (1.29), none of the above expressions gives a consistent estimator of the variance-covariance matrix, even under H_o. The distribution of the test statistic under H_o needs to be investigated. The nominal size may be quite different from the observed one.

To investigate the size distortion under normality, we use the distributions of quadratic forms in normal random variables.[5] In particular, we use the following Lemma.[6]

Lemma 3 *(in Lemma 3.2 in Vuong, 1989). Let $x \sim N_K(0, V)$, with rank $(V) \leq K$, and let A be an $K \times K$ symmetric matrix. Then the random variable $x'Ax$ is distributed as a weighted sum of chi-squares with parameters (K, γ), where γ is the vector of eigenvalues of AV.*

This implies that $x'Ax$ is a chi-squared statistic with r degrees of freedom, where $r = rank(A)$, if and only if AV is idempotent (Muirhead, 1982, Theorem 1.4.5).

If $A = V^{-1}$, i.e. in cases of no misspecification, AV is idempotent. The theorem is satisfied and result (1.30) holds. The test statistic gives correct significance levels.

If $A \neq V^{-1}$ but AV is idempotent then $rank\,(A) < K$ and/or $rank\,(V) < K$ but still (1.30) holds. We omit this case for simplicity of exposition.

If $A \neq V^{-1}$ and AV is not idempotent, implying that the eigenvalues of AV are not 0 or 1, the asymptotic distribution of the Hausman test under H_o is a weighted sum of

[5]See, among others, Muirhead (1982, Ch. 1), Johnson and Kotz (1970, Ch.29).

[6]This Lemma holds also in the asymptotic case (using the Continuous Mapping Theorem, e.g. White, 1984, Lemma 4.27).

central chi-squares

$$h \sim \sum_{i=1}^{K} d_i z_i^2$$

where $z_i^2 \sim \chi_1^2$ and d_i are the eigenvalues of AV. This implies that the significance levels of the standard Hausman test are not correct.

Consider first the limiting case where $d_1 \rightarrow K$, $d_i \rightarrow 0$, $i = 2, .., K$. Figure 1-1 illustrates numerically that

$$\Pr\left[K\chi_1^2 > \chi_{K,\alpha}^2\right],$$

where $\chi_{K,\alpha}^2$ is the critical value for a test of size α under the χ_K^2 distribution. In this illustration α is set equal to 0.05.

In general we distinguish two effects: a scale effect if $\sum\limits_{i=1}^{K} d_i \neq K$, which is predictable (e.g. if $d_i = 2\ \forall\ i,\ h \sim 2\chi_K^2$) and a dispersion effect if $d_i \neq d_j$, even if $\sum\limits_{i=1}^{K} d_i = K$. We normalize the weights and we conjecture that the dispersion effect is maximized in the limit if we put all the weight on the largest eigenvalue, say the first one.

Figure 1-1 illustrates this case, i.e. the tail area of a χ_K^2 is compared with the maximum tail area of $K\chi_1^2$. The graph shows that the size distortion is an increasing function of K. For instance, if K is equal to 14, an inappropriate use of the Hausman test will give a probability of rejecting a true hypothesis of exogeneity which is almost 4 times larger than the nominal size.

In certain simple contexts an expression for the eigenvalues of AV can be analytically derived. For instance, a common source of misspecification in the variance covariance matrix occurs when the regressors contain measurement errors.

Suppose the true model is

$$y_{it} = z_{it}^{'}\beta + \eta_i + v_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T$$

where z_{it} is a $K \times 1$ vector of theoretical variables, $\eta_i \sim iid\left(0, \sigma_\eta^2\right)$, $v_{it} \sim iid\left(0, \sigma^2\right)$

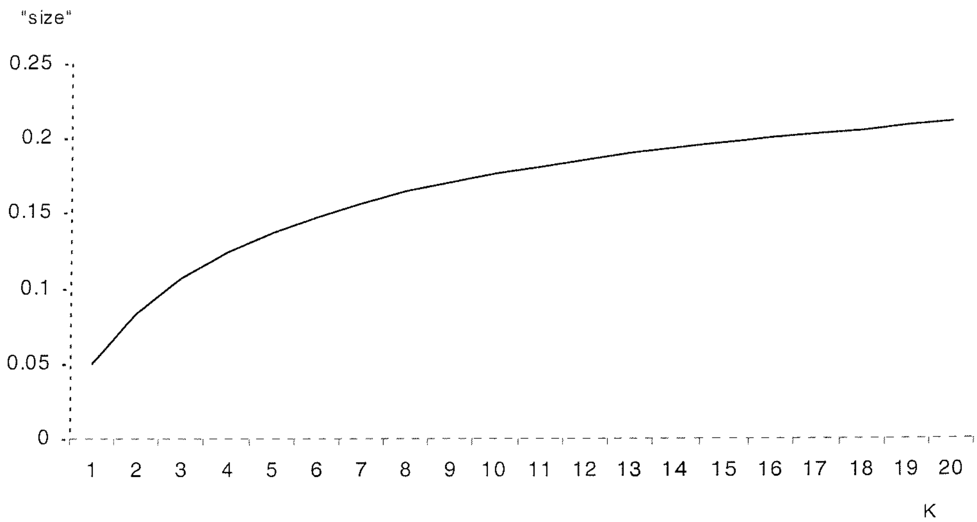


Figure 1-1: $\Pr\left[K\chi_1^2 > \chi^2_{K,\alpha=0.05}\right]$

uncorrelated with z_{it} and $Cov(\eta_i, v_{it}) = 0$. The observed variables are

$$x_{it} = z_{it} + m_{it},$$

where m_{it} is a vector of measurement errors uncorrelated with η_i and v_{it}. The estimated model is

$$y_{it} = x_{it}^{'}\beta + \eta_i + v_{it} - m_{it}^{'}\beta, \qquad i = 1, ..., N, \qquad t = 1, ..., T. \tag{1.31}$$

In the case of exact measurement, i.e. $m_{it} = 0$,

$$\begin{aligned} Var(y_{it}) &= E(\eta_i + v_{it})^2 = \sigma_\eta^2 + \sigma^2, \\ Cov(y_{it}, y_{it-s}) &= Cov(x_{it}^{'}\beta + \eta_i + v_{it}, x_{it-s}^{'}\beta + \eta_i + v_{it-s}) \\ &= \sigma_\eta^2 \qquad \forall s. \end{aligned}$$

The variance-covariance matrix is matrix (1.5). It can be written as

$$\underset{(NT\times NT)}{\Sigma} = I_N \otimes \Omega,$$

where

$$\Omega = \sigma^2 I_T + \sigma_\eta^2\, \iota\iota^{'} = \sigma^2[I_T + \vartheta_1 \iota\iota^{'}],$$

and

$$\vartheta_1 = \frac{\sigma_\eta^2}{\sigma^2}.$$

If we assume that $m_{it} \sim iid\,(0, \Sigma_M)$, we obtain

$$\begin{aligned} Var(y_{it}) &= E(\eta_i + v_{it} - \beta m_{it})^2 = \sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta, \\ Cov(y_{it}, y_{it-s}) &= Cov(x_{it}^{'}\beta + \eta_i + v_{it} - \beta' m_{it}, x_{it-s}^{'}\beta + \eta_i + v_{it-s} - \beta' m_{it-s}) \\ &= \sigma_\eta^2 \qquad \forall s \neq 0. \end{aligned}$$

So

$$\Omega = \left(\sigma^2 + \beta'\Sigma_M\beta\right) I_T + \sigma^2_\eta\, \iota\iota' = \left(\sigma^2 + \beta'\Sigma_M\beta\right) (I_T + \vartheta_2\iota\iota'),$$

and

$$\vartheta_2 = \frac{\sigma^2_\eta}{\sigma^2 + \beta'\Sigma_M\beta}.$$

Consider now the exogeneity test based, for instance, on the comparison between $\widehat{\beta}_{bg}$ and $\widehat{\beta}_{wg}$. In this case, the measurement errors render $\widehat{\beta}_{bg}$ and $\widehat{\beta}_{wg}$ inconsistent. If we assume that

$$plim(\widehat{\beta}_{bg} - \beta) = plim(\widehat{\beta}_{wg} - \beta) = \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \Sigma_M\beta = \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \Sigma_M\beta,$$

O'Brien and Patacchini (2003) show that, in the normal case,

$$\begin{aligned}
&\sqrt{N}(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}) \xrightarrow{D} N(0, [1/(T-1)\}\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \times \\
&\Big[(\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}\Big] \times \\
&\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} + \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \times \\
&[T\sigma^2_\eta\Sigma_{ZMZ} + (\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZMZ} + \sigma^2_\eta T\Sigma_M + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}] \\
&\left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}),
\end{aligned}$$

where

$$\Sigma_{ZQZ} = plim(1/N)\sum_{i=1}^{N} Z_i'QZ_i, \qquad \Sigma_{ZMZ} = plim(1/N)\sum_{i=1}^{N} Z_i'MZ_i,$$

$$Z_i = \begin{bmatrix} z_{i1}' \\ z_{i2}' \\ \vdots \\ z_{iT}' \end{bmatrix}$$

and Q and M are defined in (1.6). The Hausman test

$$\begin{aligned} h &= (\widehat{\beta}_{wg} - \widehat{\beta}_{bg})' \left[\widehat{Var}(\widehat{\beta}_{wg}) + \widehat{Var}(\widehat{\beta}_{bg})\right]^{-1} (\widehat{\beta}_{wg} - \widehat{\beta}_{bg}) \\ &= \sqrt{N}(\widehat{\beta}_{wg} - \widehat{\beta}_{bg})' \left[N\widehat{Var}(\widehat{\beta}_{wg}) + N\widehat{Var}(\widehat{\beta}_{bg})\right]^{-1} \sqrt{N}(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}) \end{aligned}$$

will have the same asymptotic distribution as

$$h_a = \sqrt{N}(\widehat{\beta}_{wg} - \widehat{\beta}_{bg})' p\lim \left[N\widehat{Var}(\widehat{\beta}_{wg}) + N\widehat{Var}(\widehat{\beta}_{bg})\right]^{-1} \sqrt{N}(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}).$$

O'Brien and Patacchini (2003) also show that

$$\begin{aligned} & N\widehat{Var}(\widehat{\beta}_{wg}) \\ & \xrightarrow{p} \{\sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M \left[\frac{1}{(T-1)}\Sigma_{ZQZ} + \Sigma_M\right]^{-1} \Sigma_M\beta\} \times \\ & [\Sigma_{ZQZ} + (T-1)\Sigma_M]^{-1} \end{aligned}$$

and

$$\begin{aligned} & N\widehat{Var}(\widehat{\beta}_{bg}) \\ & \xrightarrow{p} \{T\sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M [\Sigma_{ZMZ} + \Sigma_M]^{-1} \Sigma_M\beta\} \times \\ & [\Sigma_{ZMZ} + \Sigma_M]^{-1}. \end{aligned}$$

Thus in terms of the notation of Lemma 3, for the asymptotic distribution,

$$\begin{aligned} V = & [1/(T-1)\} [\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} \times \\ & \left[(\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}\right] \times \\ & [\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} + [\Sigma_{ZMZ} + \Sigma_M]^{-1} \times \\ & [T\sigma_\eta^2\Sigma_{ZMZ} + (\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZMZ} + \sigma_\eta^2 T\Sigma_M + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}] \\ & [\Sigma_{ZMZ} + \Sigma_M]^{-1}). \end{aligned}$$

and

$$A = \begin{bmatrix} \{\sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M\left[\frac{1}{(T-1)}\Sigma_{ZQZ} + \Sigma_M\right]^{-1}\Sigma_M\beta\} \times [\Sigma_{ZQZ} + (T-1)\Sigma_M]^{-1} \\ +\{T\sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M[\Sigma_{ZMZ} + \Sigma_M]^{-1}\Sigma_M\beta\} \times [\Sigma_{ZMZ} + \Sigma_M]^{-1} \end{bmatrix}^{-1}.$$

Consider first the case when $\beta = 0$.

$$\begin{aligned} V &= [1/(T-1)]\,[\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} \times \\ &\quad \left[\sigma^2\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M\right] \times [\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} \\ &\quad + [\Sigma_{ZMZ} + \Sigma_M]^{-1} \times [T\sigma_\eta^2\Sigma_{ZMZ} + \sigma^2\Sigma_{ZMZ} + \sigma_\eta^2 T\Sigma_M + \sigma^2\Sigma_M]\,[\Sigma_{ZMZ} + \Sigma_M]^{-1} \\ &= [1/(T-1)]\sigma^2\,[\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} \\ &\quad + [\Sigma_{ZMZ} + \Sigma_M]^{-1}\,(T\sigma_\eta^2 + \sigma^2)[\Sigma_{ZMZ} + \Sigma_M]\,[\Sigma_{ZMZ} + \Sigma_M]^{-1} \\ &= [1/(T-1)]\sigma^2\,[\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} \\ &\quad +(T\sigma_\eta^2 + \sigma^2)\,[\Sigma_{ZMZ} + \Sigma_M]^{-1}, \end{aligned}$$

$$A = \left[\sigma^2\,[\Sigma_{ZQZ} + (T-1)\Sigma_M]^{-1} + \{T\sigma_\eta^2 + \sigma^2\} \times [\Sigma_{ZMZ} + \Sigma_M]^{-1}\right]^{-1}.$$

So $AV = I$. As a check, when $\Sigma_M = 0$,

$$V = \sigma^2[1/(T-1)\}\,[\Sigma_{ZQZ}/(T-1)]^{-1} + [T\sigma_\eta^2 + \sigma^2]\,[\Sigma_{ZMZ}]^{-1}$$

$$A = \left[\sigma^2\Sigma_{ZQZ}^{-1} + \{T\sigma_\eta^2 + \sigma^2\}\Sigma_{ZMZ}^{-1}\right]^{-1}$$

which can be compared with the results contained in Appendix 3.1.

Now let $\Sigma_Q = \Sigma_{ZQZ}/(T-1)$, $\sigma^{*2} = \sigma^2 + \beta'\Sigma_M\beta$, $c{=}\Sigma_M\beta$, $\sigma^{**2} = \sigma^{*2} + T\sigma_\eta^2$, so

$$\begin{aligned} V &= [1/(T-1)]\,[\Sigma_Q + \Sigma_M]^{-1}\left[\sigma^{*2}[\Sigma_Q + \Sigma_M] + cc'\right][\Sigma_Q + \Sigma_M]^{-1} + \\ &\quad [\Sigma_{ZMZ} + \Sigma_M]^{-1}\left[\sigma^{**2}[\Sigma_{ZMZ} + \Sigma_M] + cc'\right][\Sigma_{ZMZ} + \Sigma_M]^{-1} \\ &= [1/(T-1)]\left[\sigma^{*2}[\Sigma_Q + \Sigma_M]^{-1} + dd'\right] + \left[\sigma^{**2}[\Sigma_{ZMZ} + \Sigma_M]^{-1} + ee'\right] \end{aligned}$$

where $d = [\Sigma_Q + \Sigma_M]^{-1} c$, and $e = [\Sigma_{ZMZ} + \Sigma_M]^{-1} c$. These are just the inconsistencies, which we are assuming equal.

$$A = \begin{bmatrix} 1/(T-1)\{\sigma^{*2} - c' [\Sigma_Q + \Sigma_M]^{-1} c\} \times [\Sigma_Q + \Sigma_M]^{-1} \\ +\{\sigma^{**2} - c' [\Sigma_{ZMZ} + \Sigma_M]^{-1} c\} \times [\Sigma_{ZMZ} + \Sigma_M]^{-1} \end{bmatrix}^{-1}$$

The simplest case to examine is when $\Sigma_Q = \Sigma_{ZMZ} \Leftrightarrow p\lim \widehat{\beta}_{wg} = p\lim \widehat{\beta}_{bg}$ for all β; let $\Sigma_{QM} = \Sigma_Q + \Sigma_M = \Sigma_{ZMZ} + \Sigma_M$. Noting $d = e$, we have

$$V = \sigma^{+2} \Sigma_{QM}{}^{-1} + 2dd'$$

where

$$\begin{aligned} \sigma^{+2} &= [1/(T-1)]\sigma^{*2} + \sigma^{**2} \\ &= [T/(T-1)]\sigma^{*2} + T\sigma_\eta^2 \end{aligned}$$

$$A = [\sigma^{++2} \Sigma_{QM}^{-1}]^{-1}$$

and

$$\begin{aligned} \sigma^{++2} &= [1/(T-1)]\{\sigma^{*2} - c'\Sigma_{QM}^{-1}c\} + \sigma^{**2} - c'\Sigma_{QM}^{-1}c, \\ &= [T/(T-1)][\sigma^{*2} - c'\Sigma_{QM}^{-1}c] + T\sigma_\eta^2. \end{aligned}$$

AV has the same eigenvalues as

$$A^{1/2} V A^{1/2} = \frac{\sigma^{+2}}{\sigma^{++2}} I + \frac{2}{\sigma^{++2}} \Sigma_{QM}{}^{1/2} dd' \Sigma_{QM}{}^{1/2}$$

and has $K-1$ eigenvalues of

$$k = \sigma^{+2}/\sigma^{++2}$$

and one of

$$
\begin{aligned}
k + (2/\sigma^{++2}) d' \Sigma_{QM} d &= k + (2/\sigma^{++2}) c' \Sigma_{QM}^{-1} c \\
&= k + (2/\sigma^{++2}) \boldsymbol{\beta}' \Sigma_M \Sigma_{QM}^{-1} \boldsymbol{\Sigma}_M \beta.
\end{aligned}
$$

Thus the size distortion depends on scalar quantities,

$$
\begin{aligned}
k &= \frac{\sigma^{+2}}{\sigma^{++2}} = \frac{1}{1-k^*}, \\
k^* &= \frac{\sigma^{+2} - \sigma^{++2}}{\sigma^{+2}} = \frac{\boldsymbol{\beta}' \Sigma_M \Sigma_{QM}^{-1} \boldsymbol{\Sigma}_M \beta}{[T/(T-1)]\{\sigma^2 + \beta' \Sigma_M \beta\} + T\sigma_\eta^2}
\end{aligned}
$$

and the larger root is

$$
\frac{\sigma^{+2}}{\sigma^{++2}} + \frac{2}{\sigma^{++2}} k^* \sigma^{+2} = \frac{1}{1-k^*} \left[1 + 2k^*\right].
$$

$$
\begin{aligned}
\beta' \Sigma_M \Sigma_{QM}^{-1} \Sigma_M \beta &= \beta' \Sigma_M^{1/2} [\Sigma_M^{1/2} (\Sigma_Q + \Sigma_M)^{-1} \Sigma_M^{1/2}] \Sigma_M^{1/2} \beta \\
&= \beta' \Sigma_M^{1/2} [\Sigma_M^{-1/2} \Sigma_Q \Sigma_M^{-1/2} + I]^{-1} \Sigma_M^{1/2} \beta.
\end{aligned}
$$

If we now consider

$$
\gamma = \Sigma_M^{1/2} \beta,
$$

from model (1.31), γ is the vector of parameters in the model

$$
\begin{aligned}
y_i &= [Z_i + M_i] \, \Sigma_M^{-1/2} \Sigma_M^{1/2} \beta + \eta_i i + v_i \\
&= Z_i^* \gamma + M_i^* \gamma + \eta_i i + v_i
\end{aligned}
$$

where

$$\boldsymbol{\nu}_i = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{bmatrix}, M_i = \begin{bmatrix} m'_{i1} \\ m'_{i2} \\ \vdots \\ m'_{iT} \end{bmatrix},$$

the rows of M_i are $NID(0, I)$, $M_i^* = M_i\Sigma_M^{-1/2}$, $Z_i^* = Z_i\Sigma_M^{-1/2} \Rightarrow Z_i = Z_i^*\Sigma_M^{1/2} \Rightarrow Z_i'M^+Z_i = \Sigma_M^{1/2}Z_i^{*\prime}M^+Z_i^*\Sigma_M^{1/2}$, M^+ is defined in (1.6).

Thus,

$$\begin{aligned} k^* &= \gamma'[\Sigma_M^{-1/2}\Sigma_{ZMZ}\Sigma_M^{-1/2}+I]^{-1}\gamma/\sigma^{+2} \\ &= \gamma'\left[\Sigma_{Z^*MZ^*} + I\right]^{-1}\gamma/\left[[T/(T-1)]\{\sigma^{*2} + \gamma'\gamma\} + T\sigma_\eta^2\right], \end{aligned} \tag{1.32}$$

where $\Sigma_{Z^*MZ^*} = \Sigma_M^{-1/2}\Sigma_{ZMZ}\Sigma_M^{-1/2}$.

The components of the variance of y_{it} are

$$Var(y_{it}) = \gamma'\gamma + \sigma_\eta^2 + \sigma^2.$$

So an interpretation of our result is that if one takes one component of the variance, $\gamma'\gamma$, downweights it by the between sums of squares of the unobserved 'true' variables (in the model with standardised measurement errors), to produce $\gamma'\left[\Sigma_{Z^*MZ^*} + I\right]^{-1}\gamma$, then the 'size' distortion depends on k^*, as in (1.32), and the asymptotic distribution of the Hausman test is not χ_K^2, but a weighted sum of K χ_1^2, $K-1$ weights being $1/(1-k^*)$, with one of $[1 + 3k^*/(1-k^*)]$. It also follows that a lower bound to the distortion is provided by multiplying a χ_K^2 by $1/(1-k^*)$.

A number of qualifications are in order. This only occurs if the inconsistency of within and between estimators is equal, and, further, the within groups sum of squares matrix,

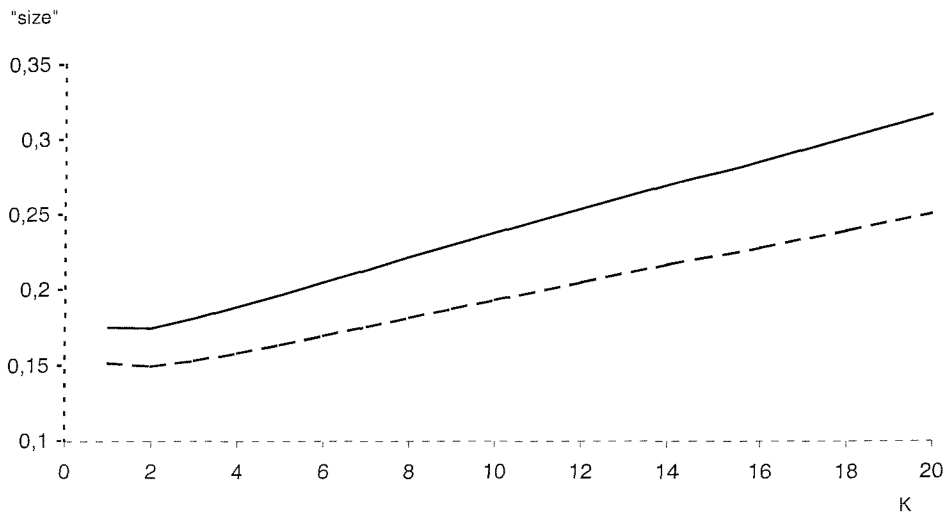


Figure 1-2: 'size' *vs* K

and between groups sum of squares matrix, are equal:

$$\Sigma_{ZMZ}, = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} Z_i M^+ Z_i, = \Sigma_Q, = \frac{1}{T-1} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} Z_i Q^+ Z_i.$$

The equality of $plim(\widehat{\beta}_{bg} - \beta)$ and $plim(\widehat{\beta}_{wg} - \beta)$ is required to ensure that the asymptotic 'size' is not 1. (Thus the Hausman test can be regarded as a (consistent) test of equality of these 'inconsistencies'). The equality of Σ_{ZMZ} and Σ_Q simplifies the result and is an aid to interpretability. We also assume that the rows of M_i, the measurement errors, are $NID(0, \Sigma_M)$. Some assumption about fourth moments is required, and this appears the simplest.

We can plot the size distortion for assumed values of $T, K, \gamma'\gamma, \gamma' \left[\Sigma_{Z^*MZ^*} + I\right]^{-1} \gamma, \sigma_\eta^2$ and σ^2. If $T = 5$ or $10, 1 \leq K \leq 10, \gamma'\gamma = 1, \sigma_\eta^2 = \sigma^2 = 0.1$, and $\gamma' \left[\Sigma_{Z^*MZ^*} + I\right]^{-1} \gamma = 0.5$, we have Figure 1-2, evaluated by Monte Carlo (1 million replications).

We can relax the assumption that $\Sigma_Q = \Sigma_{ZMZ}$ by observing that V is of the form

$$V = k_1 B + k_2 C + d^* d^{*\prime}$$

and A is of the form

$$A = (k_3 B + k_4 C)^{-1}$$

where

$$\begin{aligned}
B &= [\Sigma_Q + \Sigma_M]^{-1}, \quad C = [\Sigma_{ZMZ} + \Sigma_M]^{-1} \\
k_1 &= [1/(T-1)]\sigma^{*2}, \quad k_2 = \sigma^{**2} \\
d^* &= \{1 + 1/(T-1)\}^{1/2} d = \{T/(T-1)\}^{1/2} d, \\
k_3 &= 1/(T-1)\{\sigma^{*2} - c' B^{-1} c\}, < k_1 \\
k_4 &= \{\sigma^{**2} - c' C^{-1} c\}, < k_2
\end{aligned}$$

and B and C are positive definite. We see that A is "too small", and the test will be oversized.

$$\begin{aligned}
V &= B^{1/2}[k_1 I + k_2 B^{-1/2} C B^{-1/2} + B^{-1/2} d^* d^{*\prime} B^{-1/2}] B^{1/2} \\
A^{-1} &= B^{1/2}[k_3 I + k_4 B^{-1/2} C B^{-1/2}] B^{1/2}
\end{aligned}$$

Let

$$D = B^{-1/2} C B^{-1/2} = P \Lambda P'$$

where P is orthogonal, Λ diagonal, with as diagonal elements λ_i the eigenvalues of D. Then

$$\begin{aligned}
V &= B^{1/2} P [k_1 I + k_2 \Lambda + P' B^{-1/2} d^* d^{*\prime} B^{-1/2} P] P' B^{1/2} \\
A &= \left[B^{1/2} P [k_3 I + k_4 \Lambda] P' B^{1/2}\right]^{-1} = B^{-1/2} P [k_3 I + k_4 \Lambda]^{-1} P' B^{-1/2}
\end{aligned}$$

and thus

$$\begin{aligned} AV &= B^{-1/2}P[k_3 I + k_4\Lambda]^{-1}[k_1 I + k_2\Lambda + P'B^{-1/2}d^* d^{*\prime}B^{-1/2}P]P'B^{1/2} \\ &= B^{-1/2}P[diag([k_3+k_4\lambda_i]^{-1})\{diag(k_1+k_2\lambda_i) \\ &\quad +P'B^{-1/2}d^* d^{*\prime}B^{-1/2}P\}]P'B^{1/2} \end{aligned}$$

which has the same eigenvalues as

$$\begin{aligned} &\{diag(\frac{k_1+k_2\lambda_i}{k_3+k_4\lambda_i}) \\ &+diag([k_3+k_4\lambda_i]^{-1}\{P'B^{-1/2}d^* d^{*\prime}B^{-1/2}P\} \end{aligned}$$

The second matrix has rank 1, and the eigenvalues of the whole matrix are bounded between the smallest of $k_{0,i} = (k_1+k_2\lambda_i)/(k_3+k_4\lambda_i)$ and the largest of $k_{0,i}+d^{*\prime}B^{-1}d^*/(k_3+k_4\lambda_i)$. λ_i are the eigenvalues of $D = B^{-1/2}CB^{-1/2}$, or of $B^{-1}C = [\Sigma_Q + \Sigma_M][\Sigma_{ZMZ} + \Sigma_M]^{-1}$. $d = [\Sigma_Q + \Sigma_M]^{-1} c = [\Sigma_{ZMZ} + \Sigma_M]^{-1} c = Bc = Cc$

$$\begin{aligned} d^{*\prime}B^{-1}d^* &= \{T/(T-1)\}\mathbf{c}'B\mathbf{c} = \{T/(T-1)\}\beta'\Sigma_M[\Sigma_{ZMZ}+\Sigma_M]^{-1}\Sigma_M\beta \\ &= \{T/(T-1)\}\gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma \end{aligned}$$

$$\begin{aligned} k_1 &= [1/(T-1)]\sigma^{*2}, \sigma^{*2} = \sigma^2 + \beta'\Sigma_M\beta = \sigma^2 + \gamma'\gamma \\ k_2 &= \sigma^{**2} = \sigma^{*2} + T\sigma_\eta^2 \\ k_3 &= 1/(T-1)\{\sigma^{*2} - c'B^{-1}c\} = 1/(T-1)\left[\sigma^2 + \gamma'\gamma - \gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma\right] < k_1 \\ k_4 &= \{\sigma^{**2} - c'C^{-1}c\} = \sigma^2 + \gamma'\gamma + T\sigma_\eta^2 - \gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma < k_2 \end{aligned}$$

Thus

$$\begin{aligned} \sigma^{+2} &= [1/(T-1)]\sigma^{*2} + \sigma^{**2} = k_1 + k_2 \\ \sigma^{++2} &= [1/(T-1)]\{\sigma^{*2} - c'\Sigma_{QM}^{-1}c\} + \sigma^{**2} - c'\Sigma_{QM}^{-1}c = k_3 + k_4 \end{aligned}$$

$$
\begin{aligned}
k_{0,i} &= \frac{k_1 + k_2\lambda_i}{k_3 + k_4\lambda_i} = \frac{k_1 + k_2 + k_2(\lambda_i - 1)}{k_3 + k_4 + k_4(\lambda_i - 1)} \\
&= \frac{\sigma^{+2}[1 + k_2(\lambda_i - 1)/\sigma^{+2}]}{\sigma^{++2}[1 + k_4(\lambda_i - 1)/\sigma^{++2}]} = k\frac{[1 + k_2(\lambda_i - 1)/\sigma^{+2}]}{[1 + k_4(\lambda_i - 1)/\sigma^{++2}]}
\end{aligned}
$$

Thus comparing this case with the $B = C$ case, we are introducing more variability into the eigenvalues, which as we have seen , may well increase the 'size' of the test. (Thus the 'size' is sensitive to the relative magnitude of the intra-group and inter-group variations of the covariates, Σ_{ZQZ} and Σ_{ZMZ}). Our conclusion is somewhat dispiriting: a significant Hausman statistic may arise from measurement error, as it is implicitly comparing the inconsistencies: but cannot be used to test if the inconsistencies are equal, as the 'size' may considerably exceed its nominal value, even when the inconsistencies *are* equal.

1.5 A Power Comparison

The possible serious size distortion of the standard Hausman test motivates the formulation of the *HR-test*. Using the White (1984) estimators for the variance-covariance matrix, the test is robust to the presence of common sources of misspecification of the variance-covariance matrix, i.e. to arbitrary patterns of within groups dependence. In other words, using the notation in Lemma 3, AV is idempotent and the nominal size is equal to the observed one. We now use a simulation experiment to investigate the relative power of the standard Hausman test and the *HR-test*. We are interested in a quantitative assessment of the possible power loss that may incur in using a robust version of the test, in absence of misspecification.

The postulated data generation process is the following.

We consider the model

$$
y_{it} = \gamma x_{it} + \pi z_{it} + \eta_i + \xi_{it}, \quad i = 1, ..., N; \; t = 1, ..., T,
$$

where the disturbance term consists of two independent components: a unit-specific effect, η_i and a white noise component, ξ_{it} :

$$u_{it} = \eta_i + \xi_{it}.$$

The null hypothesis of the Hausman test is

$$Cov(x_{it}, u_{it}) = 0 \text{ and } Cov(z_{it}, u_{it}) = 0.$$

We assume z exogenous variable and we generate x correlated with u, so that the null hypothesis above is not satisfied. We consider

$$x_{it} = \varrho w_{it} + \varepsilon_{it}, \tag{1.33}$$

w is an exogenous variable and (u, ε) are drawn from a bivariate normal distribution with a specified correlation structure.

The values for the exogenous variables and the range of values for the parameters are taken from the empirical case of study analyzed in Chapter 2. Using UK data, the following model is estimated.

$$\log M_{it} = a + \alpha \log U_{it} + \beta \log V_{it} + e_{it}, \qquad i = 1, ..., 275;\ t = 1, ..., 63,$$

where M is the number of hirings, U and V are the stocks of unemployed and vacancies respectively, a is a constant term, e indicates a disturbance term. The estimates of α and β, 0.5 and 0.4, have been used in the simulation experiment for γ and π respectively. Also, the best prediction for the stock of vacancies is

$$\log V_{it} = 1.2 \log NV_{it}, \qquad i = 1, ..., 275;\ t = 1, ..., 63$$

where NV is the number of monthly notified vacancies (flow variable). In our experiment, the real values for U and NV have been used as exogenous variables, i.e. respectively

z and w. The endogenous variable, V, i.e. x, has been constructed according to the structure (1.33):

$$x_{it} = 1.2w_{it} + \varepsilon_{it}.$$

The equation estimated is

$$y_{it} = 0.5x_{it} + 0.4z_{it} + u_{it},$$

where (u, ε) are constructed as draws from a bivariate normal distribution with the specified correlation coefficient rho of $(0, 0.05, 0.10, \ldots, 0.95)$.

Six sample sizes, typically encountered in applied panel data studies are used. The experiment is repeated 5000 times for each sample size and level of correlation. Figures 1-3 to 1-5 contain the results of the simulation experiment. The power is expressed in percentages.

The tables displayed compare *H_pow*, the power of the standard Hausman statistic (*H-test*):

$$hs = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)' \left(\widehat{V}_{wg} + \widehat{V}_{bg}\right)^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)$$

with *HR_pow,* the power of the robust Hausman statistic (*HR-test*) obtained using the auxiliary regression detailed in Section 1.3:

$$hr = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)' \left[\widehat{Var\left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)}\right]^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right),$$

with different sample sizes. Figures 1-6 to 1-11 contained in Appendix 5.1 illustrate the relative power functions. The significance level has been fixed at 5%. $rho\hat{}$ is the estimated level of correlation between x and u conditioned upon w. For each level of rho, H_pow and HR_pow indicate the percentage of times we reject a false hypothesis if we use the *H-test* or the *HR-test* respectively. In Tables 1, 2 and 3 the number of cross-sectional units is held fixed at 25 and the number of time periods is varied between 4, 10 and 20 respectively. In Tables 4, 5 and 6 the number of cross-sectional units is held fixed at 275 and the number of time periods is varied respectively between 4, 10 and

Table 1: N=25, T=4

rho	rho^	H_pow	HR_pow
0.00	0.00	4.90	4.80
0.05	0.03	5.10	4.90
0.10	0.06	7.90	7.40
0.15	0.09	9.20	9.30
0.20	0.12	14.40	13.80
0.25	0.15	19.90	20.90
0.30	0.17	25.50	26.80
0.35	0.20	32.20	32.50
0.40	0.23	34.50	38.50
0.45	0.26	43.60	45.80
0.50	0.29	50.10	57.40
0.55	0.32	70.10	70.80
0.60	0.35	78.20	79.90
0.65	0.37	87.90	89.70
0.70	0.40	94.10	92.70
0.75	0.43	98.50	98.90
0.80	0.46	99.90	100.00
0.85	0.49	100.00	100.00
0.90	0.52	100.00	100.00
0.95	0.55	100.00	100.00

Table 2: N=25, T=10

rho	rho^	H_pow	HR_pow
0.00	0.00	4.60	4.50
0.05	0.04	6.50	5.40
0.10	0.08	8.10	6.10
0.15	0.11	12.50	9.20
0.20	0.15	16.40	13.90
0.25	0.17	20.60	20.10
0.30	0.21	25.40	27.50
0.35	0.25	31.50	32.50
0.40	0.28	40.10	43.30
0.45	0.32	50.20	55.50
0.50	0.35	57.20	61.90
0.55	0.39	70.20	72.70
0.60	0.42	82.40	85.40
0.65	0.46	88.60	90.00
0.70	0.49	99.80	96.70
0.75	0.53	99.90	99.40
0.80	0.56	99.90	99.90
0.85	0.60	100.00	99.90
0.90	0.64	100.00	100.00
0.95	0.67	100.00	100.00

Figure 1-3: Simulation Results

Table 3: N=25, T=20

rho	rho^	H_pow	HR_pow
0.00	0.00	4.80	4.70
0.05	0.04	6.80	5.90
0.10	0.07	9.00	8.10
0.15	0.10	17.80	16.50
0.20	0.14	27.80	27.00
0.25	0.18	36.10	36.40
0.30	0.21	46.20	48.10
0.35	0.25	66.20	66.50
0.40	0.28	79.00	79.60
0.45	0.32	87.20	87.90
0.50	0.35	95.00	93.90
0.55	0.39	97.80	97.70
0.60	0.42	99.10	98.70
0.65	0.46	99.90	99.80
0.70	0.50	99.90	100.00
0.75	0.53	100.00	100.00
0.80	0.57	100.00	100.00
0.85	0.60	100.00	100.00
0.90	0.64	100.00	100.00
0.95	0.67	100.00	100.00

Table 4: N=275, T=4

rho	rho^	H_pow	HR_pow
0.00	0.00	4.90	5.00
0.05	0.03	6.30	6.40
0.10	0.06	9.60	8.80
0.15	0.09	18.20	17.60
0.20	0.11	29.10	28.90
0.25	0.15	45.10	48.10
0.30	0.17	57.20	62.50
0.35	0.20	72.40	78.20
0.40	0.23	86.00	89.10
0.45	0.26	93.60	96.20
0.50	0.29	97.90	98.00
0.55	0.32	99.80	99.80
0.60	0.34	99.80	100.00
0.65	0.37	100.00	100.00
0.70	0.40	100.00	100.00
0.75	0.43	100.00	100.00
0.80	0.46	100.00	100.00
0.85	0.49	100.00	100.00
0.90	0.52	100.00	100.00
0.95	0.55	100.00	100.00

Figure 1-4: Simulation Results

Table 5: N=275, T=10

rho	rho^	H_pow	HR_pow
0.00	0.00	5.00	4.90
0.05	0.03	9.80	6.40
0.10	0.06	26.10	15.10
0.15	0.09	61.00	34.00
0.20	0.12	87.80	55.10
0.25	0.15	97.80	74.10
0.30	0.18	98.90	86.50
0.35	0.20	99.80	93.40
0.40	0.23	99.90	97.90
0.45	0.26	100.00	98.90
0.50	0.29	100.00	99.90
0.55	0.32	100.00	100.00
0.60	0.35	100.00	100.00
0.65	0.38	100.00	100.00
0.70	0.41	100.00	100.00
0.75	0.44	100.00	100.00
0.80	0.47	100.00	100.00
0.85	0.50	100.00	100.00
0.90	0.53	100.00	100.00
0.95	0.55	100.00	100.00

Table 6: N=275, T=20

rho	rho^	H_pow	HR_pow
0.00	0.00	5.10	4.70
0.05	0.03	18.40	6.40
0.10	0.06	59.70	18.90
0.15	0.09	91.10	40.10
0.20	0.12	99.80	62.40
0.25	0.15	99.90	75.50
0.30	0.18	99.90	87.40
0.35	0.20	100.00	94.10
0.40	0.23	100.00	98.90
0.45	0.26	100.00	100.00
0.50	0.29	100.00	100.00
0.55	0.32	100.00	100.00
0.60	0.35	100.00	100.00
0.65	0.38	100.00	100.00
0.70	0.41	100.00	100.00
0.75	0.44	100.00	100.00
0.80	0.47	100.00	100.00
0.85	0.50	100.00	100.00
0.90	0.53	100.00	100.00
0.95	0.56	100.00	100.00

Figure 1-5: Simulation Results

20. Tables 1 to 4 show that the performance of the *HR-test* is comparable with the one of the *H-test*, even better for values of *rho* greater than 0.3. In larger samples (Tables 5 and 6) the performance of the *H-test* is superior but the power loss of the *HR-test* is not serious. The *HR-test* gives a very high rejection frequency for the false hypothesis of absence of correlation between x and u, starting from levels of correlation around 0.3 (86.5% and 87.4% respectively in Tables 5 and 6) and it detects the endogeneity problem almost surely as soon as *rho* is higher than 0, 4 (97.9% and 98.9% respectively in Tables 5 and 6). Taking the results as a whole, if one excludes cases of small values of *rho*, the simulation experiment provides evidence that the performance of the *HR-test* in terms of power is satisfying in large samples and even better than the one given by the *H-test* in small samples.

In addition, it is worthwhile noting that a version of the Hausman test implemented in most econometric software, which is generally used in empirical studies, is the one based on the comparison between $\widehat{\beta}_{wg}$ and $\widehat{\beta}_{BN}$, i.e.

$$h_2 = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)' \left(\widehat{V}_{wg} - \widehat{V}_{BN}\right)^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right).$$

The problem with this approach is that, in finite samples, the difference between the two estimated variances of the estimators (i.e. $\widehat{V}_{wg} - \widehat{V}_{BN}$) may not be positive definite. In this cases, the use of a code implementing a different Hausman statistic or the formulation of the Hausman test using an auxiliary regression (e.g. the one proposed by Davidson and McKinnon (1993, p. 236), which is now already implemented in some statistical packages, e.g. a Stata 7 extension, or the (robust) one presented in this chapter) are the only possibilities to get a test outcome.

1.6 Conclusions

This chapter has presented a methodological revision of the use of the Hausman test for correlated effects with panel data. The relevance of the discussion is both theoretical and empirical. From a theoretical point of view, it is shown that the assumptions in Lemma 2.1. in Hausman (1978) are sufficient but not necessary. In particular, it is demonstrated that the attainment of the absolute Fisher lower bound can be replaced by the attainment of a relative minimum variance bound. From an empirical point of view, the main implication of this chapter is a *caveat* on the use of the standard Hausman test framework for correlated effects in applied panel data studies. The assumptions underlying the construction of the Hausman statistic (Hausman, 1978) may be rarely satisfied in empirical work. An analytical investigation of the size of the test shows that, at least in some cases, the distortion is substantial. The econometrics of panel data offers a variety of estimators for the same parameters. Our recommendation is to use the Hausman test framework for the comparison of appropriate pairs of panel data estimators and to construct a version of the test robust to deviations from the classical errors assumption, as proposed in this chapter. This test, the *HR-test*, gives correct significance levels in common cases of misspecification of the variance-covariance matrix of the errors and has a power comparable to the Hausman test when no evidence of misspecification is present. The power of the *HR-test* is even higher in small samples. It can be easily implemented using a standard econometric package.

1.7 Appendix 1.1

Lemma 4 *If t_1 and t_2 are unbiased estimators of $\theta \in R^p$, with t_1 minimum variance (MV) at least in the set*

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}$$

then

$$Cov(t_1, t - t_1) = \mathbf{0}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{0}$ a null matrix, and $\mathbf{A} \in R^{p \times p}$ is fixed.

Proof.

$$\begin{aligned} t &= \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2 = t_1 + (\mathbf{I} - \mathbf{A})(t_2 - t_1) \\ &= t_1 + \mathbf{B}d, \text{ say, } \mathbf{B} \in R^{p \times p} \end{aligned}$$

$$\begin{aligned} Var(t) &= E\{[t_1 - \theta + \mathbf{B}d]\,[t_1 - \theta + \mathbf{B}d]'\} \\ &= Var(t_1) + Cov(t_1, d)\mathbf{B}' + \mathbf{B}Cov(d, t_1) + \mathbf{B}Var(d)\mathbf{B}'. \end{aligned}$$

Thus we can write

$$Var(t) - Var(t_1) = \mathbf{CB}' + \mathbf{BC}' + \mathbf{BDB}'.$$

The minimum variance property of t_1 implies this difference is positive semi-definite, and thus for every $\lambda \in R^p$, and $\mathbf{B} \in R^{p \times p}$,

$$Q = \lambda' \left(\mathbf{CB}' + \mathbf{BC}' + \mathbf{BDB}'\right) \lambda \geq 0.$$

However, for the particular case of

$$\mathbf{B} = -\mathbf{CD}^{-1}$$

$$\begin{aligned} Q &= \lambda'(-\mathbf{C}\mathbf{D}^{-1}\mathbf{C}' - \mathbf{C}\mathbf{D}^{-1}\mathbf{C}' + \mathbf{C}\mathbf{D}^{-1}\mathbf{D}\mathbf{D}^{-1}\mathbf{C}')\lambda \\ &= \lambda'(-\mathbf{C}\mathbf{D}^{-1}\mathbf{C}')\lambda \end{aligned}$$

which satisfies the required inequality if and only if

$$\mathbf{C} = \mathbf{0}.$$

Further, for any $\mathbf{B} \in R^{p \times p}$

$$\begin{aligned} t - t_1 &= \mathbf{B}d, \\ Cov(t_1, t - t_1) &= \mathbf{C}\mathbf{B}' = \mathbf{0}. \end{aligned}$$

■

Remark 5 *We exclude the case where* $\mathbf{D}$ *is singular, as in that case replacing* $\mathbf{D}^{-1}$ *with a pseudo-inverse* $\mathbf{D}^{+}$ *such that* $\mathbf{D}^{+}\mathbf{D}\mathbf{D}^{+} = \mathbf{D}^{+}$ *reveals that all that is required is* $\mathbf{C}\mathbf{D}^{+}\mathbf{C}' = \mathbf{0}$, *or that* $\mathbf{C}$ *has rows orthogonal to the eigenvectors of* $\mathbf{D}$ *corresponding to the non-zero roots. As an example, consider the case where some elements of* t_1 *and* t_2 *coincide. It is simplest to exclude the coincident elements, and apply the argument above to the reduced vectors so formed.*

Remark 6 *This lemma implies that the MV unbiased estimator is uncorrelated with its difference from any other unbiased estimator, and the MV linear unbiased estimator is uncorrelated similarly.*

We next show that a set of the form $\mathcal{T}$ in Lemma 1 contains a minimum variance estimator. First, it is convenient to re-write the basis of the set in terms of t_1 and t_3, where $Cov(t_3, t_1) = \mathbf{0}$.

Lemma 7 *If* t_1 *and* t_2 *are unbiased estimators of* $\theta \in R^p$ *with covariance matrix* $\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$, *the set*

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}$$

can also be defined in terms of t_1 *and*

$$t_3 = \mathbf{B}t_1 + (\mathbf{I} - \mathbf{B})t_2$$

where

$$Cov(t_3, t_1) = \mathbf{0}$$

as

$$\mathcal{T} = \{t : t = \mathbf{C}t_1 + (\mathbf{I} - \mathbf{C})t_3\}$$

with

$$\mathbf{B} = -\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}, \mathbf{I} - \mathbf{B} = \mathbf{V}_{11}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}$$

$$Var(t_3) = -\mathbf{D}\mathbf{V}_{11}^{-1}\mathbf{D}' + \mathbf{D}\mathbf{V}_{21}^{-1}\mathbf{V}_{22}\mathbf{V}_{12}^{-1}\mathbf{D}', \mathbf{D} = \left[\mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}\right]^{-1}$$

$$\mathbf{C} = \mathbf{A}(\mathbf{V}_{11} - \mathbf{V}_{21}) + \mathbf{V}_{21})\mathbf{V}_{11}^{-1}, \mathbf{I} - \mathbf{C} = (\mathbf{I} - \mathbf{A})(\mathbf{V}_{11} - \mathbf{V}_{12})\mathbf{V}_{11}^{-1}$$

$$Var(t) = \mathbf{C}\mathbf{V}_{11}\mathbf{C}' + (\mathbf{I} - \mathbf{C})Var(t_3)(\mathbf{I} - \mathbf{C})'$$

Proof.

$$\begin{aligned} Cov(t_3, t_1) &= E\{[\mathbf{B}t_1 + (\mathbf{I} - \mathbf{B})t_2 - \theta][t_1 - \theta]'\} \\ &= \mathbf{B}\mathbf{V}_{11} + (\mathbf{I} - \mathbf{B})\mathbf{V}_{21} \\ &= -\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}\mathbf{V}_{11} + \mathbf{V}_{11}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}\mathbf{V}_{21} \end{aligned}$$

Now

$$\begin{aligned} \left[\mathbf{V}_{11}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}\mathbf{V}_{21}\right]^{-1} &= \mathbf{V}_{21}^{-1}(\mathbf{V}_{11} - \mathbf{V}_{21})\mathbf{V}_{11}^{-1} \\ &= \mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1} \end{aligned}$$

and

$$\left[\mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{11}\right]^{-1} = \mathbf{V}_{11}^{-1}(\mathbf{V}_{11}-\mathbf{V}_{21})\mathbf{V}_{21}^{-1}$$
$$= \mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}.$$

It follows that

$$\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{21} = \mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{11} \tag{1.34}$$

and thus

$$Cov(t_3, t_1) = \mathbf{0}$$

To find $Var(t_3)$, as

$$t_3 = \mathbf{B}t_1 + (\mathbf{I}-\mathbf{B})t_2$$

$$Var(t_3) = \mathbf{B}\mathbf{V}_{11}\mathbf{B}' + (\mathbf{I}-\mathbf{B})\mathbf{V}_{21}\mathbf{B}' + \mathbf{B}\mathbf{V}_{12}(\mathbf{I}-\mathbf{B})' + (\mathbf{I}-\mathbf{B})\mathbf{V}_{22}(\mathbf{I}-\mathbf{B})'$$

$$\mathbf{B}\mathbf{V}_{11}\mathbf{B}' = \mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1'}\mathbf{V}_{21}'$$

$$(\mathbf{I}-\mathbf{B})\mathbf{V}_{21}\mathbf{B}' = -\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1'}\mathbf{V}_{21}'$$

Identity (1.34) implies equality between these expressions.

$$\mathbf{B}\mathbf{V}_{12}(\mathbf{I}-\mathbf{B})' = -\mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{12}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1'}\mathbf{V}_{11}'$$

Transposing (1.34), this becomes the same as the expression for $\mathbf{B}\mathbf{V}_{11}\mathbf{B}'$.

$$(\mathbf{I}-\mathbf{B})\mathbf{V}_{22}(\mathbf{I}-\mathbf{B})' = \mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{22}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1'}\mathbf{V}_{11}'$$

This suggests writing the matrix in (1.34) as

$$\mathbf{D} = \left[\mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}\right]^{-1}$$

to give

$$Var(t_3) = -\mathbf{D}\mathbf{V}_{11}^{-1}\mathbf{D}' + \mathbf{D}\mathbf{V}_{21}^{-1}\mathbf{V}_{22}\left(\mathbf{V}_{21}'\right)^{-1}\mathbf{D}'$$

■

Remark 8 *Again, we are assuming non-singularity, in particular of* $\mathbf{V}_{21}$*. One could apply the steps above to zero a single non-zero element of* $\mathbf{V}_{21}$*, by shrinking* t_1 *and* t_2 *to the corresponding elements. Repeated application would then replace* $\mathbf{V}_{21}$ *with a null matrix.*

We can now show that $\mathcal{T}$ always contains a minimum variance unbiased estimator.

Lemma 9 *If* t_1 *and* t_2 *and* $\mathcal{T}$ *are as in Lemma 7 but with* $\mathbf{V}_{12} = \mathbf{0}$ *then* t *has the minimum variance in* $\mathcal{T}$ *if*

$$\mathbf{A} = [\mathbf{V}_{11}^{-1} + \mathbf{V}_{22}^{-1}]^{-1}\mathbf{V}_{11}^{-1}$$

Proof. Let this value of t be t_M, the corresponding $\mathbf{A}$ be $\mathbf{A}_M$, and $\mathbf{V}_M = Var(t_M)$. Let

$$\mathbf{A}_M = \mathbf{E}\mathbf{V}_{11}^{-1}, \Rightarrow \mathbf{I} - \mathbf{A}_M = \mathbf{E}\mathbf{V}_{22}^{-1}$$

We have

$$\begin{aligned} Var(t_M) &= \mathbf{E}\mathbf{V}_{11}^{-1}\mathbf{V}_{11}\mathbf{V}_{11}^{-1}\mathbf{E} + \mathbf{E}\mathbf{V}_{22}^{-1}\mathbf{V}_{22}\mathbf{V}_{22}^{-1}\mathbf{E} \\ &= \mathbf{E}\left[\mathbf{V}_{11}^{-1} + \mathbf{V}_{22}^{-1}\right]\mathbf{E} = \mathbf{E}. \end{aligned}$$

Moreover,

$$\begin{aligned} Cov(t_M, t_1 - t_2) &= Cov(\mathbf{A}_M t_1 + (\mathbf{I} - \mathbf{A}_M)t_2, t_1 - t_2) \\ &= E[\{\mathbf{E}\mathbf{V}_{11}^{-1}(t_1 - \theta) + \mathbf{E}\mathbf{V}_{22}^{-1}(t_2 - \theta)\}\{t_1' - t_2'\}] \\ &= E[\mathbf{E}(\mathbf{V}_{11}^{-1}\mathbf{V}_{11} - \mathbf{V}_{22}^{-1}\mathbf{V}_{22})] = \mathbf{0}. \end{aligned}$$

If $t \in \mathcal{T}$,

$$\begin{aligned} t &= \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2 \\ &= (\mathbf{A}_M + \mathbf{A} - \mathbf{A}_M)t_1 + (\mathbf{I} - \mathbf{A}_M - \mathbf{A} + \mathbf{A}_M)t_2 \\ &= t_M + (\mathbf{A} - \mathbf{A}_M)(t_1 - t_2) \end{aligned}$$

Thus

$$Var(t) = Var(t_M) + (\mathbf{A} - \mathbf{A}_M)Var(t_1 - t_2)(\mathbf{A} - \mathbf{A}_M)'$$

and thus $Var(t)$ exceeds $Var(t_M)$ by a positive semi-definite difference, and thus t_M is the minimum variance estimator in $\mathcal{T}$. ■

Finally, we establish the large sample equivalent of Lemma 1.

Lemma 10 *Consider* $t'_* = [t'_1, t_2'], \theta'_* = [\theta', \theta']$

$$\sqrt{n}(t_* - \theta_*) \xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix})$$

where $\mathbf{V}_{11}$ *is the 'asymptotic variance', Avar, of* t_1 *and* $\mathbf{V}_{12}$ *is the 'asymptotic covariance' of* t_1 *and* t_2, $Acov(t_1, t_2)$. *If* t_1 *is asymptotically minimum variance at least in the class*

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}, \mathbf{A} \in R^{p \times p}, \ \textit{fixed,}$$

then if $t'_d = [t'_1, [t - t_1]'], \theta'_d = [\theta', \mathbf{0}']$

$$\sqrt{n}(t_d - \theta_d) \xrightarrow{D} (0, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & Var(t) - \mathbf{V}_{11} \end{bmatrix})$$

Proof. Let $t_d = \begin{bmatrix} t_1 \\ t_2 - t_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$, so, as $\theta_d = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \theta_*$,

$$\begin{aligned}
\sqrt{n}(t_d - \theta_d) &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \sqrt{n}(t_* - \theta_*) \\
&\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} - \mathbf{V}_{11} \\ \mathbf{V}_{21} - \mathbf{V}_{11} & \mathbf{V}_{11} - \mathbf{V}_{12} - \mathbf{V}_{21} + \mathbf{V}_{22} \end{bmatrix}) \\
&\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{C} \\ -\mathbf{C}' & \mathbf{D} \end{bmatrix}), \text{ say}
\end{aligned}$$

:

$$\begin{aligned}
t &= \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2 = t_1 + (\mathbf{I} - \mathbf{A})(t_2 - t_1) \\
&= t_1 + \mathbf{B}d, \text{ say, } \mathbf{B} \in R^{p \times p}
\end{aligned}$$

$$\begin{bmatrix} t_1 \\ t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{B} \end{bmatrix} t_d$$

$$\theta_* = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{B} \end{bmatrix} \theta_d$$

$$\begin{aligned}
\sqrt{n}(\begin{bmatrix} t_1 \\ t \end{bmatrix} - \theta_*) &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{B} \end{bmatrix} \sqrt{n}(t_d - \theta_d) \qquad (1.35) \\
&\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{11} + \mathbf{C}\mathbf{B}' \\ \mathbf{V}_{11} + \mathbf{B}\mathbf{C}' & \mathbf{V}_{11} + \mathbf{B}\mathbf{D}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{C}\mathbf{B}' \end{bmatrix})
\end{aligned}$$

so we can write

$$Avar(t) - Avar(t_1) = \mathbf{C}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{B}\mathbf{D}\mathbf{B}'.$$

The minimum variance property of t_1 implies this difference is positive semi-definite, and thus for every $\lambda \in R^p$, and $\mathbf{B} \in R^{p \times p}$,

$$Q = \lambda' \left(\mathbf{CB}' + \mathbf{BC}' + \mathbf{BDB}'\right) \lambda \geq 0.$$

However, for the particular case of

$$\mathbf{B} = -\mathbf{CD}^{-1}$$

$$\begin{aligned} Q &= \lambda'(-\mathbf{CD}^{-1}\mathbf{C}' - \mathbf{CD}^{-1}\mathbf{C}' + \mathbf{CD}^{-1}\mathbf{DD}^{-1}\mathbf{C}')\lambda \\ &= \lambda'(-\mathbf{CD}^{-1}\mathbf{C}')\lambda \end{aligned}$$

which satisfies the required inequality if and only if

$$\mathbf{C} = \mathbf{0}.$$

Further, for any $\mathbf{B} \in R^{p \times p}$

$$t - t_1 = \mathbf{B}d,$$

so as

$$\begin{bmatrix} t_1 \\ t - t_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} t_1 \\ d \end{bmatrix}$$

$$\begin{aligned} \sqrt{n}\left(\begin{bmatrix} t_1 \\ t - t_1 \end{bmatrix} - \theta_d\right) &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \sqrt{n}(t_d - \theta_d) \\ &\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{BDB}' \end{bmatrix}) \end{aligned}$$

where, as $\mathbf{C} = \mathbf{0}, \mathbf{V}_{11} = \mathbf{V}_{12} = \mathbf{V}_{21}, \mathbf{D} = \mathbf{V}_{22} - \mathbf{V}_{11}$. Moreover, from (1.35)

$$Var(t) = \mathbf{BDB}' + \mathbf{V}_{11} \Rightarrow Var(t - t_1) = \mathbf{BDB}' = Var(t) - Var(t_1)$$

as required. ■

Remark 11 *The assumption that* $\mathbf{A}$ *is fixed can be replaced by a stochastic matrix* $\mathbf{A}_n$ *with* $plim(\mathbf{A}_n) = \mathbf{A}$

Remark 12 *This lemma implies that an asymptotically* MV *consistent estimator is uncorrelated in large samples with its difference from any other consistent estimator.*

1.8 Appendix 2.1

In this Appendix we give further details about the expression for $Var(y_i)$ used in Section 1.2.

As

$$Var(y_i) = \Omega_i = \sigma^2 I_T + \sigma_\eta^2 \iota\iota',$$

we can use the formula (see, e.g., Golub and van Loan (1983, p.50))

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1} U)^{-1} V^T A^{-1}$$

which simplifies for vector u, v to

$$(A + uv^T)^{-1} = A^{-1} - \frac{1}{1 + v^T A^{-1} u} A^{-1} uv^T A^{-1}.$$

It follows that, if $\theta = \sigma_\eta^2 / \sigma^2$

$$\Omega_i = \sigma^2 [I_T + \theta \iota\iota'] = \sigma^2 \left[I_T - \frac{\theta}{1 + T\theta} \iota\iota' \right]^{-1}$$

$$= \sigma^2 \left[I_T - \frac{\sigma_\eta^2}{\sigma^2 + T\sigma_\eta^2} \iota\iota' \right]^{-1}.$$

1.9 Appendix 3.1

In Section 1.2 we focused our attention on the Hausman test constructed using the contrast between the *Within Groups* and the *Balestra-Nerlove* estimators. In this Appendix we show the derivation of the Hausman statistic for the comparison between the *Within Groups* and the *Between Groups* estimators. Using the notation in Section 1.2, the *Between Groups* estimator can be written as

$$\widehat{\beta}_{bg} = (X'MX)X'MY.$$

The variance is

$$Var(\widehat{\beta}_{bg}) = \left[X'MX\right]^{-1} X'M(VarY)M'X \left[X'MX\right]^{-1}.$$

Further

$$\begin{aligned} Var(M^+ y_i) &= M^+ Var(y_i) M^{+'} = \sigma^2 M^+ \left[I_T + \theta\iota\iota'\right] M^+ \\ &= \sigma^2 M^+ \left[I_T + \theta T M^+\right] M^+ = \sigma^2 (1+\theta T) M^+, \end{aligned}$$

where $\theta = \sigma_\eta^2/\sigma^2$. Thus

$$Var(MY) = \sigma^2(1+\theta T) I_N \otimes M^+ = \sigma^2(1+\theta T)M.$$

$$\begin{aligned} Var(\widehat{\beta}_{bg}) &= \sigma^2(1+\theta T)\left[X'MX\right]^{-1} X'MX \left[X'MX\right]^{-1} \\ &= \sigma^2(1+\theta T)\left[X'MX\right]^{-1}. \end{aligned}$$

In addition

$$\begin{aligned} Cov(\widehat{\beta}_{bg}, \widehat{\beta}_{wg}) &= \left[X'MX\right]^{-1} X'M(VarY)Q'X \left[X'QX\right]^{-1} \\ &= \sigma^2 \left[X'MX\right]^{-1} X'M \left[I_{NT} + \theta TM\right] QX \left[X'QX\right]^{-1} = 0 \end{aligned}$$

So

$$\begin{aligned} Var(\widehat{\beta}_{bg} - \widehat{\beta}_{wg}) &= Var(\widehat{\beta}_{bg}) + Var(\widehat{\beta}_{wg}) \\ &= \sigma^2(1+\theta T) \left[X'MX\right]^{-1} + \sigma^2 \left[X'QX\right]^{-1}. \end{aligned}$$

Thus we have as a test

$$(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}) \left[\sigma^2(1+\theta T) \left[X'MX\right]^{-1} + \sigma^2 \left[X'QX\right]^{-1}\right]^{-1} (\widehat{\beta}_{wg} - \widehat{\beta}_{bg}).$$

1.10 Appendix 4.1

Lemma 13 *If*

$$\widehat{\beta} = (X'X)^{-1}X'y, \widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y,$$
$$X^* = XA,\ |A| \neq 0,\ \widehat{\varepsilon} = y - X\widehat{\beta},\ \widehat{\varepsilon}^* = y - X^*\widehat{\beta}^*,$$

then

$$\begin{aligned} (X^{*\prime}X^*)^{-1} &= A^{-1}(X'X)^{-1}A'^{-1} \\ \widehat{\beta}^* &= A^{-1}\widehat{\beta} \\ \widehat{\varepsilon}^* &= \widehat{\varepsilon} \end{aligned}$$

Proof.

$$(X^{*\prime}X^*)^{-1} = (A'X'XA)^{-1} = A^{-1}(X'X)^{-1}A'^{-1}.$$

$$\widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y = A^{-1}(X'X)^{-1}A'^{-1}A'X'y = A^{-1}\widehat{\beta}.$$

$$\widehat{\varepsilon}^* = y - X^*\widehat{\beta}^* = y - XAA^{-1}\widehat{\beta} = y - X\widehat{\beta} = \widehat{\varepsilon}$$

■

Lemma 14 *If*

$$\widehat{\beta}_A = (X_A'X_A)^{-1}X_A'y_A,\ \widehat{\beta}_B = (X_B'X_B)^{-1}X_B'y_B,$$

$$\widehat{\varepsilon}_A = y_A - X_A\widehat{\beta}_A,\ \widehat{\varepsilon}_B = y_B - X_B\widehat{\beta}_B$$

$$X^* = \begin{bmatrix} X_A & X_A \\ 0 & X_B \end{bmatrix},\ y^* = \begin{bmatrix} y_A \\ y_B \end{bmatrix},\ \widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y$$

$$\widehat{\varepsilon}^* = y^* - X^*\widehat{\beta}^*$$

then

$$\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_A - \widehat{\beta}_B \\ \widehat{\beta}_B \end{bmatrix}, \widehat{\varepsilon}^* = \begin{bmatrix} \widehat{\varepsilon}_A \\ \widehat{\varepsilon}_B \end{bmatrix}$$

Proof. Let

$$X = \begin{bmatrix} X_A & 0 \\ 0 & X_B \end{bmatrix}, \Rightarrow X^* = X\begin{bmatrix} I & I \\ 0 & I \end{bmatrix} = XA \text{ say}$$

$$A^{-1} = \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix}$$

Further, it is an exercise in elementary matrix algebra to show that

$$\widehat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} \widehat{\beta}_A \\ \widehat{\beta}_B \end{bmatrix},\ \widehat{\varepsilon} = y - X\widehat{\beta} = \begin{bmatrix} \widehat{\varepsilon}_A \\ \widehat{\varepsilon}_B \end{bmatrix}.$$

So applying Lemma 13,

$$\widehat{\beta}^* = A^{-1}\widehat{\beta} = \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} \widehat{\beta}_A \\ \widehat{\beta}_B \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_A - \widehat{\beta}_B \\ \widehat{\beta}_B \end{bmatrix}$$

and

$$\widehat{\varepsilon}^* = \widehat{\varepsilon} = \begin{bmatrix} \widehat{\varepsilon}_A \\ \widehat{\varepsilon}_B \end{bmatrix}$$

■

Return now to model (1.20). Results (1.21) and (1.22) in Lemma 2 directly follow from the application of Lemma 13 and 14. Next, we will prove the remaining result in Lemma 2, i.e. (1.23).[7]

Let

$$\begin{aligned} H^+ &= \frac{1}{T}i', H = I_N \otimes H^+, H'H = \frac{1}{T}M \\ \widehat{\beta}_{bg} &= [(HX)'(HX)]^{-1}(HX)'(HY) = (X'MX)^{-1}X'MY \\ \widehat{\beta}_{wg} &= [(QX)'(QX)]^{-1}(QX)'(QY) = (X'QX)^{-1}X'QY \end{aligned}$$

Further, let G^+ be Arellano and Bover's (1990) forward orthogonal deviations matrix, $(T-1) \times T$, such that

$$\begin{aligned} G^+i &= 0, G^+G^{+\prime} = I_{(T-1)}, G^{+\prime}G^+ = Q^+ = I_T - \frac{1}{T}ii' \\ G &= I_N \otimes G^+, G'G = Q, GG' = I_N \otimes I_{(T-1)} = I_{N(T-1)} \\ \widehat{\beta}_{wg} &= [(GX)'(GX)]^{-1}(GX)'(GY) = (X'QX)^{-1}X'QY \end{aligned}$$

and identifying HX and HY with X_A and Y_A, GX and GY with X_B and Y_B, we see that the artificial regression of $Y^* = \begin{bmatrix} HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} HX & HX \\ 0 & GX \end{bmatrix}$ gives coefficients

[7] Note that it does not matter which way round one does the artificial regression given that the test for the equality of the two estimators is a quadratic form on the difference between the two.

$\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{bg} - \widehat{\beta}_{wg} \\ \widehat{\beta}_{wg} \end{bmatrix}$. In this case,

$$Var(Y^*) = \begin{bmatrix} HVar(Y)H' & 0 \\ 0 & GVar(Y)G' \end{bmatrix}.$$

If $\theta = \sigma_\eta^2/\sigma^2$we have

$$\begin{aligned} GVar(Y)G' &= \sigma^2 G(I_{NT} + \theta I_N \otimes ii')G' \\ &= \sigma^2 GG' \text{ as } G^+ i = 0 \\ &= \sigma^2 I_{N(T-1)} \end{aligned}$$

and

$$\begin{aligned} HVar(Y)H' &= \sigma^2 H(I_{NT} + \theta I_N \otimes ii')H' \\ &= \sigma^2 [I_N \otimes H^+](I_{NT} + \theta I_N \otimes ii')[I_N \otimes H^{+\prime}] \\ &= \sigma^2 [I_N \otimes (H^+ H^{+\prime}) + \theta I_N \otimes (H^+ ii' H^{+\prime}). \end{aligned}$$

As

$$H^+ = \frac{1}{T} i', H^+ i = 1, H^+ H^{+\prime} = \frac{1}{T},$$

$$HVar(Y)H' = \sigma^2 [\frac{1}{T} I_N + \theta I_N] = \frac{\sigma^2}{T}(1 + T\theta) I_N.$$

Assembling our results,

$$Var(Y^*) = \begin{bmatrix} \frac{\sigma^2}{T}(1 + T\theta) I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{bmatrix}.$$

If now $\widetilde{X} = \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix}$,

$$\begin{aligned} Var(\widehat{\beta}^*) &= (X^{*\prime}X^*)^{-1}X^{*\prime}Var(Y^*)X^*(X^{*\prime}X^*)^{-1} \\ &= A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime}. \end{aligned}$$

Next, we calculate this variance by separating the different components.

$$\begin{aligned} \widetilde{X}'Var(Y^*)\widetilde{X} &= \begin{bmatrix} X'H' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} X'H' & 0 \\ 0 & XG' \end{bmatrix} \begin{bmatrix} (\theta+1/T)HX & 0 \\ 0 & GX \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} (\theta/T+1/T^2)X'MX & 0 \\ 0 & X'QX \end{bmatrix}. \end{aligned}$$

$$(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}.$$

Thus

$$\begin{aligned} &(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1} \\ &= \sigma^2 \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \times \\ &\quad \begin{bmatrix} (\theta/T+1/T^2)X'MX & 0 \\ 0 & X'QX \end{bmatrix} \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} (T\theta+1)(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} &A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\ = \ &\sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (T\theta+1)(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \sigma^2 \begin{bmatrix} (T\theta+1)(X'MX)^{-1} & -(X'QX)^{-1} \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} (T\theta+1)(X'MX)^{-1} + (X'QX)^{-1} & -(X'QX)^{-1} \\ -(X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}.
\end{aligned} \tag{1.36}
$$

We now need to find the variance-covariance matrix the artificial regression will assume. This will be proportional to

$$
\begin{aligned}
(X^{*\prime}X^{*})^{-1} &= (A'\widetilde{X}'\widetilde{X}A)^{-1} = A^{-1}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} T(X'MX)^{-1} & -(X'QX)^{-1} \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} T(X'MX)^{-1} + (X'QX)^{-1} & -(X'QX)^{-1} \\ -(X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}.
\end{aligned} \tag{1.37}
$$

By comparing (1.36) with (1.37) it appears that an artificial regression is a valuable device to estimate a suitable variance-covariance matrix. This variance is estimated using a (White) robust *OLS* estimator which uses a consistent estimator of $X^{*\prime}Var(Y^{*})X^{*}$ under the assumption that $Var(Y^{*})$ is diagonal. Next, we derive this consistent estimator. Following the steps used in the derivation of $Var(\widehat{\beta}^{*})$ above, we separate the different components.

$$
\begin{aligned}
&\widetilde{X}'Var(Y^{*})\widetilde{X} \\
&= \begin{bmatrix} X'H' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \sigma^2\Omega & 0 \\ 0 & \sigma^2\Omega \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} X'H' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \Omega & 0 \\ 0 & \Omega \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma^2 \begin{bmatrix} X'H'\Omega & 0 \\ 0 & X'G'\Omega \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} XH'\Omega HX & 0 \\ 0 & X'G'\Omega GX \end{bmatrix}.
\end{aligned}
$$

$$
(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}.
$$

Thus

$$
\begin{aligned}
& (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1} \\
&= \sigma^2 \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \times \\
& \begin{bmatrix} XH'\Omega HX & 0 \\ 0 & X'G'\Omega GX \end{bmatrix} \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} T(X'MX)^{-1}\,(XH'\Omega HX) & 0 \\ 0 & (X'QX)^{-1}\,(X'G'\Omega GX) \end{bmatrix} \times \\
& \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} T^2(X'MX)^{-1}\,(XH'\Omega HX)\,(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1}\,(X'G'\Omega GX)\,(X'QX)^{-1} \end{bmatrix}.
\end{aligned}
$$

Let

$$
B = T^2(X'MX)^{-1}\,(X'H'\Omega HX)\,(X'MX)^{-1}
$$

and

$$
D = (X'QX)^{-1}\,(X'G'\Omega GX)\,(X'QX)^{-1}.
$$

So

$$
\begin{aligned}
& A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
= & \; \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= & \; \sigma^2 \begin{bmatrix} B & -D \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= & \; \sigma^2 \begin{bmatrix} B+D & -D \\ -D & D \end{bmatrix}.
\end{aligned}
$$

The residuals from this regression of $Y^* = \begin{bmatrix} HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} HX & HX \\ 0 & GX \end{bmatrix}$ to give coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{bg} - \widehat{\beta}_{wg} \\ \widehat{\beta}_{wg} \end{bmatrix}$ can be obtained by stacking those from HY on HX above those from GY on GX. The first set will yield sum of squares

$$
\begin{aligned}
RSS_A &= (HY)'[I_N - (HX)T(X'MX)^{-1}(X'H')]HY \\
&= \frac{1}{T}Y'(M - MX(X'MX)^{-1}X'M)Y.
\end{aligned}
$$

Note that $(M - MX(X'MX)^{-1}X'M) = M_P$ is idempotent, and $M_P MX = 0$.

Note that if we write the model as

$$Y = X\beta + E$$

we get

$$MY = MX\beta + ME,$$

$$M_P MY = M_P E$$

and

$$RSS_A = \frac{1}{T}E'M_PE.$$

The expectation is given by

$$\begin{aligned} ERSS_A &= \frac{1}{T}trace\,[M_PVar(E)] = \frac{1}{T}trace\,[M_PVar(Y)] \\ &= \frac{\sigma^2}{T}trace\,[M_P\{I_{NT} + \theta I_N \otimes ii'\}]\,. \end{aligned}$$

As

$$M(I_N \otimes ii') = (I_N \otimes \frac{1}{T}ii')(I_N \otimes ii') = I_N \otimes ii' = TM,$$

$$ERSS_A = \frac{\sigma^2}{T}(1+\theta T)trace(M_P) = \frac{\sigma^2}{T}(1+\theta T)(N-K).$$

Similarly, if

$$\begin{aligned} RSS_B &= (GY)'[I_{NT} - GX(X'QX)^{-1}X'G']GY \\ &= Y'[Q - QX(X'QX)^{-1}X'Q)]Y, \end{aligned}$$

$$\begin{aligned} ERSS_B &= \sigma^2 trace\,[Q_P\{I_{NT} + \theta I_N \otimes ii'\}] \\ &= \sigma^2 trace\,[Q_P] = \sigma^2[N(T-1) - K]. \end{aligned}$$

Accordingly, there is no multiple of $RSS_A + RSS_B$ with expectation σ^2. However, if in the first regression Y_A and X_A are scaled by

$$k = \sqrt{T/(1+\theta T)}$$

the coefficients will be unchanged, their variance will be unchanged, $(X_A'X_A)^{-1}$ will be scaled by $1/k^2 = (1+\theta T)/T$. So instead of

$$[(HX)'HX]^{-1} = T(X'MX)^{-1}$$

we will now have

$$(X_A'X_A)^{-1} = (1+\theta T)(X'MX)^{-1}.$$

Further,

$$k^2 ERSS_A = \frac{T}{(1+\theta T)}\frac{\sigma^2}{T}(1+\theta T)(N-K) = \sigma^2(N-K)$$

and $(k^2 RSS_A + RSS_B)/(NT-2K)$ is an unbiased estimator of σ^2.

Thus given a consistent estimator $\widehat{\theta}$ of θ, and thus $\widehat{k}$ of k, we can construct the Hausman test by carrying out the artificial regression of $Y^* = \begin{bmatrix} \widehat{k}HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} \widehat{k}HX & \widehat{k}HX \\ 0 & GX \end{bmatrix}$, and constructing a Wald test on the first K coefficients. In practice, as consistent estimator of θ one can use the one obtained under the assumption of spherical disturbances.

1.11 Appendix 5.1

This appendix contains the graphs of the power curve of the standard Hausman test (*H-test*) versus the one of the robust formulation presented in Section 1.3 (*HR-test*) with different sample sizes.

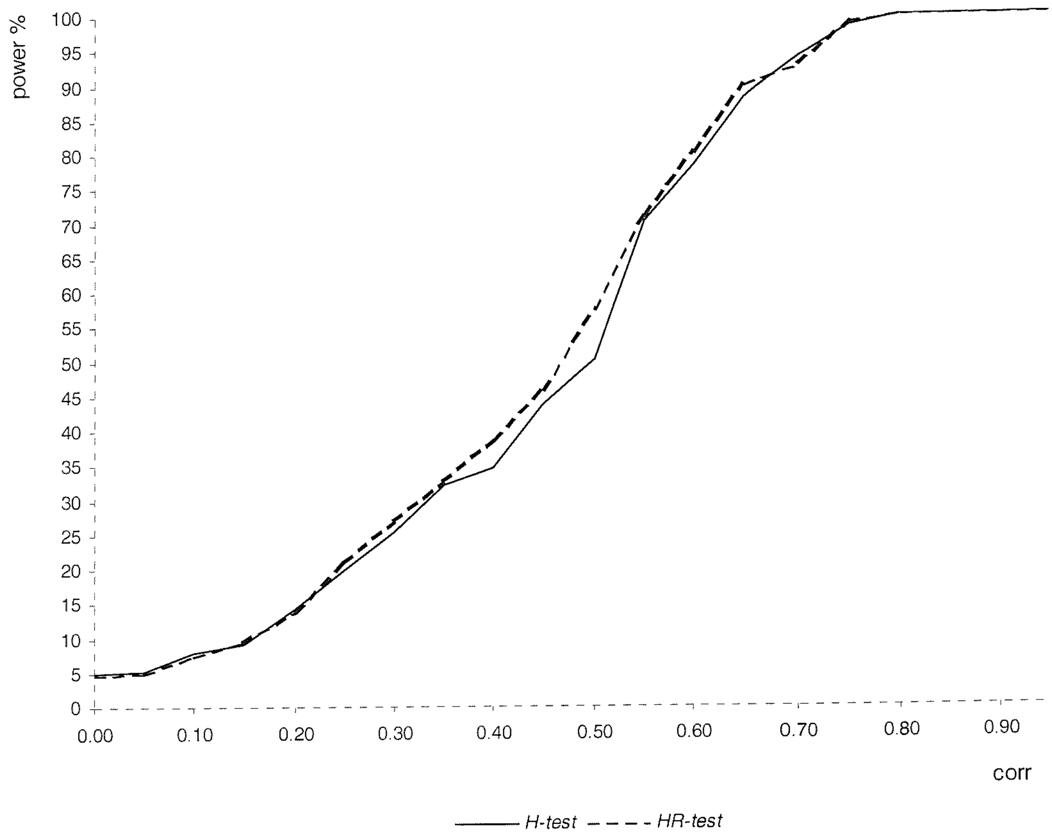


Figure 1-6: Power function comparison when N=25, T=4

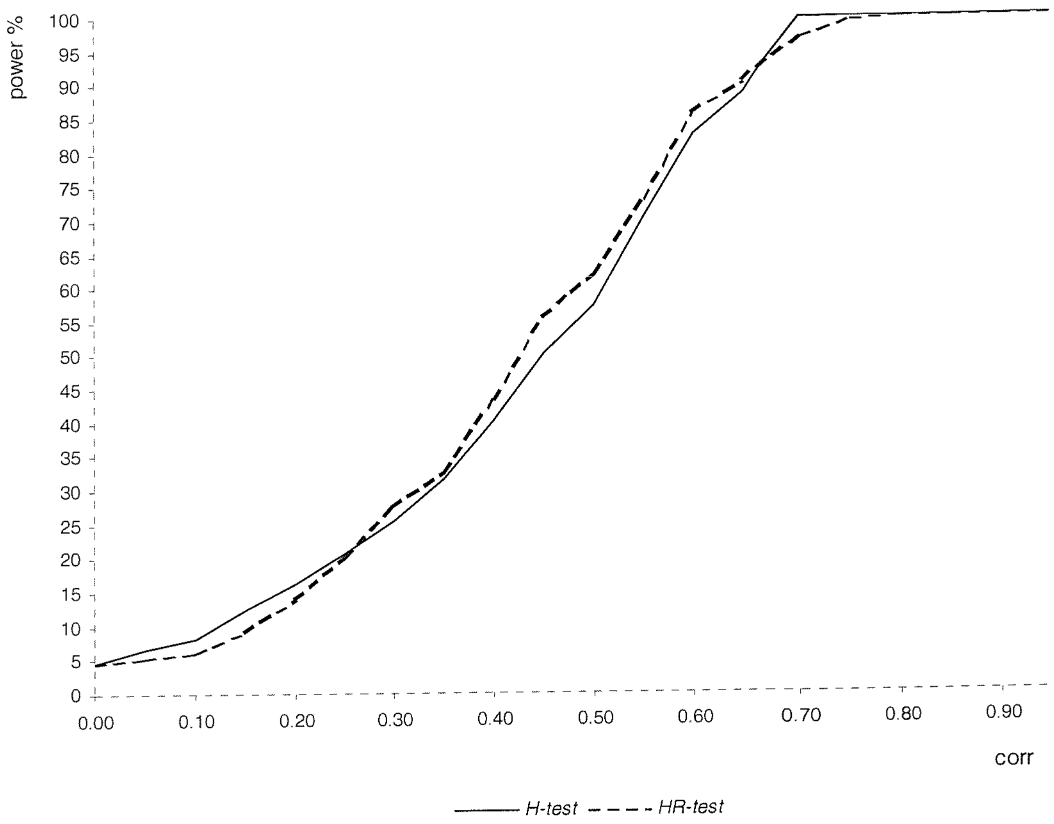


Figure 1-7: Power function comparison when N=25, T=10

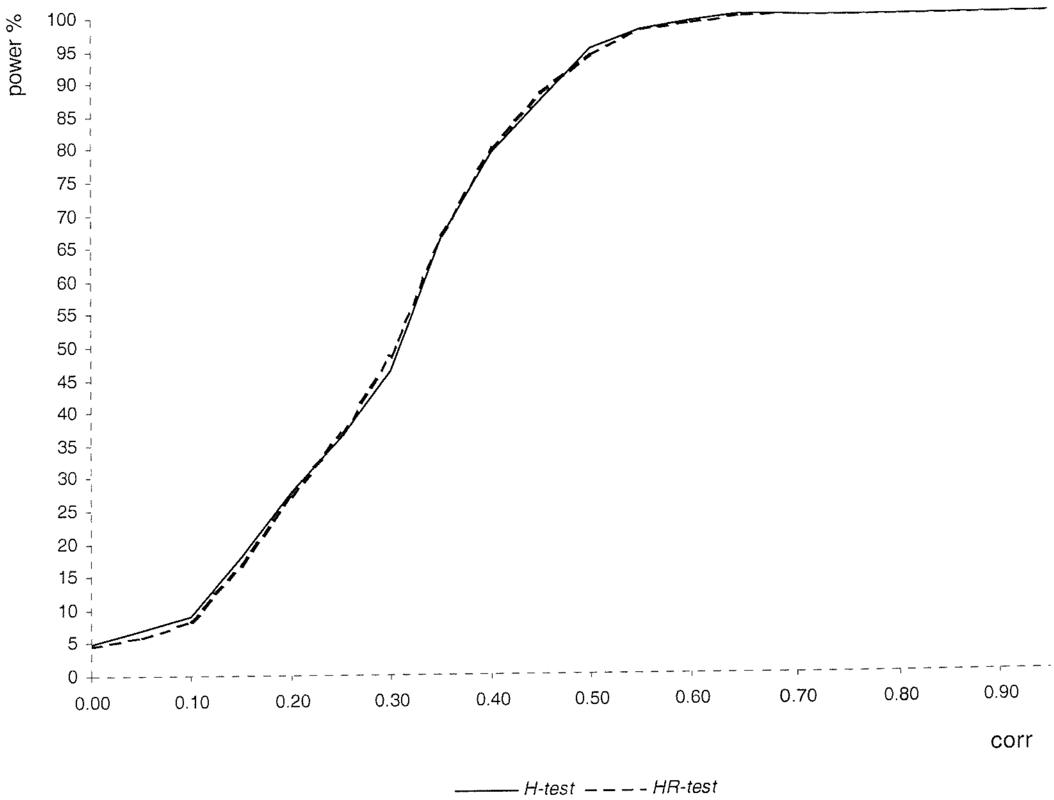


Figure 1-8: Power function comparison when N=25, T=20

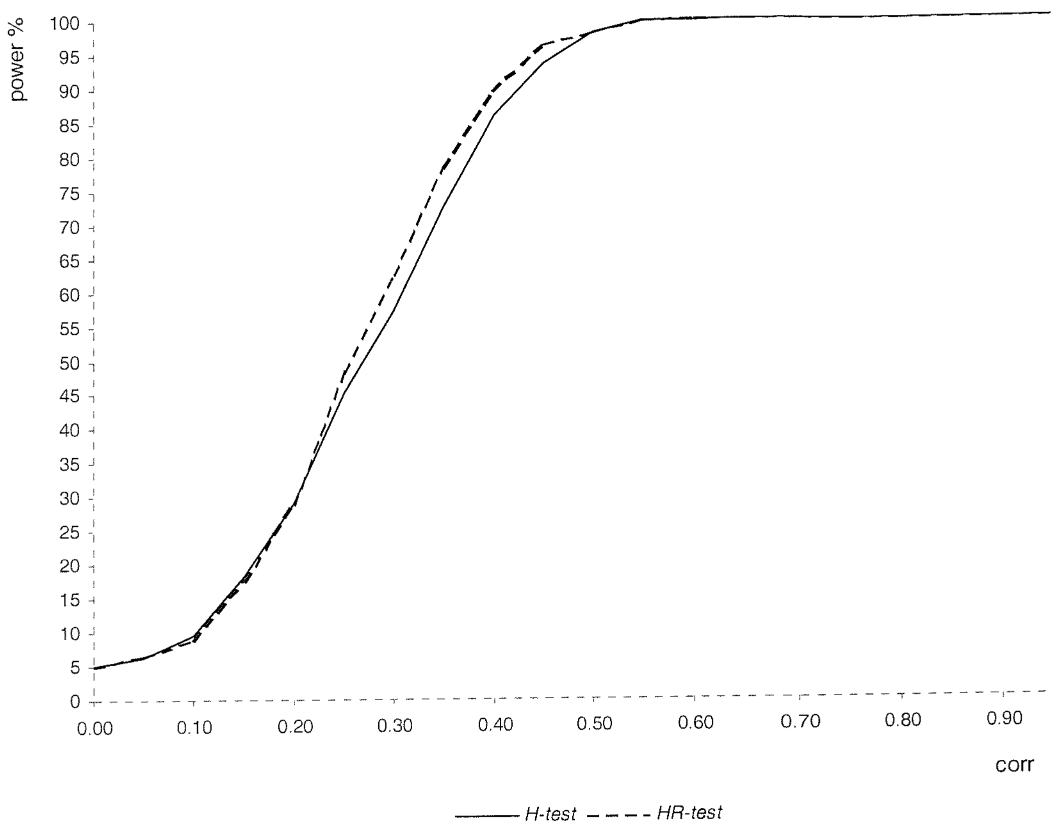


Figure 1-9: Power function comparison when N=275, T=4

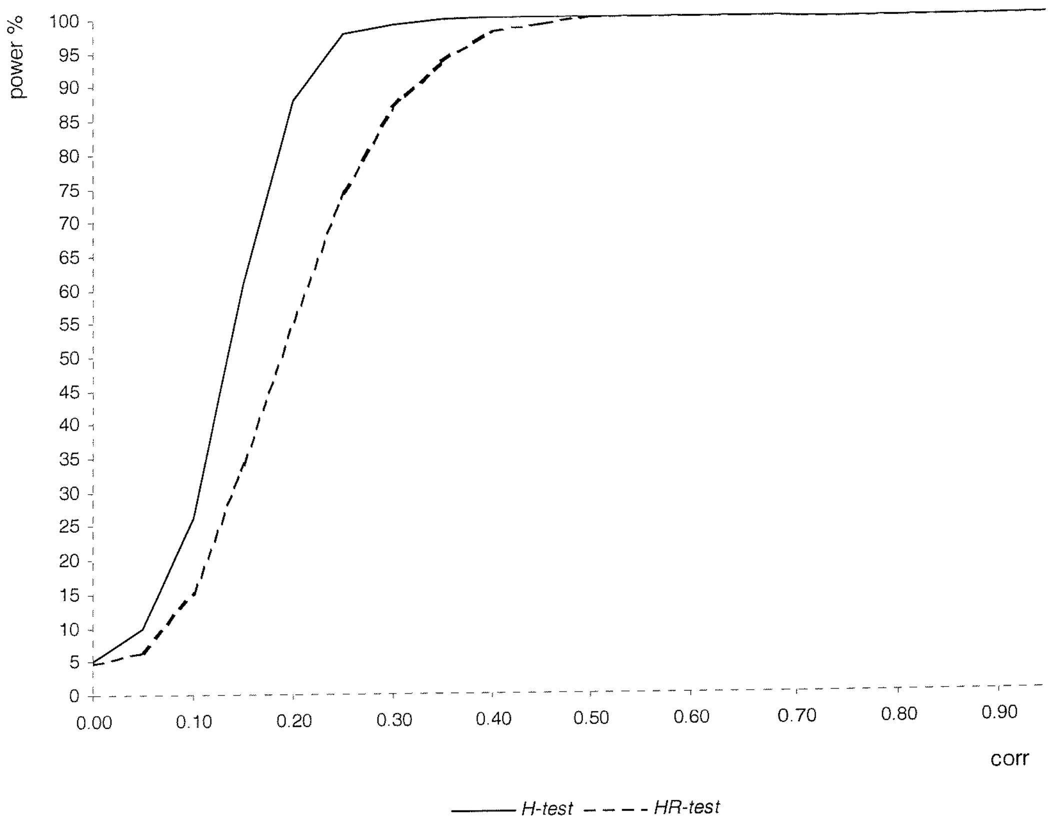


Figure 1-10: Power function comparison when N=275, T=10

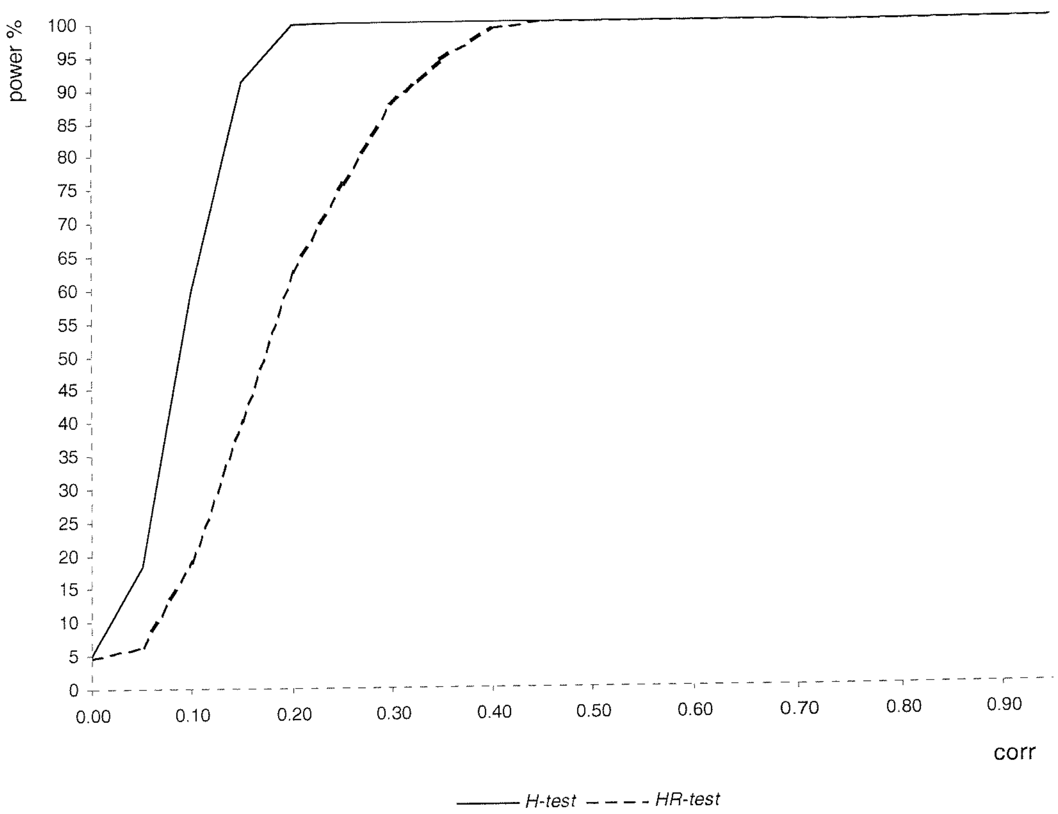


Figure 1-11: Power function comparison when N=275, T=20

Chapter 2

Unobservable Factors and Panel Data Sets: the Case of Matching Unemployed and Vacancies Data

This chapter presents a sequential procedure aiming to distinguish between an endogeneity problem caused by correlation between regressors and individual effects and an endogeneity problem due to measurement errors-in-variables. The relevance of the choice of the model specification is underlined. The statistical performance of the sequential test is assessed using simulated data. Considerations on the significance level and power of the testing procedure are presented. This procedure is then used to investigate the effects of unobservable factors like measurement errors and unobserved heterogeneity that, as is well-known, contaminate two of the variables extensively used in labor market research, namely the stock of unemployed and the stock of vacancies. Using a matching function framework, we compare different pairs of panel data estimators organized in a specific sequence.

2.1 Introduction

A statistical test which has large implications for applied studies is the Durbin-Wu-Hausman test, or DWH test (Durbin (1954), Wu (1973), Hausman (1978)). (In contrast to Chapter 1, where emphasis was put on Hausman's exact assumptions, in this chapter we are discussing this class of test in more generality, and will use the DWH abbreviation.) In panel data modelling, it is widely used as a test for correlated effects, i.e. to investigate the presence of unobserved heterogeneity across units correlated with the explanatory variables. It is based on the contrast between an *OLS* estimator on the model in levels and an *OLS* estimator on the model in differences and it is the common practice to choose between different model specifications. However, the DWH test detects the presence of any possible endogeneity problem (Davidson and MacKinnon, 1989), not necessarily induced by a correlation between the regressors and the individual effects. Almost always in the widespread use of the DWH test for correlated effects in static panel data modelling, the consistency of the *Within Groups* and the *Balestra-Nerlove* estimators under the null is not questioned. However, it might not be the case in presence, for instance, of measurement errors.

An analysis of the causes that lead to a failure of the consistency of an estimator is quite delicate because it is often related to unobservable factors often difficult to detect and to treat properly. Nevertheless, in modelling economic data, it is essential to acquire some further knowledge about different sources of bias and to assess what is the most important problem to control for. The appropriate estimators vary in the different cases. An inaccuracy in the choice of the model specification may lead to unreliable results. In spite of the large related theoretical literature, the problem continues to receive surprisingly little attention in empirical work.

The purpose of this chapter is two-fold. Firstly, it aims to emphasize the misleading inference results one can get by testing for correlated effects without conditioning on the existence or non existence of measurement errors. Secondly, the chapter presents a two-step testing procedure for panel data aiming to distinguish between an endogeneity problem caused by correlation between regressors and individual effects and an endogeneity problem due to measurement errors. (We refer to this as an endogeneity problem because measurement errors (usually) induce correlation between right-hand side observed variables and the disturbances.) The important feature of the methodology is the search for appropriate DWH tests robust to deviations from the classical errors assumption.

Chapter 1 presents the implementation of a robust test for correlated effects, i.e. for the comparison of the *Within Groups* and the *Between Groups* estimators. In this chapter we construct robust tests for the comparison of other pairs of panel data estimators. The motivation underlying the construction of a robust version of a DWH test is that the hypotheses underlying the construction of the statistic (Lemma 2.1 in Hausman, 1978) are often too strong in most of the empirical cases. It is usually a delicate task to rank the different estimators in terms of efficiency. The robust version of the DWH test presented in Chapter 1 is based on the use of an auxiliary regression to estimate a suitably constructed covariance matrix and on the application of the panel data counterpart of the White (1984) robust standard errors. If applicable, the same devices are used for the construction of robust tests for the comparison of the other estimators considered in the two-step testing procedure. This technique allows us to estimate covariances matrices between estimators that cannot be ranked in terms of efficiency. The attractive feature of this methodology for applied works when compared to related techniques in the literature (e.g. Lee, 1996) is that it can be implemented in standard statistical packages.

The chapter is organized as follows. Section 2.2 explains to what extent bi-dimensional data sets may help us to deal with different kinds of unobservable factors and the possible effects of poor attention to the phase of model evaluation. Section 2.3 illustrates a two-step testing procedure for linear panel data models. This may be considered as a guide towards the choice of the most reliable model specification. The statistical performance of this sequential test is assessed using simulated data. The results are contained in Section 2.4. Section 2.5 presents an empirical application of the methodology to a longitudinal data set of travel-to-work areas (TTWAs) in the UK, observed monthly for the period

1996 - 2000. The stock of unemployed and the stock of vacancies, originating in different sides of the labor market, are very likely to be affected by different types of unobservable factors. Our aim is to investigate the importance of correlated effects and measurement errors by analyzing their effects on the estimators of the parameters in a empirical search-matching function model. We suggest the most reliable model specification in the case of study considered. Section 2.6 concludes.

2.2 Unobservable Factors and Panel Data Sets

It is well-known that measurement errors are extremely relevant in data collection. Even though the problem has given rise to a certain amount of theoretical interest, most applied econometric studies do not address this issue. In the analysis of uni-dimensional data sets, i.e. time series or cross sections, classical errors-in-variables models have not been applied widely mainly because it is often not possible to find valid instrumental variables among the variables included in those models. External variables are required in order to identify the structural parameters of interest. Furthermore, measurement errors with different structures and other unobservable factors, like unobserved heterogeneity, may affect our data. Some assumptions of the classical errors in variables model are often not sustainable in many empirical cases.

When a panel data set is at hand it may be possible to handle these issues, since instrumental variables can be found within the model. Moreover, pooling cross sectional and time series observations, the econometrics of panel data offers a variety of different estimators for the same parameter, and the behavior of such estimators in the presence of unobserved factors affecting the data can be analyzed. Therefore, it is possible to acquire some knowledge about the kind of errors of specification involved by checking whether they can actually account for the sign and order of magnitude of the observed discrepancies between estimators. Pursuing this approach, we present a panel data sequential test designed to check for the presence of relevant sources of bias in the data. As the presence of such unobserved factors may invalidate the estimation results, it is essential to use suitable estimators when different sources of bias are discovered. Our procedure checks for the presence of correlated effects and measurement errors and indicates which estimators are likely to give the most reliable results for the analysis of a certain data set.

We focus our attention on the problems of unobserved heterogeneity and measurement errors because in the analysis of the empirical search-matching function framework presented in Section 2.5 unobserved heterogeneity and measurement errors are the only expected important sources of bias. An application of the methodology to panel data sets in different contexts requires that problems arising from the possible presence of other relevant sources of bias, such as sample selection, missing data from attrition, non strict exogeneity, have to be addressed and solved previously. Yet, after controlling for other relevant sources of bias, to distinguish the effects of unobserved heterogeneity and measurement errors remains a subtle issue.

The motivation underlying the implementation of the sequential test is that the standard procedure used in empirical work in order to discriminate between different estimators is often misleading and almost always unsatisfactory. The common practice consists in the application of a DWH test where the two estimators involved in the implementation of the test are the *Within Groups* and the *Balestra-Nerlove*; both *OLS* estimators constructed on different transformations of the data. A possible failure of the consistency of the two estimators under the null, not related to the source of endogeneity being tested, is almost never considered in empirical studies. However, if for instance we are in presence of measurement errors least square estimators not only lose their efficiency but also their consistency. We may end up comparing two inconsistent estimators. Moreover, measurement errors can have different impact using different transformations of the data. For instance, if we use first differences then the bias can be magnified (Griliches and Hausman, 1986). As a consequence, the probability limits of two estimators calculated on different transformations of the data may be different. Thus, in the presence of strong measurement errors, OLS on the model in levels and OLS on the model in first differences (or deviations) would turn out to be different regardless of whether unobserved heterogeneity really matters. We may end up attributing the bias of our results to unobservable individual characteristics while it could be that measurement errors are playing a major role. Consequently the specification of the model adopted could be inappropriate.

2.3 A Two-Step Testing Procedure

We present a sequential testing procedure, which is intended to distinguish the effects of unobserved heterogeneity and measurement errors on the estimators of the parameters in a panel data model in order to choose the most reliable specification. It consists on using different DWH tests in a particular sequence.

The IV approach used to address the problem of measurement errors is not designed to detect non-linear errors in variables (when the explanatory variable with measurement error enters in nonlinear form, e.g. to the square, in the model specification). In this case an IV approach is not valid because the measurement error is not additively separable from the explanatory variables and we cannot find instruments correlated to the explanatory variables but uncorrelated to the new error term (Hausman, Newey and Powell, 1995).

The outline of the procedure is illustrated in Diagram 2.1.

Diagram 2.1: Sequential Procedure

1	WG	vs	IVD
	No UH bias, but ME bias		No UH bias, no ME bias

Reject ↙ Not Reject ↘

2.A	IVD	vs	IVL
	No ME bias, no UH bias		No ME bias, but UH bias

Reject ↙ Not Reject ↘

2.B	WG	vs	OLSL
	No UH bias, but ME bias		UH bias, and ME bias

Reject ↙ Not Reject ↘

Diagnosis A   Diagnosis B   Diagnosis C   Diagnosis D

Diagnosis A → Both unobserved heterogeneity and measurement errors are important sources of bias

Diagnosis B → The most important issue to control for is a measurement problem of the specified type

Diagnosis C → Unobserved heterogeneity plays a major role

Diagnosis D → No evidence of the specified "unobservability"

UH bias: bias due to correlated effects

ME bias: bias due to measurement errors

The different estimators account for one or another (or both) sources of bias.

At a first stage, the *Within Groups* estimator (WG), an *OLS* on the model in deviations from the individual time-means which controls only for correlation between the regressors and the unobserved heterogeneity (UH bias), is compared with a *Generalized Instrumental Variables* estimator on the model in deviations from the individual time-means (IVD), which controls for both measurement error bias (ME bias) and unobserved heterogeneity bias. A significant difference in the two estimators gives evidence of measurement problems in the data. In this case, we investigate if unobserved individual characteristics matter also, by comparing a *Generalized Instrumental Variables* estimator on the model in levels (IVL), which controls only for measurement error bias, with a *Generalized Instrumental Variables* estimator on the model in deviations from the individual time-means, which controls for both measurement error bias and unobserved heterogeneity bias. If we find a significant difference in these two estimators we can infer that unobserved heterogeneity is also an important potential source of bias. If this difference is not significant we can conclude that the most important issue to control for is a measurement problem. On the other hand, if the test performed at the first step gives us insignificant results, we can conclude that measurement bias is not a major issue and we continue our diagnostic procedure comparing *OLS* on the model in levels (OLSL) and *OLS* on the model in deviations. The *OLS* estimator on the model in levels does not control for any kind of bias while the *OLS* on the model in deviations, i.e. the *Within Groups* estimator, rules out the heterogeneity bias. A significant difference in the two estimators gives us evidence of unobserved heterogeneity bias in our data set.

It is worthwhile noting how much the sequence of these tests matters. If we compare at the first step, as is common in empirical work, *OLS* on the model in levels (or the *Between Groups* estimator) and *OLS* on the model in deviations (i.e. the *Within Groups* estimator) we cannot distinguish what is the source of the bias because measurement errors have different effects in models in levels and in deviations from the mean, as previously emphasized.

A robust version of the Hausman test is not directly implemented in standard econometric software. In what follows, we formalize the above outlined methodology in a way that can be implemented in any standard statistical packages.[1] We begin by giving examples of the way panel data sets can be used to construct valid instruments in presence of a specified type of measurement error and by providing insights into the inference implications of the chosen design for the testing procedure.

2.3.1 Robust Testing for Measurement Errors

The first step of the diagnostic procedure (step 1 in Diagram 2.1) requires us to compare the IVD, *IV* estimator on data in deviations from individual time-means, with the WG, *OLS* estimator in data on deviations from individual time-means, in order to investigate the importance of measurement errors.

Particular care is required in the choice of the instruments we use. In order to apply the DWH framework, we have to compare two estimators that are both consistent under the null hypothesis (one more efficient) and one consistent and the other inconsistent under the alternative. If the null hypothesis of non existence of measurement errors is satisfied, the WG is more efficient than an IVD but the instruments have to be chosen in a way such that the consistency of the IVD estimator has to hold when the null hypothesis is violated. Measurement errors may arise under different forms, each of them having different effects on the estimators that are used. It is not possible to construct a reliable test for the presence of arbitrary measurement errors. Panel data sets can help us with this issue because they provide a variety of different types of instrumental variables. However, the choice of the instruments has to be related to a specific structure of the measurement errors in order to guarantee their validity.

Suppose, for instance, we consider the presence of measurement errors with a period

[1] The *Stata 7* routines, that have been written for the empirical application of the methodology presented in Section 2.5 are available upon request.

specific component. Consider the errors-in-variables panel data model

$$y_{it} = \beta x_{it} + \eta_i + \varepsilon_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{2.1}$$

where η_i is the unobserved heterogeneity term, we do not observe x_{it} but x_{it}^* and

$$x_{it}^* = x_{it} + m_{it}, \tag{2.2}$$

$$m_{it} = \theta_t + \xi_{it}, \tag{2.3}$$

$$x_{it}^*, \ \varepsilon_{jt} \quad \text{independent for all } t \text{ and } i \neq j. \tag{2.4}$$

The process of the measurement error, i.e. m_{it}, consists of an i.i.d. time-specific effect with zero mean, i.e. θ_t, and of a white noise component, i.e. ξ_{it}. Substituting, we obtain

$$\begin{aligned} y_{it} &= \beta x_{it}^* + \delta_t + \eta_i + e_{it}, \\ \delta_t &= -\beta\theta_t, \\ e_{it} &= -\beta\xi_{it} + \varepsilon_{it}. \end{aligned}$$

If we define v_{it}, the new composite disturbance component, as

$$v_{it} = \delta_t + e_{it},$$

the basic assumption for the consistency of the *OLS* estimators, i.e. $E(v_{it} \mid x_{it}^*) = 0$, does not hold any longer. x_{it}^* is endogenous because of the component of the measurement error θ_t, that acts in the model through δ_t. This yields

$$Cov(v_{it}, x_{it}^*) = Cov(-\beta\theta_t - \beta\xi_{it} + \varepsilon_{it}, x_{it} + \theta_t + \xi_{it}) \neq 0.$$

Moreover, the problem remains if we transform the model in deviations from the individual time-means to purge the model from the possible correlation between the regressor and the unobserved heterogeneity:

$$Cov(v_{it} - \overline{v_{i.}} \mid x_{it} - \overline{x_{i.}}) \neq 0,$$

where

$$\overline{v_{i.}} = \sum_{t=1}^{T} v_{it}, \quad \overline{x_{i.}} = \sum_{t=1}^{T} x_{it}.$$

Hence, the WG will be inconsistent. However, by virtue of assumption (2.4)

$$E(x_{jt} - \overline{x_{j.}} \mid v_{it} - \overline{v_{i.}}) = 0, \qquad i \neq j.$$

Therefore, if we use as an instrument for the within variation of individual i, the within variation of individual j, we obtain an IVD which is consistent in presence of measurement error having the specified structure. Thus, a DWH type test for the comparison between the WG and the above constructed IVD can be applied and provides reliable results about the presence of measurement errors with a time-specific component.

It is worthwhile noting that it is not possible to distinguish the effects of a measurement error with an individual-specific component from the ones arising from unobserved heterogeneity. In the sequential procedure proposed, this issue is investigated in a second step. If at this further stage we corroborate the importance of unobserved heterogeneity bias, we can use the results of the first step to choose between a two-way and one-way panel data model. Specifically, a rejection of the test at the second stage means that fixed effects may be strong. A rejection of the test at the first stage means that measurement errors with a period specific component are an important issue. The combination of these results leads us to the choice of a two-way panel data specification, i.e. $y_{it} = \beta x_{it} + \eta_i + \delta_t + e_{it}$, instead of the one-way, i.e. $y_{it} = \beta x_{it} + \eta_i + e_{it}$.

Another case that can be worth investigating in the first step is the presence of measurement errors that follow a moving average or autoregressive process. In this context, the instrumental variables have to be chosen according to the structure of the dynamic process. For instance, if we consider measurement errors that follow a moving average process of order one, possible valid instruments for a variable at time t are all the lag values of the same variable of at least lag $(t-2)$. In what follows, lagged instruments are used.

From the discussion above, it is clear that in the first step of the diagnostic procedure it is important to use different IV estimators, each of them robust to a particular structure of measurement error. In fact, if we reject the null at the first stage, given a specific structure of measurement error, this does not mean that we accept the alternative, that is we accept as valid the given structure assumed for the measurement error. To be able to assess which is the most probable structure of the measurement error affecting the data at hand, we should implement some valid tests between an estimator robust to a very flexible measurement error structure against a particular estimator based on a given measurement error structure. Although this is a preliminary step that can be added to the illustrated sequential test, it is not considered in the design of the procedure. The aim of the methodology is to check the robustness of inference results to different assumed structures of measurement error and not to test the specification of the measurement error.

We now turn to the implementation of a robust DWH type test for assessing the presence of measurement errors in variables, that is for the comparison between the WG and an arbitrary IVD. The formulation of such a test using a standard econometric package is not straightforward. Unlike the implementation of a robust DWH type test that will be considered in Sections 2.3.2 and 2.3.3, here we do not directly compare OLS estimators applied on different orthogonal transformations of the data. In other words, it is not just necessary to manipulate the data according to the different transformations, insert the new variables in a auxiliary regression and then run OLS using White (1984) robust standard errors. The procedure needs to be adjusted. Some preliminaries are needed.

In static models, the most efficient *Generalized Instrumental Variables* estimator is obtained by projecting the variables to be instrumented onto the space generated by the instruments. This is a case where the instruments are orthogonal to the initial errors and especially correlated with the initial regressors. It can be shown that, given the properties of the projection matrix, when estimating coefficients it is equivalent to running OLS in a regression where the regressors are the projected variables.[2]

Consider model (1.18). First transform the data according to the *within groups* transformation, i.e. deviations from the mean

$$y_i^* = X_i^* \beta + \mu_i^*, \qquad i = 1, ..., N.$$

Then choose the instrumental matrix according to the structure of the measurement errors we want to test for, say Z. Project the variables we want to instrument in the space generated by Z

$$\widetilde{X_i^*} = P_Z X_i^*$$

where

$$P_Z = Z(Z'Z)^{-1}Z'.$$

If we assemble the data in a $NT \times 1$ vector of dependent variables, Y^* and in a $NT \times NK$ matrix of regressors $\widetilde{X^*}$[3]

$$\widehat{\beta}_{ivd} = (\widetilde{X^*}'\widetilde{X^*})^{-1}\widetilde{X^*}'Y^*.$$

For the single individual i, construct the system

$$\begin{cases} \widetilde{y_i^*} = \widetilde{X_i^*}\beta + \widetilde{\mu_i^*} \\ y_i^* = X_i^*\beta + \mu_i^*. \end{cases}$$

[2]For further details and an extensive discussion on these issues see Bowden and Turkington (1984).

[3]Recall that this is only a different reformulation of the IV estimators because the projection matrix is idempotent, i.e. $P_Z' = P_Z$ and $P_Z' P_Z = P_Z$,

$$\widehat{\beta}_{ivd} = (X^{*'} P_Z X^*)^{-1} X^{*'} P_Z Y^* = (X^{*'} P_Z' P_Z X^*)^{-1} X^{*'} P_Z' P_Z Y^* = (\widetilde{X^*}'\widetilde{X^*})^{-1}\widetilde{X^*}'Y^*.$$

Estimating by OLS the first group of equations ($i = 1, ..., N$), i.e. the ones in levels, we obtain the IVD, i.e. $\widehat{\beta}_{ivd}$. Estimating by OLS the second group ($i = 1, ..., N$), i.e. equations in deviations, we obtain the WG, i.e. $\widehat{\beta}_{wg}$. However, the use of an artificial regression, as the one used in Appendix 4.1, is not suitable. In that case, because the *between groups* transformation and the *within groups* transformations are orthogonal, the variance covariance matrix in the auxiliary model was block-diagonal. It was then estimated using the White (1984) robust estimators. When different transformations of the data are used, the structure of the variance covariance matrix in the auxiliary regression model can be more complicated. The fact that the equation sets are not orthogonal is not taken into consideration and the White's estimators are not robust to the presence of inter-groups correlation. The use of a Newey-West robust OLS estimator would not help either. The variance covariance matrix exhibits a pattern of cross sectional dependence (i.e. particular form of non stationarity persistent when N goes to infinity) that is not supported by these estimators. Therefore, a consistent estimator for the variance of the difference of the two estimates needs to be constructed step by step. Appendix 1.2, Part I contains further clarification of these points and implements an appropriate procedure.

2.3.2 Robust Testing for Correlated Effects without Measurement Errors

If the results of the test at the first stage provide evidence that measurement errors can be neglected, the widespread practice to test for correlated effects using the comparison between OLS in levels, i.e. the OLSL, and in deviations, i.e. the WG, is correct (step 2.B in Diagram 2.1). A standard Hausman test can be applied. However it is recommended to use a robust version of the test in order to control for the possible presence of non spherical disturbances, as it is explained in Chapter 1. The details of the construction of such a test are also contained in Chapter 1 (Section 1.3). The chosen OLS estimator on the model in levels is the *Between Groups* estimator, hereafter BG.

2.3.3 Robust Testing for Correlated Effects with Measurement Errors

If the results of the test at the first stage provide evidence of important measurement error bias, testing for correlated effects using the comparison between *OLS* in levels and in deviations is not correct and may lead to unreliable results. IV estimators can be used instead.

An implementation of the DWH test for the presence of correlated effects in presence of measurement errors consists in comparing the IVD constructed in the first step of the procedure with the same *Generalized Instrumental Variables* estimator on the model in levels (IVL) (step 2.A in Diagram 2.1). If correlated effects are present, the IVL, which controls only for a specific structure of the measurement error is not consistent while the IVD remains consistent because the transformation of the data used purges the model from the effects of individual-specific components. Therefore, the DWH framework can be applied. Note that, as in the first step, it is important to construct a number of different tests for correlated effects. By analyzing the results of a combination of the tests of first and second stage, it is possible to choose the most reliable (robust) model specification.

A DWH test for correlated effects in presence of measurement error can be set out as follows. As in the comparison between the BG and the WG (Section 2.3.2), we deal with two different estimators that are obtained applying the same estimation method on data transformed in different ways. In Section 2.3.2 we choose the BG as an *OLS* estimator for the model in levels. In this section we choose an IV estimator on the model in averages (*between groups* transformation) as IV estimator for the model in levels. This choice allows us to deal again with two orthogonal transformations of the data. Thus, the use of an artificial regression will lead to the desired outcome. We obtain the IVL applying the IV methodology to the average equations in levels and the IVD applying the IV methodology to the equations in deviations from the mean.

Model (1.18), a system of T equations in levels, can be transformed into $(T-1)$

equations in deviations and one in averages. We obtain

$$\begin{cases} y_i^* = X_i^*\beta + \mu_i^* \\ \overline{y_i} = \overline{X_i}\beta + \overline{\mu_i}. \end{cases}$$

Estimating by IV the first group of equations, i.e. the ones in deviations from individual time means, we obtain the IVD, i.e. $\widehat{\beta}_{ivd}$. Estimating by IV the average equation we obtain the IVL, i.e. $\widehat{\beta}_{ivl}$.

Let

$$\beta_{ivl} = plim\left(\widehat{\beta}_{ivl}\right)$$

and

$$\beta_{ivd} = plim\left(\widehat{\beta}_{ivd}\right).$$

Rewrite the system as

$$\begin{cases} y_i^* = X_i^*\beta_{ivd} + \mu_i^* - X_i^*\beta_{ivl} + X_i^*\beta_{ivl} \\ \overline{y}_i = \overline{X}_i\beta_{ivl} + \overline{\mu}_i. \end{cases}$$

Rearranging, we obtain

$$\begin{cases} y_i^* = X_i^*\left(\beta_{ivd} - \beta_{ivl}\right) + X_i^*\beta_{ivl} + \mu_i^* \\ \overline{y}_i = \overline{X}_i\beta_{ivl} + \overline{\mu}_i. \end{cases}$$

Call

$$Y_i^+ = \begin{pmatrix} y_i^* \\ \overline{y}_i \end{pmatrix}, \quad W_i^+ = \begin{pmatrix} X_i^* & X_i^* \\ 0 & \overline{X}_i \end{pmatrix},$$

$$\beta^+ = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_{ivd} - \beta_{ivl} \\ \beta_{ivl} \end{pmatrix}, \quad \mu_i^+ = \begin{pmatrix} \mu_i^* \\ \overline{\mu}_i \end{pmatrix}.$$

The augmented auxiliary model would be

$$Y_i^+ = W_i^+ \beta^+ + \mu_i^+, \qquad i = 1, ..., N. \tag{2.5}$$

Estimating the model by IV, we obtain directly the variance of the difference of the two estimators in the upper left part of the covariance matrix of β^+. Unfortunately, in standard econometric packages White's robust variance estimator may not be implemented for IV panel data estimation. In this case, a practical possible solution can be to obtain the IV estimators as OLS estimators on a further transformed model, as was necessary and explained in Section 2.3.1. After repeating the same steps for the construction of another artificial regression with these new transformed equations and estimating consistently the variance of the OLS estimators, once again a Wald test will allow us to investigate the presence of correlated effects in a reliable way. This approach is pursued in Appendix 1.2, Part II. The following Lemma is proved.

Lemma 15 Given model (2.5),

$$\widehat{\beta}_1 = \widehat{\beta}_{ivd} - \widehat{\beta}_{ivl}, \tag{2.6}$$

$$Var(\widehat{\beta}_1) = Var\left(\widehat{\beta}_{ivd} - \widehat{\beta}_{ivl}\right), \tag{2.7}$$

An appropriate estimator $\widehat{Var}(\widehat{\beta}_1)$ *consistently estimates* $Var(\widehat{\beta}_1)$.

2.4 A Monte Carlo Experiment

In this section we use simulated data in order to assess the statistical performance of the proposed procedure. In the bunch of diagnostic tests provided by standard statistical packages there is scarce attention to the fact that they are repeated test procedures using the data at hand. This yields consequences both on the properties of the chosen estimator and on the subsequent inference (see Judge and Bock, 1978). An exhaustive treatment of these issues is far beyond the scope of this chapter. Nevertheless, we present a Monte

Carlo experiment aiming to examine how the critical regions of the test depends on the fact that it is a two stage procedure.

The postulated data generation process is the following. We consider the linear panel data regression model

$$y_{it} = \varphi x_{it} + \pi z_{it} + \eta_i + \delta_t + u_{it}, \quad i = 1, ..., N; \; t = 1, ..., T,$$

where the disturbance term consists of three independent components: a unit-specific effect, η_i, a time-specific effect, δ_t, and a white noise component, i.e. u_{it}. This context can be considered as a case where there is the joint occurrence of individual effects and measurement errors with a time specific component. We assume that z is an exogenous variable and we generate x using the same structure for the disturbance term

$$x_{it} = w_{it} + v_i + \vartheta_t + \varepsilon_{it}, \tag{2.8}$$

where w is an exogenous variable. The random variables (η_i, v_i) and (δ_t, ϑ_t) are constructed as draws from a bivariate normal distribution with correlation rho_1 and rho_2 respectively. Each of these takes values of $(0, 0.05, 0.10, \ldots 0.95)$.

The values for the exogenous variables and the range of values for the parameters are taken from the analysis of the empirical case of study presented in Section 2.5. Using UK data, the following model is estimated.

$$\log M_{it} = a + \alpha \log U_{it} + \beta \log V_{it} + e_{it}, \qquad i = 1, ..., 276; \; t = 1, ..., 63,$$

where M is the number of matches/hirings, U and V are the stocks of unemployed and vacancies, a is a constant term, e indicates a disturbance term. Different estimation methods have been applied. The different estimates of the coefficient of the unemployed, α, and for the vacancies, β, have been used in the simulation experiment for φ and π respectively. They take values of $(0.5, 0.4; \; 0.5, 1.5; \; 0.4, 0.7; \; 0.3, 1.4)$. Also, the best prediction for the stock of vacancies is

$$\log V_{it} = 1.2 \log NV_{it}, \qquad i = 1, ..., 276; \ t = 1, ..., 63$$

where NV is the number of monthly notified vacancies (flow variable). In our experiment design, the real values for U and NV have been used as exogenous variables, i.e. respectively z and w. The endogenous variable, V, i.e. x, has been constructed using the structure (2.8).

The sample size available in the empirical case of study, i.e. $N = 276$, $T = 63$, is considered. The experiment is repeated 5000 times for each combination of rho_1 and rho_2. Four hundred data sets have been generated. For each repetition, the test statistics used at the first step, T_1 and at the second step, T_2, has been calculated as explained in the development of the procedure in Section 2.3 and the instrumental variables chosen are the ones used in the empirical case of study in Section 2.5, i.e. we use as instrument for a variable at time t the value of the same variable at time $(t-3)$. We summarize as follows.

Let us indicate

$$\gamma = \begin{pmatrix} \varphi \\ \pi \end{pmatrix},$$

and $\widehat{\gamma}_{ivd}$ the IVD, $\widehat{\gamma}_{wg}$ the WG, $\widehat{\gamma}_{ivl}$ the IVL and $\widehat{\gamma}_{bg}$ the chosen OLSL, that is the *Between Groups* estimator.

At the first stage we test the null hypothesis

$$H_0 : rho_2 = 0$$

against the alternative

$$H_1 : rho_2 \neq 0,$$

using the statistic

$$T_1 = \widehat{q}_1' \widehat{V}\left(\widehat{q}_1\right) \widehat{q}_1,$$

where $\widehat{q}_1 = \widehat{\gamma}_{ivd} - \widehat{\gamma}_{wg}$ and $\widehat{V}(\widehat{q}_1)$ is a consistent estimator of the variance (see Appendix 1.2).

At the second stage we test the null hypothesis

$$H_0 : rho_1 = 0$$

against the alternative

$$H_1 : rho_1 \neq 0,$$

using the statistic

$$T_2 = I(T_1 > c_1)T_{2R} + I(T_1 \leq c_1)T_{2A},$$

where c_1 and c_2 are specified critical values, $I(a)$ is the indicator function which takes value equal to one if the event (a) is realized, zero otherwise.

$$T_{2R} = \widehat{q}_2'\widehat{V}(\widehat{q}_2)\widehat{q}_2,$$

where $\widehat{q}_2 = \widehat{\gamma}_{ivd} - \widehat{\gamma}_{ivl}$ and $\widehat{V}(\widehat{q}_2)$ is a consistent estimator of the variance.

$$T_{2A} = \widehat{q}_3'\widehat{V}(\widehat{q}_3)\widehat{q}_3$$

where $\widehat{q}_3 = \widehat{\gamma}_{wg} - \widehat{\gamma}_{bg}$ and $\widehat{V}(\widehat{q}_3)$ is a consistent estimator of the variance.

The sequential procedure proposed in this chapter consists on performing T_1 followed by T_2.

Let us consider the size first. In principle, the alpha level for multiple tests could be corrected by using the Bonferroni Inequality (Bonferroni, 1936). It is an inequality from probability theory that gives us an upper bound for the probability that one or more separate tests will lead to a specified type I error. It can be stated as follows. For any sets of events A_K,

$$\Pr(A_1, A_2, ..., A_K) \geq 1 - \sum_{i=1}^{K} \Pr\left(A_i^C\right),$$

c*	r*(%)	r^1(%)	r^2(%)	r^3(%)	r^4(%)	95% precision(%)
0.58	56.25	55.61	55.79	57.04	56.08	+/- 1.2
1.39	25	24.03	24.02	25.81	25.34	+/- 1.4
2.77	6.25	6.20	6.28	6.33	6.32	+/- 1.2
4.61	0.01	0.00	0.01	0.04	0.03	+/- 0.8

Figure 2-1: Theoretical versus empirical frequencies

where A_i is an event and A_i^C its complement.

Rearranging

$$\sum_{i=1}^{K} \Pr\left(A_i^C\right) \geq 1 - \Pr(A_1, A_2, ..., A_K).$$

Therefore, if we fix the significance level of the joint test as α, the size of each separate test, say α', has to satisfy the following inequality

$$K\alpha' \geq \alpha.$$

Thus, the Bonferroni inequality implies that we should set

$$\alpha' = \frac{\alpha}{K}$$

if we wish the probability that a type I error occurs to be no more than α in the multiple test procedure. In our two stage test, if we choose the alpha level of the overall procedure equal to 0.05, the size of each test should be set at 0.025. However, the Bonferroni inequality does not require any knowledge about the structure of dependence between the test statistics used at the different steps. This is a strength but at the same time a weakness of the principle. If the dependence is further investigated, the inference can be sharpened.

In order to study the structure of dependence between T_1 and T_2, we consider their empirical distribution function. The asymptotic marginal distribution of both test statistics is a chi-squared with two degrees of freedom (χ_2^2). Figure 2-1 reports in the first column the four quantiles c^* that leave a probability of 0.75, 0.50, 0.25, 0.10 in the tail of a χ_2^2 distribution, in the second column r^*, the square of $P(\chi_2^2 > c^*)$, that is the *p-value* that would be obtained asymptotically if the two test statistics were independent, and in the remaining four columns the estimated frequencies obtained when φ and π are set to the four pairs of values used in the design of the experiment: $(0.5, 0.4)$, $(0.5, 1.5)$, $(0.3, 1.4)$, $(0.4, 0.7)$. They are indicated by $r\hat{}1$, $r\hat{}2$, $r\hat{}3$, $r\hat{}4$ respectively. Under H_0, the simulation experiment suggests that the hypothesis of independence between the two test statistics cannot be rejected. The last column shows the Monte Carlo precision as a 95% confidence interval. As a matter of fact, for different c^*, we get that the estimated frequency of rejecting both hypotheses is roughly equal to the square of $P(\chi_2^2 > c^*)$. Therefore, the significance levels of the two tests can be fixed independently such that their product is equal to the desired alpha level of the sequential procedure.

Let us now turn our attention to power considerations. The property of independence of the test statistics is not preserved under any alternative hypothesis.[4] We investigate the performance of the test in terms of power by fixing the alpha level of the sequential test equal to 0.05, the one of each single test equal to $\sqrt{0.05}$ and calculating the probability that both T_1 and T_2 fall in the critical region for each combination of rho_1 and rho_2. The resulting triples of points have been plotted in a three-dimensional space. Figure 2-2 shows the power function when $\varphi = 0.5$ and $\pi = 0.4$.[5]

For low levels of rho_1 and rho_2 (in the range of $(0.05, 0.5)$) the power of the test is not very high (always below 60%) but it increases sharply as soon as either rho_1 or rho_2 takes values greater than 0.55. It is worthwhile noting that, as the relevance of the procedure is mainly empirical, we are interested in getting a good performance of the test in terms of power for high levels of rho_1 or rho_2. In other words, the procedure should help us to avoid the use of badly biased estimators, i.e. we need to be able

[4]We do not report all the results for all combinations of values of rho_1 and rho_2 for brevity. They are available upon request.

[5]Similar pictures can be obtained by plotting the power function for the other values of the parameters.

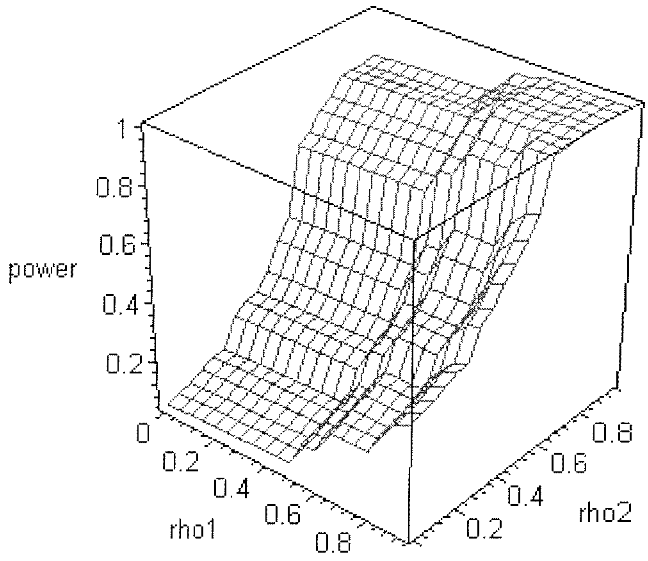


Figure 2-2: Power function: three dimensional plot of points

to reject almost surely the hypothesis of independence when the level of correlation is high. Therefore, the performance of the test is satisfactory. However, an analysis of the graph reveals that the power function is more sensitive to deviations from the hypothesis of absence of measurement errors in variables ($rho_2 = 0$) than from the hypothesis of non correlation between regressors and individual effects ($rho_1 = 0$). This asymmetric behavior in the performance of the multiple test in terms of power should be taken into consideration when fixing the significance levels of the single tests. The power function illustrated in the graph refers to the case when the size of the two separate tests is the same. A different case can lead to a more balanced performance. Further research may be needed. For instance, the construction of a sequential test that assesses the importance of correlated effects at the first step and of measurement errors at the second step could be considered. This possibility has not been explored here because in the empirical context the test is applied (empirical job matching function framework in Section 2.5) there are a priori reasons to assume the presence of measurement errors ($rho_2 \neq 0$) and therefore the expected performance of the test is adequate. As noted in Section 2.2, a different application of the test requires the procedure to be tailored to the new context and the statistical performance of the test may vary in each case.

Finally, it is of interest to report some results on the estimated probability that the sequential test detects the postulated major source of endogeneity. They are illustrated in Figure 2-3. They refers to the case when $\alpha = 0.5$ and $\beta = 0.4$.

Table 1 reports the probability p_{12} defined as

$$p_{12} = \Pr(T_1 > c^*, T_2 \leq c^*)$$

when $rho_2 = 0.9$ and rho_1 takes all the twenty values between 0 and 0.95. $rho_1\hat{}$ is the estimated level of correlation between the regressors x_{it} and η_i and $rho_2\hat{}$ is the estimated level of correlation between x_{it} and δ_t. If we reject T_1 and we cannot reject T_2, we are expected to use an instrumental variable estimator using the model in levels. The table shows that the probability to get such an outcome from the application of the sequential test is a decreasing function of the level of rho_1. According to what is desirable, the probability of choosing an estimator in levels is a decreasing function of the importance of unit-specific effects.

Table 2 reports the probability p_{21} defined as

$$p_{21} = \Pr(T_1 \leq c^*, T_2 > c^*),$$

when $rho_1 = 0.9$ and rho_2 takes all the twenty values between 0 and 0.95. The same definition for $rho_1\hat{}$ and $rho_2\hat{}$ in Table 1 applies. If we cannot reject T_1 and we reject T_2, we are expected to use a least square estimator after having transformed the variables in deviations from the mean. Also in this case the table shows a desirable result. The probability of obtaining such an outcome from the application of the sequential test is a decreasing function of the level of rho_2. This implies that the probability of requiring an estimator in deviations is a decreasing function of the importance of factors different from individual-specific effects (that can be magnified when transforming the variables). Sim-

Table 1:
rho2=0.9; rho1 varies

rho1^	rho2^	p12
0.001	0.517	0.45
0.026	0.508	0.45
0.051	0.515	0.42
0.094	0.509	0.38
0.111	0.501	0.33
0.152	0.51	0.3
0.179	0.51	0.22
0.202	0.51	0.13
0.228	0.505	0.1
0.268	0.515	0.09
0.293	0.503	0.09
0.319	0.52	0.06
0.347	0.512	0.03
0.384	0.514	0.03
0.413	0.513	0.03
0.44	0.51	0.02
0.476	0.509	0.02
0.501	0.51	0.01
0.53	0.511	0.01
0.555	0.522	0.01

Table 2:
rho1=0.9; rho2 varies

rho1^	rho2^	p21
0.525	0.002	0.44
0.525	0.035	0.43
0.522	0.047	0.42
0.534	0.091	0.33
0.523	0.109	0.23
0.53	0.142	0.22
0.527	0.152	0.21
0.529	0.203	0.18
0.528	0.219	0.15
0.525	0.262	0.15
0.528	0.286	0.12
0.528	0.307	0.1
0.53	0.33	0.09
0.525	0.364	0.05
0.529	0.405	0.02
0.526	0.43	0.01
0.53	0.453	0.01
0.534	0.479	0.01
0.525	0.517	0.01
0.527	0.539	0.01

Figure 2-3: Simulated performance

ilar tables can be obtained for the other points in the parameters space. The qualitative performance of the test is unchanged. Taking the results as a whole, the application of the sequential test implemented seems to give some appropriate directions for the choice of the most reliable model specification.

2.5 The Empirical Framework

The main purpose of our analysis is to investigate the importance and typology of unobservable factors that, as is well-known, affect two of the variables frequently used in labor market research, namely the stock of unemployed and the stock of vacancies. We use a job search matching framework. The matching function is a modeling device that summarizes the search process that eventually brings workers and firms into productive matches. The simplest form of the matching function (Pissarides, 2000) is

$$M = m(U, V), \tag{2.9}$$

where M is the number of jobs formed at any moment in time, U is the number of unemployed workers looking for work and V is the number of vacant posts. In recent years the concept of a matching function has been extensively used to explain the working of the labor market.[6] However the majority of the studies are theoretical. Moreover, while the theoretical emphasis is typically on the behavior of microeconomic units, most of the empirical applications have used aggregate data. In recent years, a small number of empirical studies investigating the empirical relevance of the concept at less aggregate levels have been produced. The central question addressed is whether the matching function exhibits constant returns to scale, which is one of the basic assumptions in the theoretical literature. Although our primary aim is neither an empirical test of such a stylized relation nor an inspection of the returns to scale exhibited, we will comment

[6]See, for instance, Petrongolo and Pissarides (2001) for a review.

on these issues while analyzing the results obtained for the data set under investigation. Our main focus is to emphasize the relevance of the sequential test in applied work by showing that a neglect or improper treatment of unobservable factors, like testing for correlated effects without conditioning on the existence or non existence of measurement errors, can lead to extremely unreliable inference results.

2.5.1 Description of the Data and Definition of the Variables

A longitudinal data set of travel-to-work areas (TTWAs) in the UK observed monthly for the period 1996-2001 has been used. All data are available from the National On-line Manpower Information Service (NOMIS) located at the University of Durham. In the United Kingdom the travel-to-work-areas are considered the standard approximations to self-contained labor markets, i.e. areas in which people both live and work. They are geographic regions with a minimum of 3500 working people where at least 75% of those living (working) in the area should also work (live) there. We consider the most recent TTWAs' definition, based on the journey to work statistics from the 1991 Census of Population. A total of 297 TTWAs are designated in England, Scotland and Wales. Only areas with non missing values are included in the sample used for estimation, reducing the cross section dimension from 297 to 277 areas. Furthermore, because London looks like an outlier compared to the other TTWAs, we performed the analysis without London, so that we ended up with 276 areas.

The NOMIS database contains detailed informations from both sides of the labor market. Unemployed and vacancies data collected by Nomis are registration data provided by local employment agencies (Job Centres). They are administrative data that have the advantage of being readily available on a regular basis, at high frequencies, and at a very disaggregate spatial level. Temporal aggregation is an important issue in the estimation of a matching function because it involves estimating flows from stock variables. High-frequency data can in principle mitigate this bias. Also aggregation over space can be misleading. The estimation of a matching function combining cross sectional and time series observations where the cross section units are the regions may still lead to unreliable results. If the regions have a different size and matching does not take place under constant return to scale, estimates may be affected by a spurious scale effect (regional boundaries do not coincide with labor market boundaries). Furthermore, the different local labor markets may be interrelated (for instance, because of common shocks). By working with TTWAs, a quite fine level of spatial disaggregation (the finest available for the vacancies data) the further sources of bias should also be mitigated. We use as a proxy for the total number of unemployed the monthly count of claimants who are claiming unemployment benefits on the unemployment count date (second Thursday of each month) and as a proxy for the jobs that are vacant the monthly stock count of notified vacancies that have not been filled at the end of the previous month. The number of vacancies that are monthly filled by job seekers is our measure of total matches/hirings. We do not arbitrarily adjust the data following, for instance, the correction proposed by Coles and Smith (1996). It is believed that the Job Centres numbers represent approximately one-third of the vacancies and one-quarter of the placings in a TTWA. It is certainly true that registered vacancies are only one channel from which firms recruit personnel but we are not aware of the exact proportions. However, if for instance the ratio between measured number of vacancies (or hires) and true number of vacancies (or hires) is not constant across areas, we would introduce a systematic measurement error by correcting the data this way. Our approach is to work with the raw data and try to control for the most important unobservable factors affecting our data set.

2.5.2 Empirical Analysis and Results

We start by considering a standard Cobb-Douglas specification of the matching function (2.9) in log-linear form:

$$\log M_{it} = a + \alpha \log U_{it} + \beta \log V_{it} + \eta_i + u_{it}, \qquad i = 1, ..., N, \quad t = 1, ..., T. \tag{2.10}$$

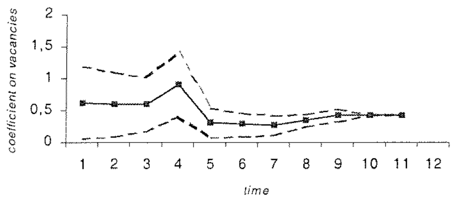
We indicate by M_{it} the number of hirings in area i during month t; U_{it} and V_{it} the stocks of registered unemployed and of vacancies in area i at the beginning of period t; η_i is a TTWA fixed effect controlling for regional characteristics, including the size of the TTWA; and u_{it} is a white noise error term. The constant term, a, is meant to capture the efficiency of the matching technology. In this framework, α and β are the elasticities of hirings to unemployment and of hirings to vacancies respectively. Chart 1 contains the graphs plotting different panel data estimates of α and β calculated recursively by adding six months periods. Assuming normality of the estimators, we draw the bands corresponding to a confidence interval of 95%. The hypothesis of constancy is not rejected. If we neglect the odd values of the estimators in the first two years, perhaps affected by administrative changes in the way data had been collected,[7] both elasticities appear to be constant in all the models adopted. Therefore the restrictive Cobb-Douglas specification does not seem to be rejected by the data.[8]

[7]This may be related to the government change in 1997.

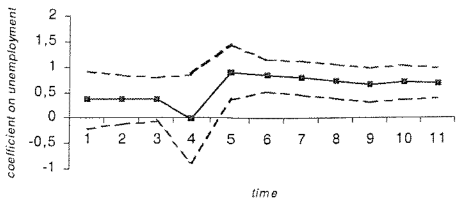
[8]A more flexible transcendental logarithmic model of the matching technology has also been analyzed. The results go beyond the main purpose of this thesis. Therefore they are not reported here, but they are available on request.

Chart 2.1: Rolling Elasticities Estimators

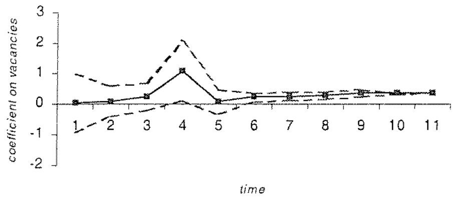
—— $\widehat{\beta}$ $(\widehat{\alpha})$, - - - - - $\widehat{\beta} \pm 2\sigma$ $(\widehat{\alpha} \pm 2\sigma)$



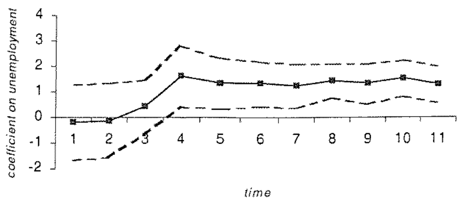
OLS estimator



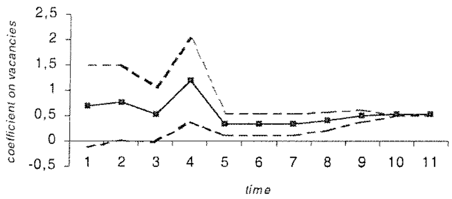
OLS estimator



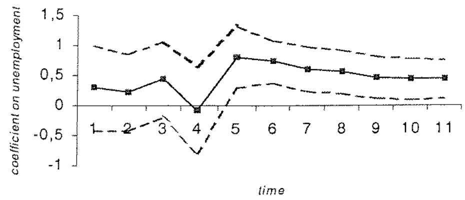
WG estimator



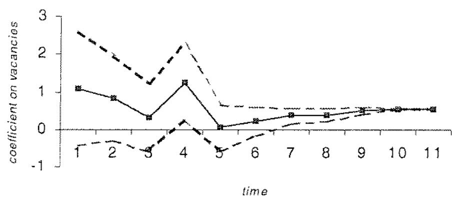
WG estimator



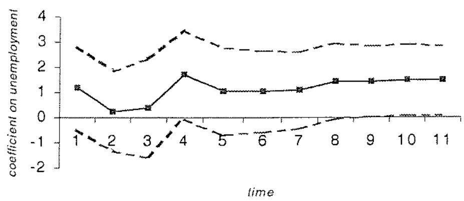
IVL estimator



IVL estimator



IVD estimator



IVD estimator

Using the sequential procedure presented in Section 2.3, we investigate the importance of unobserved heterogeneity and measurement errors affecting the unemployed and vacancies data by comparing different panel data estimators of the coefficients of the stocks of unemployed and vacancies. Table 3 reports the results for the different panel data estimators involved in the development of the sequential testing procedure. In our analysis we use IV estimators that control for autocorrelation in the process of the measurement error. We use as instrument for a variable at time t the value of the same variable at time $(t-3)$. This is reasonable from a logical point of view because the instrument is the value of the variable at the end of the previous quarter and from a technical prospective because it allows us to control for the presence of measurement errors that follow a moving average process of order one, using differences, or one and two, using levels.

Table 3: Model (2.10), Estimation Results*

Dependent Variable: Log Filled Vacancies

	OLSL	WG	IVL	IVD
Log vacancies	0.4295	0.3502	0.5171	0.5425
	(98.48)	(55.47)	(101.27)	(59.11)
Log unempl	0.6943	1.4224	0.4450	1.5323
	(4.42)	(1.99)	(2.64)	(1.63)
Const	−2.7228	−7.8714	−1.2091	−9.4665
	(−2.28)	(−1.46)	(−0.95)	(−1.34)

*- t-test in parentheses,

-robust standard errors are used,

- TTWA fixed effects included.

We begin by applying the sequential test separately for the two coefficients. For sake of clarity, we reproduce Diagram 2.1 for reference.

Diagram 2.1: Sequential Procedure

WG	vs	IVD
No UH bias, but ME bias		No UH bias, no ME bias

Reject ↙ Not Reject ↘

IVD	vs	IVL
No ME bias, no UH bias		No ME bias, but UH bias

WG	vs	OLSL
No UH bias, but ME bias		UH bias, and ME bias

Reject ↙ Not Reject ↘ Reject ↙ Not Reject ↘

Diagnosis A Diagnosis B Diagnosis C Diagnosis D

Diagnosis A → Both unobserved heterogeneity and measurement errors are important sources of bias

Diagnosis B → The most important issue to control for is a measurement problem of the specified type

Diagnosis C → Unobserved heterogeneity plays a major role

Diagnosis D → No evidence of the specified "unobservability"

UH bias: bias due to correlated effects

ME bias: bias due to measurement errors

The different estimators account for one or another (or both) sources of bias.

At the first stage of the procedure, a robust DWH test for the equality of WG and IVD gives us evidence of strong measurement errors for both unemployed and vacancies data. The null hypothesis of equality of the two estimators is strongly rejected in both cases ($\chi_1^2 = 23.21$, $p = 0.00$; $\chi_1^2 = 29.03$, $p = 0.00$ for unemployed and vacancies respectively).[9]

In the second stage of the procedure we investigate the relevance of correlated effects. Area specific effects on hirings may arise, for instance, as a result of variations in the matching technology across TTWAs. These technological differences are likely to be correlated with area size and hence with the area level of unemployed and vacancies. As explained in Section 2.3, a test for correlated effects in presence of measurement errors in variables consists of comparing IVL and IVD. Applying a robust DWH test, we cannot reject the hypothesis of equality of the two estimators for the unemployed coefficient ($\chi_1^2 = 1.18$, $p = 0.2773$), but we reject this hypothesis for the vacancies coefficient ($\chi_1^2 = 46.04$, $p = 0.00$). Different estimation methods controlling for a specific kind of bias show different effects on the coefficient of the two variables: our results suggest that vacancies and unemployed data are contaminated by unobservable factors of different types. We can conclude that area specific unobservable factors, such as local policies towards the demand or the supply side of the labor market, influence the stock of vacancies but play only a minor role in the determination of the number of unemployed. However, measurement errors remain an important issue to control for. In fact, if we perform the sequential testing procedure jointly for the coefficients of the stock of unemployed and vacancies, we reject the null hypothesis at the first step ($\chi_2^2 = 30.89$, $p = 0.00$) but we cannot reject the null hypothesis at the second step ($\chi_2^2 = 4.01$, $p = 0.1347$). Therefore the most reliable estimator is the IVL.

The lack of a rigorous statistical analysis may lead to completely different results. Firstly, if we follow the common practice and we use a standard DWH test to choose between the random effects and the fixed effects model (comparison between the estimates

[9] Strictly, the analysis in Chapter 1 suggests that we can only interpret these two tests separately if there is measurement error only in one variable, and the two variables are uncorrelated. However, the correlation of 0.12 here seems limited enough to support the interpretation.

in columns one and three of Table 3), we are forced to use the second one (the difference of the estimates is systematic: $\chi_2^2 = 64.13$, $p = 0.00$). The result is that the values of the two elasticities are in contrast to the underlying matching function economic theory (the hypothesis of constant returns to scale is rejected $\chi_1^2 = 25.92$, $p = 0.00$). Furthermore, an interpretation of the results based on a visual inspection of the table may also be misleading. Table 3 shows that while for the vacancies coefficients the discrepancies between different estimators on the same transformation of the data (OLSL versus IVL and WG versus IVD) are higher than the ones between the same estimators on different transformations of the data (OLSL versus WG and IVL versus IVD), the unemployed coefficients show opposite and more marked patterns. There is a huge difference between OLSL and WG and between IVL and IVD. Therefore the more immediate interpretation is to consider the bias due to measurement errors to be the most important problem for the vacancies coefficient, and the bias due to correlation between the regressors and the unobserved heterogeneity as the most important one for the unemployed coefficient. As explained above, this interpretation is not confirmed by the DWH tests. For instance, the particularly marked patterns of the unemployed elasticities may be due not only to the effects of area-specific factors that are neglected in the estimators of the model in levels, but also to the presence of strong measurement errors whose effects are magnified in the models in deviations, as is confirmed by the application of the diagnostic procedure.

In order to investigate the robustness of the results to different structures of the measurement error, we estimate the model with yearly time dummies. It is worth noting that the introduction of time dummies is usually used to capture time components of a (efficiency of the matching function) but it allows also for the effects of unobservable factors constant across areas and changing over time. A panel data model which controls for time differences in the technology of matching and one which assumes measurement errors with a time component in an additive structure have the same specification. Either way we investigate whether differences in the intercept may account for differences in the previous estimators (slope). The effects of year specific unobservable factors common to all areas can be relevant in the framework we are considering because it is very likely that the stocks of unemployed and vacancies are influenced by nation-wide policies different over time. We estimate the following model:

$$\log M_{it} = a + \alpha \log U_{it} + \beta \log V_{it} + \eta_i + \delta_t + u_{it}, \qquad i = 1, ..., N, \quad t = 1, ..., T. \quad (2.11)$$

where we use the same notation of Model (2.10). In addition, δ_t is a time specific effect controlling for the influence of temporal factors constant over areas. Table 4 reports the corresponding results from the different panel data methods of estimation presented in Table 3.

Table 4: Model (2.11), Estimation Results*

Dependent Variable: Log Filled Vacancies

	OLSL	WG	IVL	IVD
Log vacancies	0.4299	0.3509	0.5173	0.5434
	(98.51	(55.49)	(101.30)	(59.05)
Log unempl	0.6635	0.0010	0.4017	1.4881
	(4.14)	(0.00)	(2.35)	(0.49)
Const	−2.7442	3.0251	−0.9392	−9.7612
	(−1.36)	(0.21)	(−0.45)	(−0.41)

*- t-test in parentheses,

- robust standard errors are used,

- TTWA fixed effects and year dummies included.

Applying the diagnostic procedure detailed in Section 2.3 separately for the coefficients of the unemployed and vacancies, the hypothesis of equality of WG and IVD is rejected at the first stage for both variables ($\chi^2_1 = 27.64$, $p = 0.00$; $\chi^2_1 = 35.79$, $p = 0.00$ for unemployment and vacancies respectively). At the second stage, we cannot reject the hypothesis of equality of IVL and IVD for the unemployed coefficient ($\chi_1^2 = 1.11$, $p = 0.2921$) but we do reject this hypothesis for the vacancies coefficient ($\chi_1^2 = 94.09$, $p = 0.00$). Therefore the results of the robust DWH tests are not different from the ones obtained for Model (2.10), both in the first and in the second stage. Also, when the tests for the equality of the different estimators are performed jointly for the two coefficients, we still reject the null hypothesis at the first step ($\chi_2^2 = 38.31$, $p = 0.00$) and we do not reject the null at the second step ($\chi_2^2 = 4.44$ $p = 0.1086$). Hence the results of the sequential test procedure are robust to either form of measurement errors. The most reliable estimator remains the IVL.

Also in this case, the common approach to test for correlated effects without conditioning to the existence or non existence of measurement errors is misleading. The hypothesis of equality of the estimators in the first two columns of Table 4 is rejected ($\chi_2^2 = 51.17$, $p = 0.00$) and the recommended estimator would be the WG. Conclusions based only on a visual comparison between Tables 3 and 4 may also be misleading. A comparison of Tables 3 and 4 shows that while the coefficients for the vacancies are almost untouched, there is a striking drop in the WG for the unemployed coefficient that cannot be compared to the slight decrease of all the other estimators. The coefficient also loses its significance. It seems that, having controlled for area-specific and nation-wide time specific factors, the effects of the stock of unemployed on the number of hirings are negligible. In other words one could infer that the unemployed data are almost completely explained by these factors. However, this interpretation needs some care. The IVD, robust to measurement errors, does not show such huge drop as the WG but its value is only slightly decreased, as are the estimators for the model in levels. In presence of strong measurement errors in variables, the estimates in the first two columns of Tables 3 and 4, namely *OLS* estimators, are not reliable. They neglect such unobservable factors and may be misleading. Once more it is worth noting that the DWH tests for unobservable heterogeneity in presence of measurement errors have not been applied for the comparison of *OLS* estimators but *IV* estimators have been used (third and fourth columns of Tables 3 and 4), as provided by the sequential procedure in Section 2.3.

In the search-matching framework analyzed the most reliable estimator seems to be the *IV* estimator on the models in levels (IVL) that control for measurement errors. This choice is robust to the introduction of time dummies. This chosen estimator presents also more reasonable results from a theoretical point of view. The hypothesis of constant returns to scale is not rejected (Model (2.10): $\chi_1^2 = 0.05$, $p = 0.8230$, Model (2.11): $\chi_1^2 = 0.23$, $p = 0.6315$).

2.6 Conclusions

The main implication from these findings is a *caveat* on the empirical use of estimation results in presence of strong unobservable factors in the data set. *OLS* estimators are almost never reliable but the availability of panel data sets and the use of estimators that control for unobservable heterogeneity bias, as widespread practice, does not always lead to the most reliable results. It is crucial to investigate what is the most important source of bias that affects the data set we are using. Different kind of unobservable variables may affect data at different levels of disaggregation. Panel data sets can be helpful in handling these issues. Pooling cross sectional and time series observations, the econometrics of panel data offers a variety of different estimators for the same parameter, and the behavior of such estimators in the presence of unobserved factors may be analyzed. Therefore, it is possible to acquire some knowledge about the kind of errors of specification involved, by checking whether they can actually account for the sign and order of magnitude of the observed discrepancies between estimators. Pursuing such an approach, we implement a sequential test aiming to distinguish the effects of unobserved heterogeneity and measurement errors on the estimators of the parameters in a panel data model. Size and power are investigated in a simulation experiment.

An application of the methodology to investigate widely discussed issues in labor economics is presented. Using a job search-matching framework, we study the effects of unobservable factors on the estimated elasticity of hirings to the stock of unemployed and to the stock of vacancies using a panel data set of TTWAs in the UK followed monthly from 1996 to 2001. A different dependent variable, i.e. unemployment outflows, has also been investigated but no more satisfactory results obtained. Our findings reveal that the data on unemployed and vacancies are affected by strong systematic measurement errors. In this particular case, unobservable cross-sectional differences, naturally associated with different labor market institutions across TTWAs, seem to be important in the determination of the number of vacancies but do not affect strongly the unemployed stock. However, it is the presence of measurement errors with an unknown structure that plays a major role. Our inference results are robust to the presence of correlated measurement errors (that follow a moving average process of order one or two, according to the data transformation used) with or without a period specific component. Indeed, we choose as instruments for a variable at time t its past value at lag $(t-3)$, and subsequently we correct with time dummies. Models controlling for unobserved heterogeneity bias may aggravate the measurement error bias. Therefore the most reliable estimators are instrumental variables on the model in levels. The hypothesis of constant returns to scale cannot be rejected. This investigation does not rule out the possibility that an empirical analysis of the matching function may lead to dissimilar results using a different data set. For instance, using data disaggregated by age or educational level it is likely that unobservable heterogeneity bias may be the most important issue to control for.

The analysis of the illustrated case of study reveals that conclusions lacking a rigorous statistical analysis of the effects of possible unobservable factors might be misleading. This implies that empirical results in contrast with the underlying economic theory do not always need a new theoretical explanation to be accommodated. They might be simply the results of an invalid inference. In presence of strong unobservable factors, as it is the case in analyzing the working of the labor market, the choice of the specification of the econometric model to be used is the most important and delicate phase. In our opinion it is often undervalued in empirical studies.

2.7 Appendix 1.2

The variables and matrices used in this Appendix, not directly reconsidered in Chapter 2, have been defined in Chapter 1, Appendix 4.1.

Part I

We need to compare an IV estimator and an OLS estimator on the model in deviations, i.e. $\widehat{\beta}_{ivd}$ and $\widehat{\beta}_{wg}$ respectively. In this context, an artificial regression of the type used in Appendix 4.1 does not help in constructing a test robust for the presence of non-spherical errors. In what follows, we explain why it is the case and we indicate an alternative procedure.

Consider the artificial regression of $Y^* = \begin{bmatrix} GY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} P_Z GX & P_Z GX \\ 0 & GX \end{bmatrix}$. By applying Lemma 13 and 14 in Appendix 4.1, it gives coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{ivd} - \widehat{\beta}_{wg} \\ \widehat{\beta}_{wg} \end{bmatrix}$. The disturbances have a covariance matrix $E\left(\varepsilon^* \varepsilon^{*\prime}\right) = \sigma^2 \begin{bmatrix} P_Z & P_Z \\ P_Z & I_{N(T-1)} \end{bmatrix}$, as $GG' = I_{N(T-1)}$.

In this case, the transformations of the data used in the two sets of equations are not orthogonal and $Var(Y^*)$ is not diagonal. We have

$$Var(Y^*) = \begin{bmatrix} GVar(Y)G' & GVar(Y)G' \\ GVar(Y)G' & GVar(Y)G' \end{bmatrix} = \begin{bmatrix} \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \\ \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \end{bmatrix}.$$

If now $\widetilde{X} = \begin{bmatrix} P_Z GX & 0 \\ 0 & GX \end{bmatrix}$,

$$\begin{aligned} Var(\widehat{\beta}^*) &= (X^{*\prime}X^*)^{-1}X^{*\prime}Var(Y^*)X^*(X^{*\prime}X^*)^{-1} \\ &= A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime}. \end{aligned}$$

Next, we calculate this variance by separating the different components.

$$\widetilde{X}'Var(Y^*)\widetilde{X} = \begin{bmatrix} X'G'P_Z' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \\ \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \end{bmatrix} \begin{bmatrix} P_Z GX & 0 \\ 0 & GX \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} X'G'P_Z' & X'G'P_Z' \\ X'G' & X'G' \end{bmatrix} \begin{bmatrix} P_Z GX & 0 \\ 0 & GX \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} X'G'P_Z'GX & X'G'P_Z'GX \\ X'G'P_Z'GX & X'QX \end{bmatrix}.$$

$$(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}.$$

Thus

$$(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}$$

$$= \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \times$$

$$\begin{bmatrix} X'G'P_Z'GX & X'G'P_Z'GX \\ X'G'P_Z'GX & X'QX \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} I & I \\ (X'QX)^{-1}[X'G'P_Z'GX] & I \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} & (X'QX)^{-1} \\ (X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}$$

and

$$A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime}$$

$$= \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & (X'QX)^{-1} \\ (X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \tag{2.12}$$

$$= \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} - (X'QX)^{-1} & 0 \\ (X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} - (X'QX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}. \tag{2.13}$$

If we run the artificial regression, the postulated variance-covariance matrix is different.

It will be proportional to

$$\begin{aligned}
(X^{*\prime}X^{*})^{-1} &= (A'\widetilde{X}'\widetilde{X}A)^{-1} = A^{-1}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'G'P_Z'GX)^{-1} & -(X'QX)^{-1} \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'G'P_Z'GX)^{-1} + (X'QX)^{-1} & -(X'QX)^{-1} \\ -(X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}.
\end{aligned}$$

The fact that the equation sets in the double length regression constructed are not orthogonal is not taken into consideration. A wrong answer will also come from the White's estimators. They are not robust to the presence of inter-group correlation. The use of a Newey-West robust *OLS* estimator would not help either. The variance covariance matrix exhibits a pattern of cross sectional dependence (i.e. particular form of non stationarity persistent when N goes to infinity) that is not supported by these estimators. Therefore, a consistent estimator for the variance of the difference of the two estimates (upper left part of matrix (2.13)) needs to be constructed step by step.

We need to recover the matrices involved and a consistent estimate of σ. Recall that for the first set we are performing an *IV* estimation by running *OLS* on a transformed model. It is known that the *OLS* residuals do not provide a consistent estimator of the variance of the initial disturbances, because the transformed model has a non spherical variance-covariance matrix. The sum of squares of the residuals coming from the initial model with the *IV* estimator should be used instead.

However, notice that

$$\widehat{\varepsilon}_{iv} = y - X\widehat{\beta}_{iv}$$

can be written as

$$\begin{aligned}
\widehat{\varepsilon}_{iv} &= y - X\widehat{\beta}_{ols} + X\widehat{\beta}_{ols} - X\widehat{\beta}_{iv} \\
&= \widehat{\varepsilon}_{ols} + X\left(\widehat{\beta}_{ols} - \widehat{\beta}_{iv}\right)
\end{aligned}$$

and therefore

$$\widehat{\varepsilon}'_{iv}\widehat{\varepsilon}_{iv} = \widehat{\varepsilon}'_{ols}\widehat{\varepsilon}_{ols} + \left(\widehat{\beta}_{ols} - \widehat{\beta}_{iv}\right)' X'X\left(\widehat{\beta}_{ols} - \widehat{\beta}_{iv}\right).$$

The sum of squares of the residuals coming from the initial model with the IV estimator is equal to the OLS sum of squares plus a function of the contrast between the two estimators, which is what we want to test eventually. This contaminates the variance estimate. Therefore, in order to get a consistent estimator of the variance we can rely only on the second set of equations in the auxiliary model. We run OLS on the second set of equations and use White robust standard errors. (They produce a consistent estimator of $X'Var(Y)X$ under the assumption that $Var(Y) = \Sigma$, a block diagonal matrix.) A possible assumption is that Σ, with N blocks $Var(y_i)$, each $T \times T$, on the main diagonal, has $Var(y_i) = \Omega_i + \sigma^2_{\eta,i}\iota\iota'$, although it becomes apparent that this is too general. Denote σ^2_1 the variance in the first set of equations ($\widehat{\sigma}^2_1$ the estimate) and σ^2_2 the variance in the second set of equations ($\widehat{\sigma}^2_2$ the estimate).

From the first set, we get

$$\begin{aligned}
X'Var(GY)X &= X'G'P'_Z\left[\sigma^2\Omega\right]P_Z GX \\
&= \sigma^2\left(X'G'P'_Z\Omega P_Z GX\right).
\end{aligned}$$

So

$$\begin{aligned}
&(X'X)^{-1}X'Var(Y)X(X'X)^{-1} \\
= \;& \sigma^2_1\left(X'G'P'_Z GX\right)^{-1}\left(X'G'P'_Z\Omega P_Z GX\right)\left(X'G'P'_Z GX\right)^{-1}.
\end{aligned}$$

In order to get the matrix of interest, we will divide the estimate of this variance by the obtained $\widehat{\sigma}_1^2$.

Denote $\Psi = \left(X'G'P_Z'GX\right)^{-1}\left(X'G'P_Z'\Omega P_ZGX\right)\left(X'G'P_Z'GX\right)^{-1}$.

Similarly, run *OLS* on the second set of equations and use White robust standard errors. We get

$$\begin{aligned} X'Var(GY)X &= X'G'\left[\sigma^2\Omega\right]GX \\ &= \sigma^2\left(X'G'\Omega GX\right). \end{aligned}$$

So

$$\begin{aligned} &(X'X)^{-1}X'Var(Y)X'(X'X)^{-1} \\ = \;& \sigma_2^2\left(X'QX\right)^{-1}\left(X'G'\Omega GX\right)\left(X'QX\right)^{-1}. \end{aligned}$$

Denote $\Theta = \left(X'QX\right)^{-1}\left(X'G'\Omega GX\right)\left(X'QX\right)^{-1}$.

A robust and consistent estimator of the precision matrix in the Wald test is[10]

$$\left[\widehat{Var}\left(\widehat{\beta}_{ivd} - \widehat{\beta}_{wg}\right)\right]^{-1} = \widehat{\sigma}_2^2\left(\Psi - \Theta\right).$$

Part II

We need to compare an instrumental variable estimator on the model transformed according to the *between groups* transformation and an instrumental variable estimator on the model transformed according to the *within groups* transformation.

The artificial regression of $Y^* = \begin{bmatrix} HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} P_ZHX & P_ZHX \\ 0 & P_ZGX \end{bmatrix}$ gives coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{ivl} - \widehat{\beta}_{ivd} \\ \widehat{\beta}_{ivd} \end{bmatrix}$. Results (2.6) and (2.7) in Lemma 15 directly follow from the application of Lemma 13 and 14 in Appendix 4.1. Moreover, as in the construction of

[10]Note that the precision matrix may not always be positive definite in finite samples.

the artificial regression in Appendix 4.1, we use again two orthogonal transformations.

Also in this case

$$Var(Y^*) = \begin{bmatrix} HVar(Y)H' & 0 \\ 0 & GVar(Y)G' \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{bmatrix}.$$

As $H = I_N \otimes H^+ = I_N \otimes \frac{1}{T} i'$ is $N \times NT$, the instrument set for variables transformed by H, Z_H say, is not the same as for variables transformed using G, which is $N(T-1) \times NT$, Z say. One may set $Z_H = HZ$, but this is not necessary. If now $\widetilde{X} = \begin{bmatrix} P_{Z_H} HX & 0 \\ 0 & P_Z GX \end{bmatrix}$,

$$\begin{aligned} Var(\widehat{\beta}^*) &= (X^{*\prime}X^*)^{-1}X^{*\prime}Var(Y^*)X^*(X^{*\prime}X^*)^{-1} \\ &= A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime}. \end{aligned}$$

Next, we calculate this variance by separating the different components, as we did in Part I of this Appendix.

$$\begin{aligned} \widetilde{X}'Var(Y^*)\widetilde{X} &= \begin{bmatrix} X'H'P'_{Z_H} & 0 \\ 0 & X'G'P'_Z \end{bmatrix} \begin{bmatrix} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{bmatrix} \begin{bmatrix} P_{Z_H}HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} X'H'P'_{Z_H} & 0 \\ 0 & X'G'P'_Z \end{bmatrix} \begin{bmatrix} (\theta + 1/T)P_{Z_H}HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} (\theta+1/T)X'H'P'_{Z_H}HX & 0 \\ 0 & X'G'P'_Z GX \end{bmatrix}. \end{aligned}$$

$$(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1} \end{bmatrix}.$$

Thus

$$\begin{aligned} &(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1} \\ &= \sigma^2 \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1} \end{bmatrix} \times \\ &\begin{bmatrix} (\theta+1/T)X'H'P'_{Z_H}HX & 0 \\ 0 & XG'P'_Z GX \end{bmatrix} \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1} \end{bmatrix} \end{aligned}$$

$$= \sigma^2 \begin{bmatrix} (\theta + 1/T)(X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_ZGX)^{-1} \end{bmatrix}$$

and

$$\begin{aligned}
& A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1'} \\
= \; & \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (\theta + 1/T)(X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_ZGX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (\theta + 1/T)(X'H'P'_{Z_H}HX)^{-1} & -(XG'P'_ZGX)^{-1} \\ 0 & (X'G'P'_ZGX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (\theta + 1/T)(X'H'P'_{Z_H}HX)^{-1} + (X'G'P'_ZGX)^{-1} & -(X'G'P'_ZGX)^{-1} \\ -(X'G'P'_ZGX)^{-1} & (X'G'P'_ZGX)^{-1} \end{bmatrix} \quad (2.14)
\end{aligned}$$

We now need to find the variance-covariance matrix the artificial regression will assume. This will be proportional to

$$\begin{aligned}
(X^{*\prime}X^*)^{-1} &= (A'\widetilde{X}'\widetilde{X}A)^{-1} = A^{-1}(\widetilde{X}'\widetilde{X})^{-1}A^{-1'} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (XG'P'_ZGX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & -(X'G'P'_ZGX)^{-1} \\ 0 & (X'G'P'_ZGX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} + (X'G'P'_ZGX)^{-1} & -(X'G'P'_ZGX)^{-1} \\ -(X'G'P'_ZGX)^{-1} & (X'G'P'_ZGX)^{-1} \end{bmatrix} \quad (2.15)
\end{aligned}$$

By comparing (2.14) with (2.15) it appears that an artificial regression is a valuable device to estimate a suitable variance-covariance matrix. We also need to consider the (White) robust *OLS* estimator which uses a consistent estimator of $X^{*\prime}Var(Y^*)X^*$ under the assumption that $Var(Y) = \Sigma$ is block diagonal.

$$\begin{aligned}
\widetilde{X}'Var(Y^*)\widetilde{X} &= \begin{bmatrix} X'H'P'_{Z_H} & 0 \\ 0 & X'G'P'_Z \end{bmatrix} \begin{bmatrix} H\Sigma H' & 0 \\ 0 & G\Sigma G' \end{bmatrix} \begin{bmatrix} P_{Z_H}HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\
&= \begin{bmatrix} X'H'P'_{Z_H}H\Sigma H' & 0 \\ 0 & X'G'P'_Z G\Sigma G' \end{bmatrix} \begin{bmatrix} P_{Z_H}HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\
&= \begin{bmatrix} X'H'P'_{Z_H}H\Sigma H'P_{Z_H}HX & 0 \\ 0 & X'G'P'_Z G\Sigma G'P_Z GX \end{bmatrix}.
\end{aligned}$$

Denote for simplicity $\Gamma = X'H'P'_{Z_H}H\Sigma H'P_{Z_H}HX$, $\Pi = X'G'P'_Z G\Sigma G'P_Z GX$. Thus

$$\begin{aligned}
&(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1} \\
&= \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1} \end{bmatrix} \times \\
&\quad \begin{bmatrix} \Gamma & 0 \\ 0 & \Pi \end{bmatrix} \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1}\Gamma & 0 \\ 0 & (X'G'P'_Z GX)^{-1}\,\Pi \end{bmatrix} \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (X'H'P'_{Z_H}HX)^{-1}\Gamma(X'H'P'_{Z_H}HX)^{-1} & 0 \\ 0 & (X'G'P'_Z GX)^{-1}\,\Pi\,(X'G'P'_Z GX)^{-1} \end{bmatrix}
\end{aligned}$$

Denote for simplicity
$U = (X'H'P'_{Z_H}HX)^{-1}\Gamma(X'H'P'_{Z_H}HX)^{-1}$, $V = (X'G'P'_Z GX)^{-1}\,\Pi\,(X'G'P'_Z GX)^{-1}$.

$$\begin{aligned}
&A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} U & -V \\ 0 & V \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} = \begin{bmatrix} U+V & -V \\ -V & V \end{bmatrix}.
\end{aligned}$$

In the case $Var(y_i) = \sigma^2 I_T + \sigma_\eta^2 \iota\iota'$, the residuals from this regression of $Y^* = \begin{bmatrix} Y \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} P_Z HX & P_Z HX \\ 0 & P_Z GX \end{bmatrix}$ to give coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{ivl} - \widehat{\beta}_{ivd} \\ \widehat{\beta}_{ivd} \end{bmatrix}$ can be obtained by stacking those from Y on $P_Z HX$ above those from GY on $P_Z GX$. Similarly to the artificial regression considered in Section 1.3 the first set of equations needs to be scaled by

$$k = \sqrt{T/(1+\theta T)}$$

as otherwise there is no multiple of the residual sum of squares of the artificial regression (i.e. $RSS_A + RSS_B$) with expectation σ^2. However, because in this case we are performing an IV estimation by running OLS on a transformed model, the OLS residuals do not provide a consistent estimator of the variance of the initial disturbances. Both in the estimation of θ and in the test statistic, the sum of squares of the residuals has to be calculated using the IV estimate of β and the untransformed right hand side variables. The Hausman test can be calculated by carrying out the artificial regression of $Y^* = \begin{bmatrix} \widehat{k}HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} \widehat{k}HX & \widehat{k}HX \\ 0 & GX \end{bmatrix}$ and constructing a Wald test, W, on the first K coefficients, using the following correction:

$$W_{iv} = W_{ols} \frac{\frac{(RSS_A + RSS_B)_{iv}}{[NT-2K]}}{\frac{(RSS_A + RSS_B)_{ols}}{[NT-2K]}} = W_{ols} \frac{(RSS_A + RSS_B)_{iv}}{(RSS_A + RSS_B)_{ols}},$$

where quantities with subscript iv are referred to the initial model and the ones with subscript ols are referred to the transformed model.

Returning to the case $Var(y_i) = \Omega_i + \sigma_{\eta,i}^2 \iota\iota'$, the device of scaling the first set of equations in the artificial regression requires U and V to be in simple scalar ratio. The case of simple heteroscedasticity over time, $Var(y_i) = \Omega + \sigma_\eta^2 \iota\iota', \Omega$ diagonal, gives $\Sigma = Var(Y) = I_N \otimes (\Omega + \sigma_\eta^2 \iota\iota')$ and $Var(HY) = I_N \otimes (\overline{\sigma}^2/T + T\sigma_\eta^2)$, where $\overline{\sigma}^2 = \iota'\Omega'\iota/T$. But $Var(GY) = I_N \otimes (G^{+\prime}\Omega G^+)$, and the presence of the different P terms removes any simple relationship. The orthogonality of HY and GY ensures that it is straightforward to combine the results from separate estimations of $\widehat{\beta}_{ivl}$ and $\widehat{\beta}_{ivd}$ to calculate W.

Chapter 3

Latent Variables in Dynamic Panel Data Models

This chapter implements optimal minimum distance estimators to estimate consistently autoregressive models for panel data with the joint occurrence of unobserved heterogeneity and systematic measurement errors-in-variables. Efficiency considerations are presented. The proposed estimators are applied in a selected case of study where also cross-sectional dependence needs to be taken into consideration. The resulting empirical model estimated is a spatio-temporal panel data model with unobserved heterogeneity and systematic measurement errors.

3.1 Introduction

Many models in social sciences suggest that current behavior of the agents depends upon past behavior (persistence, habit formation, partial adjustment, etc....). Panel data have the unique ability to allow us to model and compare the dynamics of different individuals. But how do we deal with measurement errors and the unobserved heterogeneity issue in a dynamic framework? Dynamic models complicate the estimation procedure because measurement errors require the use of instrumental variables techniques in order to obtain consistent estimators and in a dynamic context only predetermined instruments may be available. Furthermore, certain transformations that eliminate the unobserved heterogeneity, such as first differences or deviation from time-means, lead to inconsistent estimators when instruments are predetermined.

Consistent estimators for (stationary) autoregressive panel data models with white noise errors (assuming exact measurement of the variables used as regressors) are presented, among others, by Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998). Consistent estimators for static panel data models with measurement errors in the regressors are discussed, among others, by Biorn (2000). Both measurement errors and the presence of a lagged dependent variable make *OLS* estimators inconsistent. In both contexts, the estimation techniques proposed to overcome the problem is a minimum distance estimator. This chapter combines these two strands of the panel data literature and presents optimal minimum distance estimators to estimate consistently a model which superimposes the errors-in-variables problem and the heterogeneity problem on a dynamic framework. Furthermore, the measurement errors are not supposed to be random. A more complicated process is assumed.

The chapter is organized as follows. Section 3.2 reviews the application of the *Generalized Method of Moments* for models of covariance structures. Sections 3.3 and 3.4 describe the use of this methodology to estimate dynamic panel data models without measurement errors and static panel data models with measurement errors respectively. In Section 3.5 this estimation method is used to obtain consistent estimators of the autoregressive parameter of a (stationary) panel data dynamic model with latent variables. The different consistent estimators presented vary in terms of efficiency. The more efficient is the estimator, the more restrictive are the hypotheses underlying the construction of the model. Section 3.6 presents a concrete case of study where the analyzed econometric model specification and related estimation methodology have been applied and tailored. Section 3.7 concludes.

3.2 Covariance Structures and the GMM Criterion

The *Generalized Method of Moments* (GMM) was first proposed by Hansen (1982) and extended by White (1984). Their results assume $T \to \infty$. It is the equivalent of the *Generalized Instrumental Variables Methodology* (IV), which we used in Section 2.3, in a framework that may allow for dynamics and non-linearities. This section presents the application of the GMM to panel data as used, among others, by Arellano and Bond (1991) and Chamberlain (1982). Here, the results assume $N \to \infty$. In this framework the attention is focused on the second moments of the data. Multiple sample moments are combined into a single estimate of the population moments. Consider the i-th observation of the sample

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix},$$

and assume (for simplicity of exposition)

$$E(y_i = 0), Var\left(y_i\right) = E\left(y_i y_i^{'}\right) = \Omega_i\left(\vartheta\right),$$

where ϑ is a vector of parameters to be estimated. The statistical model can be expressed as a list of orthogonality conditions

$$E\left[\left(y_i y_i^{'}\right) - \Omega_i\left(\vartheta\right)\right] = 0 \qquad \forall i.$$

We now explain how to estimate ϑ using the GMM. In this context, this technique is often referred to as the *optimal minimum distance* estimator.

In order to recover the standard vector formulation of the GMM (Hansen, 1982), the matrices $y_i y_i^{'}$ and $\Omega_i(\vartheta)$ need to be transformed in vector form. In the simple case of a symmetric covariance matrix it is sufficient to put all distinct elements of the upper triangle or lower triangle matrix one after the other in a vector. For example, consider model (1.1) and suppose $T = 3$. We have

$$
\begin{aligned}
y_i &= \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix}, \qquad y_i y_i^{'} = \begin{pmatrix} y_{i1}^2 & y_{i1}^{'} y_{i2} & y_{i1}^{'} y_{i3} \\ \vdots & y_{i2}^2 & y_{i2}^{'} y_{i3} \\ \dots & \dots & y_{i3}^2 \end{pmatrix}, \\
\Omega(\vartheta) &= \begin{pmatrix} \sigma_\eta^2 + \sigma^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \vdots & \sigma_\eta^2 + \sigma^2 & \sigma_\eta^2 \\ \dots & \dots & \sigma_\eta^2 + \sigma^2 \end{pmatrix}.
\end{aligned}
$$

Denote

$$
\underset{6\times 1}{S} = \begin{pmatrix} y_{i1}^2 \\ y_{i1}^{'} y_{i2} \\ \vdots \\ y_{i3}^2 \end{pmatrix}, \qquad \underset{6\times 1}{W(\vartheta)} = \begin{pmatrix} \sigma_\eta^2 + \sigma^2 \\ \sigma_\eta^2 \\ \vdots \\ \sigma_\eta^2 + \sigma^2 \end{pmatrix}.
$$

The set of orthogonality conditions for each i can be written as

$$
E\left[S - W(\vartheta)\right] = 0.
$$

The estimation is based on the sample orthogonal conditions

$$
\frac{1}{N} \sum_i \left[s_i - w_i(\vartheta)\right] = 0,
$$

where s_i is the generic element of S and $w_i(\vartheta)$ is the corresponding element in $W(\vartheta)$.

Defining $\overline{s} = \frac{1}{N}\sum_i s_i$ and $w(\vartheta) = \frac{1}{N}\sum_i w_i(\vartheta)$, the sample orthogonal conditions can be written as

$$\overline{s} - w(\vartheta) = 0.$$

We can then estimate ϑ by GMM.

$$\widehat{\vartheta}_{GMM} = \arg\min_{\vartheta} \left[\overline{s} - w(\vartheta)\right]' \widehat{V}^{-1} \left[\overline{s} - w(\vartheta)\right]$$

where

$$\widehat{V} = \frac{1}{N}\sum_i \left[s_i s_i' - \overline{s}\overline{s}'\right]$$

is the sample covariance matrix of the orthogonal conditions which yields the most efficient GMM estimator.[1]

If the restrictions are linear in $\Omega(\vartheta)$, it is possible to derive analytically the estimator. Otherwise it has to be calculated by numerical optimization.

For instance, if we rewrite

$$\underset{6\times 1}{W(\vartheta)} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_\eta^2 \\ \sigma^2 \end{pmatrix} = H\vartheta,$$

using the *GMM criterion*

$$\min_{\vartheta} \left[\overline{s} - H\vartheta\right]' \widehat{V}^{-1} \left[\overline{s} - H\vartheta\right]$$

[1] Note that we are minimizing a quadratic form, but we could also minimize other measures of distance. As a result, we can obtain other estimators. However, they cannot perform better in asymptotic terms, as the GMM estimator reaches the efficient lower bound, provided the optimal weighting matrix is used and the set of orthogonality conditions remains unchanged.

we obtain

$$\widehat{\vartheta} = \left(H'\widehat{V}^{-1}H\right)^{-1} H'\widehat{V}^{-1}\overline{s}.$$

This estimator has the form of a *GLS* and it can be calculated all in one step. However in most empirical cases we do not have linear restrictions. Moving Average models (MA models), for instance, imply linear restrictions in the covariance matrix and they are easy to treat but they may not be appropriate in most real cases. It is sufficient to note that in an MA model all the variables on the right-hand-side are unobservables. For this reason the analysis of dynamic models for panel data will be restricted in this work to Autoregressive Models (AR models). In a AR model, the restrictions are not linear and the estimation is slightly more complicated but such models have a wider applicability. The next section describes how to use the GMM technique in order to obtain consistent estimators of a first order autoregressive panel data model with unobserved heterogeneity, assuming exact measurement.

3.3 Dynamic Panel Data Models with Unobserved Heterogeneity and without Measurement Errors

Arellano and Bond (1991) propose consistent estimators for dynamic panel data models using a method of moments formulation. Consider a random sample of individual time series of length T, $\{y_i, i = 1.....N\}$, with second-order moments matrix $E\left(y_i y_i^{'}\right) = \Omega_i$.

Assume that the joint distribution of y_i and the individual effect η_i, satisfies

$$y_{it} = \alpha y_{i,t-1} + \eta_i + \varepsilon_{it}, \quad |\alpha| < 1, \qquad i = 1, ..., N, \quad t = 1, ..., T, \tag{3.1}$$

where

$$E(\varepsilon_{it} \mid y_{i,t-1}) = 0, \qquad t = 1, ..., T, \tag{3.2}$$

$$E(\eta_i) = \mu,\; E(\varepsilon_{it}^2) = \sigma_t^2,\; Var(\eta_i) = \sigma_\eta^2. \tag{3.3}$$

It is worthwhile noting that this is a quite general autoregressive specification of the model. Assumption (3.2) does not rule out correlation between η_i and ε_{it}, nor the possibility of conditional heteroskedasticity, since $E(\varepsilon_{it}^2 \mid y_{i,t-2})$ need not coincide with σ_t^2. It assumes lack of serial correlation in the errors but not independence over time. However, it is the crucial assumption because it allows us to consider values of y lagged two periods or more as valid instruments in the equations in first differences. Arellano and Bond (1991) propose also a test for serial correlation of the residuals.

Consider the restrictions we need in order to estimate Ω_i. Ω_i is a $T \times T$ matrix and therefore has $\frac{1}{2}\left[T \times (T+1)\right]$ distinct elements. Model (3.1)-(3.3) has $(T+3)$ parameters (T variances, α, μ, σ_η). The number of overidentifying restrictions is

$$\frac{1}{2}\left[T \times (T+1)\right] - (T+3) = \frac{1}{2}\left[T^2 - T - 6\right].$$

Model (3.1)-(3.3) implies $m = \frac{(T-2)(T-1)}{2}$ linear moment restrictions

$$E\left[y_{it-j}\left(\Delta y_{it} - \alpha \Delta y_{i,t-1}\right)\right] = 0, \qquad j = 2, ..., (t-1),\ t = 3, ..., T, \tag{3.4}$$

where Δ is the operator that transforms the data in first differences.

Arellano and Bond (1991) propose to use a GMM approach to estimate α when $N \to \infty$ and T is small. We can express the conditions (3.4) as

$$E\left(Z_i' \Delta\varepsilon_i\right) = 0$$

$$\text{where} \quad \Delta\epsilon_i = \begin{pmatrix} \Delta\varepsilon_{i2} \\ \vdots \\ \Delta\varepsilon_{iT} \end{pmatrix} = \begin{pmatrix} \Delta y_{i2} - \alpha\Delta y_{i1} \\ \vdots \\ \Delta y_{iT} - \alpha \Delta y_{i,T-1} \end{pmatrix} = \left\{\Delta y_{it} - \alpha \Delta y_{i,t-1}\right\}, \quad t = 2, ..., T.$$

The matrix of instruments Z_i is a block diagonal matrix of the form

$$\underset{(T-2)\times m}{Z_i} = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & 0 & \cdots & 0 \\ . & . & . & & . & . & & . \\ . & . & . & & . & . & & . \\ 0 & 0 & 0 & \cdots & y_{i1} & y_{i2} & \cdots & y_{i,T-2} \end{bmatrix}.$$

Following the technique presented in Section 3.2, consider the sample moments

$$b(\alpha) = \frac{1}{N} \sum_i \left[Z_i' \Delta\varepsilon_i \right].$$

The value of α which minimizes the quadratic form $b(\alpha)' A_N \ b(\alpha)$ is the GMM estimator

$$\widehat{\alpha}_{GMM} = \arg\min_{\alpha} b(\alpha)' A_N b(\alpha),$$

where A_N is a weighting matrix. The estimators depend on the choice of A_N. As explained in Hansen (1982), all of them are consistent because they are built on the sample moment conditions but not all of them are efficient. As indicated in Section 3.2, an optimal choice of A_N is provided by the inverse of the covariance of the orthogonality conditions

$$E\left(Z_i' \Delta\varepsilon_i \Delta\varepsilon_i' Z_i \right). \tag{3.5}$$

The moments that have more variance receive less weight. In this way the precision of the estimators is higher. In practice, in order to obtain a feasible estimator we have to follow a two-step procedure. First, use a sub-optimal weighting matrix that does not depend on ϑ (e.g. the identity matrix). For the autoregressive model considered above, the one step consistent (but not efficient) estimator, $\widehat{\alpha}_1$, can be obtained by setting

$$A_N = \left[\frac{1}{N} \sum_i \left(Z_i' H Z_i \right) \right]^{-1}$$

where H is a $(T-2)$ square matrix which has twos in the main diagonal, minus ones in the first subdiagonals and zeros otherwise. The residuals, $\widehat{\varepsilon}_i$, from this preliminary estimation can then be used to obtain a consistent estimator of the optimal weighting matrix (formula (3.5)). Taking the sample counterpart

$$\widehat{A}_N = \left[\frac{1}{N}\sum_i \left(Z_i^{\prime}\Delta\widehat{\varepsilon}_i\Delta\widehat{\varepsilon}_i^{\prime}Z_i\right)\right]^{-1},$$

we obtain a two step estimator $\widehat{\alpha}_2 = \widehat{\alpha}_{GMM}$, that is efficient. Note that $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ are asymptotically equivalent if the ε_{it} are independent and homoskedastic both across units and over time. However, such a procedure does not use information which is contained in the levels of the variables. It loses what sometimes is a very substantial part of the total variation in the data. Arellano and Bover (1995) propose an estimation method that uses also equations in levels. They are concerned with panel data models that specify instruments in levels for equations in first differences and instruments in first differences for equations in levels. Arellano and Bover (1995) consider models in which it is usually assumed that all the explanatory variables are potentially correlated with the individual effects. Therefore, only estimators based on transformations of the original observations that purge the model from the $\eta_i s$ can be consistent. However, if there are instruments available that are not correlated with the individual effects, one can use also the information contained in the levels of the variables which, if exploited, could improve the efficiency of the resulting estimators. The instruments of this kind that they choose are first differences of variables that have a constant correlation with the individual effects. Arellano and Bover (1995) carry out simulations of a first order autoregressive model with individual effects. Such an experiment illustrates the potential of exploiting moment restrictions in level equations using predetermined variables in first differences. The validity of the added moment conditions need, of course, to be tested.[2] Next, we will review this combined GMM methodology.

[2]The Sargan test of the over-identifying restrictions (Sargan, 1958; Hansen, 1982) is the statistic typically used to assess the validity of the enlarged matrix of instruments.

Consider the standard autoregressive model (3.1)-(3.3). Following Arellano and Bond (1991), Arellano and Bover (1995) define the orthogonality conditions as

$$E\left(Z_i^{+}H\varepsilon_i\right) = 0,$$

where

$$H = \begin{pmatrix} D \\ I \end{pmatrix} \text{ is the transformation matrix}$$

and

$$Z_i^{+} = \begin{pmatrix} Z_i' & 0 \\ 0 & Z_{li}' \end{pmatrix} \text{ is the matrix of instruments.}$$

D is the matrix that transform the variables in differences, I is the identity matrix, Z_i is the block diagonal matrix which contains the instruments available for the equations transformed by H and Z_{li} is also a block diagonal matrix which contains the instruments available for the equations in levels. They estimate the following system

$$\begin{cases} E\left(Z_i' D\varepsilon_i\right) = 0, & \text{orthogonality conditions for the variables in first differences} \\ E\left(Z_{li}'\varepsilon_i\right) = 0, & \text{orthogonality conditions for the variables in levels.} \end{cases} \tag{3.6}$$

Note that, given the moment conditions for the equations in first differences, some of the added restrictions for the equations in levels will be redundant. For instance, consider the standard autoregressive model (3.1)-(3.3). We can rely on $\frac{(T-2)\times(T-1)}{2}$ "basic" moments, i.e. conditions (3.4), but if for instance we also assume mean-stationarity (Blundell and Bond, 1998) we have a long list of other "additional" moments coming from

$$E\left[y_{it} \mid \eta_i\right] = \eta_i, \qquad \forall t.$$

However, the only useful ones are

$$E\left[\Delta y_{it-j}\left(y_{it-j+1} - \alpha y_{it-j}\right)\right] = 0 \qquad \forall j = 1, 2, \ldots\ldots$$

All the other moments conditions implied by mean stationarity are redundant. For example, $E\left[\Delta y_{i,t-2}\left(y_{it}-\alpha y_{i,t-1}\right)\right]=0$ can be expressed as a linear combination of the difference of two "basic" moments and an "additional" one where the lag between Δy_{it} and ε_{it} is just one:

$$\begin{aligned} & E\left[\Delta y_{i,t-2}\left(y_{it}-\alpha y_{i,t-1}\right)\right]= \\ = \; & E\left[\Delta y_{i,t-2}\left(\Delta y_{it}-\alpha \Delta y_{i,t-1}\right)\right]+E\left[\Delta y_{i,t-2}\left(y_{i,t-1}-\alpha y_{i,t-2}\right)\right]=0. \end{aligned}$$

The instruments for the added conditions are only

$$\underset{(T-2)\times m}{Z_{li}} = \begin{bmatrix} \Delta y_{i2} & 0 & 0 & . & 0 & 0 \\ 0 & \Delta y_{i3} & 0 & . & . & \vdots \\ . & . & . & . & . & . \\ 0 & \cdots & . & . & 0 & \Delta y_{i,T-1} \end{bmatrix}.$$

where $m=\frac{2(T-2)}{2}=(T-2)$.

The procedure proposed by Arellano and Bover (1995), also called *System GMM,* combines all the non redundant information. The optimal estimator is

$$\widehat{\alpha}_{GMM} = \arg\min_{\alpha} b^{+}(\alpha)' \widehat{A}_N^{+} b^{+}(\alpha),$$

where

$$b^{+}(\alpha)=\frac{1}{N}\sum_i\left[Z_i^{+}H\varepsilon_i\right].$$

and

$$\widehat{A}_N=\left[\sum_i\left(Z_i^{+\prime}H\widehat{\varepsilon}_i\widehat{\varepsilon}_i^{\prime}H^{\prime}Z_i^{+\prime}\right)\right]^{-1}.$$

The next section describes how to use the GMM in order to obtain consistent estimators in presence of measurement errors and unobserved heterogeneity in static panel data models. The section begins by showing that panel data may create a context where consistent estimation may be possible even if the variables are measured with errors.[3]

3.4 Static Panel Data Models with Unobserved Heterogeneity and with Measurement Errors

In a panel data set each individual and each period is replicated. This characteristic is often called "the repeated measurement property" and it is the main reason why panel data sets can make the errors-in-variables identification problem more manageable. The measurement error problem can be reduced by taking averages which, in turn, may show sufficient variation to permit consistent estimation. Following this idea Biorn (2000) constructs estimators from period means and discusses their consistency. Next, we will review this approach in a simplified framework.

Consider a uni-variate panel regression model with unobserved heterogeneity and "classical" errors-in-variables

$$y_{it} = x_{it}^{*}\beta + \eta_i + v_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T \tag{3.7}$$

where the observed variable is $x_{it} = x_{it}^{*} + \varepsilon_{it}$ and $x_{it}^{*} \sim iid\left(0, \sigma_{*}^{2}\right)$, $\varepsilon_{it} \sim iid\left(0, \sigma_{\varepsilon}^{2}\right)$, $v_{it} \sim iid\left(0, \sigma_{v}^{2}\right)$, $\eta_i \sim iid\left(0, \sigma_{\eta}^{2}\right)$, $\forall i, t$, and ε_{it}, v_{js} and η_r are independent $\forall i, t, j, s, r$.

The observed equation is

$$y_{it} = x_{it}\beta + \eta_i + \varpi_{it}$$

where

$$\varpi_{it} = v_{it} - \varepsilon_{it}\beta.$$

Taking differences in period means, we obtain

$$\Delta_s \overline{y}_{.t} = \Delta_s \overline{x}_{.t}\beta + \Delta_s \overline{\varpi}_{.t} \tag{3.8}$$

[3]Note that we mantain the assumption of random measurement errors. In other words, we still consider errors-in-variables as "classical".

where

$$\overline{y}_{.t} = \frac{\sum\limits_i y_{it}}{N}, \quad \overline{x}_{.t} = \frac{\sum\limits_i x_{it}}{N}, \quad \overline{\varpi}_{.t} = \frac{\sum\limits_i \varpi_{it}}{N}$$

and Δ_s is the operator differencing over s periods.

Taking differences between period means, we obtain

$$(\overline{y}_{.t} - \overline{y}) = (\overline{x}_{.t} - \overline{x})\,\beta + (\overline{\varpi}_{.t} - \overline{\varpi}) \tag{3.9}$$

where

$$\overline{y} = \frac{\sum\limits_i\sum\limits_t y_{it}}{NT}, \quad \overline{x} = \frac{\sum\limits_i\sum\limits_t x_{it}}{NT}, \quad \overline{\varpi} = \frac{\sum\limits_i\sum\limits_t \varpi_{it}}{NT}.$$

The law of large numbers, under weak conditions, implies that

$$\underset{N\to\infty}{plim}\ (\overline{\varpi}_{.t}) = 0. \tag{3.10}$$

Hence consistent estimators of β can be obtained by OLS on model (3.8) or model (3.9), respectively

$$\widehat{\beta}_{\Delta_s} = \left(\sum_{t=s+1}^{T} (\Delta_s \overline{x}_{.t})' (\Delta_s \overline{x}_{.t})\right)^{-1} \sum_{t=s+1}^{T} (\Delta_s \overline{x}_{.t})' (\Delta_s \overline{y}_{.t}), \qquad s = 1, ..., T-1$$

and

$$\widehat{\beta}_{bp} = \left(\sum_{t=s+1}^{T} (\overline{x}_{.t} - \overline{x})' (\overline{x}_{.t} - \overline{x})\right)^{-1} \sum_{t=s+1}^{T} (\overline{x}_{.t} - \overline{x})' (\overline{y}_{.t} - \overline{y}).$$

These estimators simply exploit the fact that averages of a large number of repeated measurements of an error-ridden variable can give (under weak conditions) a consistent measure of the true average at the limit, provided that this true average shows variation along the remaining dimension and that the measurement error has no period specific component. However, the last assumption is often not sustainable in many empirical situations. In these cases the estimators proposed are not consistent any longer. We relax this assumption and discuss the estimation of the resulting model in Section 3.5.

Consistent estimators of β can also be constructed using a *minimum distance estimator* , that is a *GMM estimator*. The identification of β comes from the second order moments of the observable variables, $Cov(x_{it}, x_{is})$, $Cov(x_{it}, y_{is})$, $Cov(y_{it}, y_{is})$, $\forall i, t, s$ and in general depends on whether or not the imposed structure is sufficient to obtain a unique solution for β. For model (3.7) the second order moments are

$$\begin{cases} Cov(x_{it}, x_{it}) = \sigma_*^2 + \sigma_\varepsilon^2 \\ Cov(x_{it}, y_{it}) = \sigma_*^2\beta & i = 1...N; \qquad t = 1...T \\ Cov(y_{it}, y_{it}) = \beta^2\sigma_*^2 + \sigma_v + \sigma_\eta \end{cases} \tag{3.11}$$

and

$$\begin{cases} Cov(x_{it}, x_{is}) = 0 \\ Cov(x_{it}, y_{is}) = 0 \quad i = 1...N; \qquad t, s = 1...T \quad t \neq s \\ Cov(y_{it}, y_{is}) = \sigma_\eta \end{cases} \tag{3.12}$$

We have clearly lack of identification. We have a system with more unknown than equations. Conditions (3.11) and (3.12) are not sufficient to identify $\sigma_*^2, \sigma_\varepsilon^2, \beta, \sigma_v, \sigma_\eta$. Biorn (2000) considers five groups of assumptions: basic orthogonality assumptions, additional assumptions on the measurement errors, additional assumptions on the disturbances, additional assumptions on the latent variable. Biorn (2000) analyses various combinations of these groups and for each model so defined, he constructs valid and non redundant orthogonality conditions. He deals with different covariance structures and presents several IV and GMM estimators. The estimation procedures are of two kinds.

1. The equations are transformed to differences to remove individual heterogeneity and are estimated by IV or GMM. Level values of the variables in other periods are used as instrumental variables.

2. The equations are kept in level form and are estimated by IV or GMM. Differenced values of the variables in other periods are used as instrumental variables.

The idea is exactly the same underlying the *System GMM* estimation proposed by

Arellano and Bover (1995). The difference is that the *System GMM* technique combines approaches 1 and 2 in order to increase the efficiency of the estimators. Biorn (2000) discusses the efficiency and robustness of the estimators constructed using approach 1 or 2 but he does not consider estimators combining the two approach. In the next section, estimators obtained mixing the two approaches are also implemented.

3.5 Dynamic Panel Data Models with Unobserved Heterogeneity and Error-Components Structured Measurement Errors

This section presents consistent estimators for a first order autoregressive panel data model with unobserved heterogeneity (as model (3.1)-(3.3)) and measurement errors. In order to study a wider case of economic situations, the measurement errors are no longer assumed to be "classical". Specifically, we consider a model that allows for the presence of a time varying component in the process of the measurement errors. This extension requires us to adjust the estimation methods proposed in the traditional related literature (reviewed in Sections 3.3 and 3.4). Indeed, the estimation of model (3.1)-(3.3) is based on the assumption (3.2), i.e. the errors are idiosyncratic shocks that are assumed to have cross-sectional zero mean at each point in time. However this assumption can be inadequate in a number of cases. One circumstance, rather recurrent in real data and not analyzed properly from a theoretical point of view, is the presence of time varying measurement error common to all individuals.[4]

[4]Many concrete cases may fit into this framework. For example consider a model which requires a measure of the permanent income and a sample with no income measures at all but with data on the estimated market value of the family residence. This housing value can be used as a proxy of the underlying permanent income concept but the discrepancies will not be random. Indeed they may be affected by house prices, time varying but common to all families, by family size at purchase time varying among families and constant over time and by unmeasured random locational factors (Griliches, 1984). Or we can think about the estimation of money demand of firms. We can use sales as proxy for cash holdings but clearly sales are affected by the efficiency of the production process of the single firm invariant over time (at least in the short run) and varying among firms, by the state of the national

Consider model (3.1)-(3.3) and suppose that the true variables are measured with errors that present a composite structure

$$
\begin{aligned}
y^*_{it} &= \alpha y^*_{i,t-1} + \eta_i + \varepsilon_{it}, \quad |\alpha| < 1, \qquad i = 1, ..., N; \; T = 2, ..., T, && (3.13)\\
y^*_{it} &= y_{it} + m_{it}, && (3.14)\\
m_{it} &= \phi_t + \mu_i + \xi_{it} && (3.15)
\end{aligned}
$$

where y_{it} are the observed values of y^*_{it}. The process of the measurement error consists of three independent components. The first, ϕ_t, is an individual-invariant time-specific effect with mean 0 and variance σ^2_ϕ uncorrelated over time, the second, μ_i, is a time-invariant individual-specific effect with mean 0 and variance σ^2_μ and the third, ξ_{it}, is a white noise component with mean 0 and variance σ^2_ξ.

This error component structure is much more realistic if we consider measurement errors not only as observation errors in the narrow sense but also as discrepancies between theoretical variable definitions and their observable counterparts in a wider sense.

Substituting we obtain

$$
y_{it} = \alpha y_{i,t-1} + \delta_t + d_i + e_{it}, \tag{3.16}
$$

where

$$
\begin{aligned}
\delta_t &= \alpha\phi_{t-1} - \phi_t,\\
d_i &= \alpha\mu_i - \mu_i + \eta_i,\\
e_{it} &= \alpha\xi_{i,t-1} - \xi_{it} + \varepsilon_{it}.
\end{aligned}
$$

Our aim is to control for the effects of the measurement errors on our observations. For this purpose one can treat the δ_t as unknown period specific parameters. We are only

---

economy varying over time but common to all firms and also by shocks outside the control of the management.

interested in a consistent estimators of α and we correct with time dummies. If we do not and define v_{it}, the new composite disturbance component, as

$$v_{it} = \delta_t + e_{it},$$

assumption (3.2) does not hold any longer

$$E(v_{it} \mid y_{i,t-1}) = \delta_t, \qquad t = 2, ..., T. \tag{3.17}$$

Therefore, the estimators proposed by Arellano and Bond (1991) cannot be used. Furthermore in this model also the estimators constructed by Biorn (2000) for panel data with measurement errors based on period means are not valid. If the measurement errors have a period specific component the probability limit of period means taken when the number of individuals goes to infinity (equation (3.10)) would no longer be zero. Thus, the estimators would not be consistent. Note also that model (3.13)-(3.15) yields a constant autocorrelation of measurement errors independent of the lag,

$$Cov(m_{it}, m_{i,t-j}) = \sigma_\mu^2,$$

which violates one of the basic assumptions of the classical errors-in-variables model, i.e. the measurement error are uncorrelated with everything else in the model included its own past values. However, following the GMM approach consistent estimators can be derived using appropriate instruments. The assumptions of model (3.13)-(3.15) induce MA(1) disturbances in the model involving observed variables (formulation (3.16)):

$$E\left(e_{it}e_{i,t-s}\right) = E\left[\left(\alpha\xi_{i,t-1} - \xi_{it} + \varepsilon_{it}\right)\left(\alpha\xi_{i,t-s-1} - \xi_{i,t-s} + \varepsilon_{i,t-s}\right)\right] = \begin{cases} -\alpha\sigma_\xi^2 & \text{if} \quad s = 1, \\ 0 & \text{if} \quad s \succeq 2. \end{cases}$$

Therefore, once model (3.16) has been transformed in first differences[5] in order to get rid

[5]We consider first differences for simplicity of exposition. However, in order to get an invertible

of the individual effects invariant over time

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \delta_t + \Delta e_{it},$$

valid instruments are only obtained by using dependent variables that are at least three-times lagged. Consequently, the set of all appropriate moment conditions can be written as:

$$\begin{cases} E\left[y_{i,t-j}\left(\Delta y_{it} - \alpha \Delta y_{i,t-1}\right)\right] = 0 \\ E\left[\Delta e_{it}\right] = 0 \end{cases} \quad j = 3, ..., t-1; \; t = 2, ..., T, \tag{3.18}$$

where the second set of moments, always valid, might be used when a limited number of points in time is available. In fact, the minimum number of time observations needed to get consistent estimators for a model in presence of measurement errors is greater that the one required when exact measurement is assumed. For instance, for a model like (3.13)-(3.15) we need to have at least four time observations. With $T = 4$, α and $\Delta\delta_4$ are just identified from the two moment conditions

$$\begin{cases} E\left[y_{i1}\left(\Delta y_{i4} - \alpha \Delta y_{i3} - \Delta\delta_4\right)\right] = 0 \\ E\left[\Delta y_{i4} - \alpha \Delta y_{i3} - \Delta\delta_4\right] = 0 \end{cases}$$

If $T > 4$, and thus we have overidentifying restrictions, we can use the *GMM criterion,* as detailed in Section 3.2. We calculate the sample equivalent of the moment conditions by constructing the $(T-3) \times 1$ vector $(y_{i,t-3})$, $t = 4, .., T$ and defining

$$b(\alpha, \delta) = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} 1 \\ y_{i,t-3} \end{bmatrix} (\Delta e_{it}) \qquad t = 4, ..., T,$$

covariance matrix, the transformation used in practice is forward deviations from time means (Arellano and Bover, 1995).

we minimize the quadratic form

$$b(\alpha, \delta)' A_N b(\alpha, \delta).$$

As is explained in Section 3.2, an optimal choice of the weighting matrix A_N is provided by the inverse of the variance of the orthogonality conditions which can be consistently estimated using the sample variance. The resulting GMM estimator is consistent and has the smallest asymptotic covariance matrix for a GMM estimator based on the conditions $b(\alpha, \delta)$.

However, this does not preclude the possibility of finding a more efficient GMM estimator based on an enriched set of moment conditions. In the remaining of this section, we implement a different (more efficient) GMM estimator for model (3.13)-(3.15), following the approach of Arellano and Bover (1995) and Blundell and Bond (1998). We assume there are available instruments that are not correlated with the individual and time effects and we use this information to construct additional orthogonality conditions for the model in levels. Typically, as explained in Section 3.3, using an autoregressive model without measurement errors in variables (model (3.1)-(3.3)), first differenced values of the variables dated $(t-1)$ are the candidate instruments. However, if measurement errors structured as in (3.14)-(3.15) are incorporated into the model, the serial correlation in the error term induces an endogeneity problem which makes one-time lagged first differences of the variables invalid instruments. Therefore, a *System GMM* estimator for model (3.13)-(3.15) should not use a standard matrix of instruments but it has to take as instruments for the equations in levels first differenced values of the variables dated $(t-2)$.

The matrices of instruments in the system (3.6) take the forms

$$\underset{(T-2)\times m}{Z_i} = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & 0 & \cdots & 0 \\ . & . & . & & . & . & & . \\ . & . & . & & . & . & & . \\ 0 & 0 & 0 & \cdots & y_{i1} & y_{i2} & \cdots & y_{i,T-3} \end{bmatrix}$$

where $m = \frac{(T-3)\times(T-2)}{2}$, and

$$\underset{(T-3)\times m}{Z_{li}} = \begin{bmatrix} \Delta y_{i2} & 0 & 0 & . & 0 & 0 \\ 0 & \Delta y_{i3} & 0 & . & . & \vdots \\ . & . & . & . & . & . \\ 0 & \cdots & . & . & 0 & \Delta y_{i,T-2} \end{bmatrix}$$

where $m = \frac{2(T-3)}{2} = (T-3)$.

3.6 An Empirical Application

In the remainder of this chapter, we present an economic model that is tested using the statistical model analyzed in Section 3.5. The theoretical framework (Patacchini and Zenou, 2003) is a job search-matching model which attempts to shed some light on the relationships between the residential location of workers and their labor market outcomes. There seems to be a growing awareness that some trends of economic variables might be due to spatial rather than purely economic factors. This is particularly true in the labor market (see, for example, Topa, 2001 and Manning, 2003) and especially for job search activities since, in a search-matching framework, a spatial dispersion of agents creates more frictions and thus more unemployment. In his seminal contribution to search, Stigler (1961) puts geographical dispersion as one of the four immediate determinants of price ignorance. The reason is simply that distance affects various costs associated with search.

We investigate, both theoretically and empirically, the relationship between job search and space by focusing on the impact of local cost of living and local labor market tightness on search intensity.

From a theoretical point of view, few models have introduced a spatial analysis in a search-matching model. Exceptions include Seater (1979), McCormick and Sheppard (1992), Simpson (1992), Rouwendal (1998), Ortega (2000), Coulson, Laing and Wang (2001), Sato (2001), Wasmer and Zenou (2002), Smith and Zenou (2003). Contrary to these models, our focus is on search intensity and its relationship with cost of living and labor market tightness in a local labor market.

From an empirical point of view, few papers have tested spatial search models. Most of the related empirical literature (which is in fact quite small) focuses on the aggregation of the matching function across space and on the interaction between local matching and regional migration or commuting behavior (see in particular the survey by Petrongolo and Pissarides, 2001, and also Jackman and Savouri, 1992, Burda and Profit, 1996, Burgess and Profit, 2001). In this chapter we analyze a different issue, namely the relationship between the county average job-search intensity, on the one hand, and the county cost of living and/or the county labor market tightness, on the other.

To be more precise, we first develop a simple model in which optimal search intensity is a result of a trade off between short run losses due to higher cost of search effort (more interviews, commuting...) and long-run gains due to higher chance to find a job. We show that this optimal search intensity is higher in areas characterized by larger cost of living and/or higher labor market tightness.

We then test this model using county-level data in England for the period 1991-2000. The level of spatial disaggregation of the cross sectional units is finer than the standard regional one. An analysis at the regional level would not be accurate enough to test the theoretical model. However, the availability and the quality of data collected at a sub-regional level are still very poor and a suitable econometric method is needed to get results robust to problems of data quality. Furthermore, a fundamental assumption in model (3.1)-(3.3), maintained also in model (3.13)-(3.15), is that the cross-sectional units are independent. This implies that residuals from different cross-sectional regressions are independent of each other. However, this assumption is not tenable when the cross-sectional units have a specific spatial connotation, e.g. regions or counties, as in our case of study. If this condition is not met, standard errors estimators are inconsistent and thus not useful for inference. This issue is taken into account in the formulation of the statistical model used to test the implications of the theoretical model in the UK context. As a result, spatial econometric techniques (see Anselin, 1988 for a review) are combined with the estimation methodology presented in Section 3.5. The empirical model formulated is a spatio-temporal panel data model with unobserved heterogeneity and systematic measurement errors in variables.

As predicted by the theoretical model, both the county cost of living and the county labor market tightness are found to have a positive and significant effect on the county search intensity. We also find positive spatial correlation between counties (i.e. clustering of counties with similar level of search intensity) and strong spatial spillover effects.

The remainder of this chapter is organized as follows. Section 3.6.1 sketches the simple theoretical model and its main predictions. Section 3.6.2 describes the data while the statistical models and the estimation results are contained in Section 3.6.3. Section 3.7 concludes the chapter.

3.6.1 A Simple Theoretical Model

We develop a simple model that explains how search effort decisions are made. For this purpose, we focus on the unemployed workers that are looking for a job in a given area i (e.g. a county or a region).

Let us first explain the macroeconomic environment in a given area i. Time is continuous and workers live forever. All workers are identical. A vacancy can be filled according to a random Poisson process. Similarly, unemployed workers can find a job according to a random Poisson process. In aggregate, these processes imply that there is a number of contacts (or matches) per unit of time between the two sides of the market in area i that are determined by the following standard matching function:

$$M_i \equiv M(s_i u_i, v_i)$$

where u_i and v_i respectively denote the number of unemployed workers and vacancies in area i. Each unemployed worker $j = 1, ..., u_i$ living in area i has a search intensity equal to $s_{ij} \equiv s(e_{ij})$, which depends on how much effort e_{ij} he/she provides in the search process. We assume that $s'(e_{ij}) > 0$ and $s''(e_{ij}) \leq 0$. Accordingly, s_i represents the average intensity of search of the u_i unemployed workers in area i.

As usual (Pissarides, 2000), $M(.)$ is assumed to be increasing in both its arguments, concave and exhibits constant returns to scale. As a result, the probability of obtaining a job per unit of time for an unemployed worker j in area i with search intensity $s_{ij} \equiv s(e_{ij})$ is given by:

$$\frac{s(e_{ij})M(s_i u_i, v_i)}{s_i u_i} = M\left(1, \theta_i\right) s(e_{ij})$$

where $\theta_i = v_i / s_i u_i$ is a measure of labor market tightness in search intensity units in area i. By using the properties of the matching function, it is easy to see that

$$\frac{\partial M\left(1, \theta_i\right)}{\partial \theta_i} > 0$$

since more vacancies in the area increase the probability to find a job whereas more unemployed decrease this probability.

We do not determine the labor market equilibrium. Rather, we focus on the behavior of an unemployed worker who searches for a job in area i and analyze how this behavior is affected by factors related to his/her residential location, such as living costs and the tightness of the local labor market.

Let us first determine the instantaneous utility function. All workers have identical preferences representable by a Cobb-Douglas utility. For the unemployed worker j living in area i, it is given by:[6]

$$U(z_{ij}) = z_{ij}^{\alpha} \tag{3.19}$$

with $0 < \alpha \leq 1$ and where z_{ij} is a composite good consumption. The budget constraint for the unemployed worker j living in i is equal to:

$$b = C(e_{ij}) + h_i z_{ij} \tag{3.20}$$

where b denotes the unemployment benefit, which is not area specific, h_i, is the cost of living in area i (i.e. the higher this price, the more expensive is to buy consumption goods and housing in area i), and $C(e_{ij})$ is the total cost of searching for jobs. The latter encompasses the costs of buying newspapers, commuting contacting friends, phone calls, interviews.... We assume that

$$\frac{\partial C(e_{ij})}{\partial e_{ij}} > 0 \quad , \frac{\partial^2 C(e_{ij})}{\partial e_{ij}^2} > 0$$

i.e. more search effort implies more search costs and it is even more costly at the margin (convex function).

If one denotes the *unemployed state* for workers by '0', and the *employed state* by '1', then using (3.19) and (3.20), we can derive the following indirect utility for each *unemployed* worker j in area i:

$$U^0(e_{ij}, h_i) = \left[\frac{b - C(e_{ij})}{h_i}\right]^{\alpha} \tag{3.21}$$

We are now equipped to write W_{ij}^0, the expected discounted lifetime utility of an unemployed worker j living in area i (Bellman equation). In steady-state, W_{ij}^0 is given

[6]To simplify we do not include leisure into the model but it should be clear that it does not alter the results. It only complicates the analysis.

by

$$
\begin{aligned}
rW_{ij}^0 &= U^0(e_{ij}, h_i) + M(1, \theta_i)\, s(e_{ij}) \left(W_i^1 - W_{ij}^0\right) \\
&= \left[\frac{b - C(e_{ij})}{h_i}\right]^\alpha + M(1, \theta_i)\, s(e_{ij}) \left(W_i^1 - W_{ij}^0\right)
\end{aligned}
\tag{3.22}
$$

where $r \in (0,1)$ is the discount rate and W_i^1, the expected discounted lifetime utility of an employed worker in area i. Equation (3.22) has a standard interpretation. When a worker is unemployed today, he/she obtains an instantaneous (indirect) utility equals to $U^0(e_{ij}, h_i)$. Then, he/she can get a job with a probability $M(1, \theta_i)\, s(e_{ij})$ and, if so, obtains an increase in utility of $W_i^1 - W_{ij}^0$.

Let us now study the search effort decision. When making this decision, the unemployed located in an area takes as given the total unemployment level u_i in area i, the total number of vacancies v_i in area i (and thus $\theta_i = v_i / s_i u_i$ the labor market tightness), the average cost of living h_i and the expected discounted lifetime utilities W_{ij}^0 and W_i^1.

By maximizing (3.22) with respect to e_{ij}, we obtain[7]

$$
\frac{\partial W_{ij}^0}{\partial e_{ij}} = \frac{\partial U^0(e_{ij}^*, h_i)}{\partial e_{ij}} + M(1, \theta_i)\, s'(e_{ij}^*) \left(W_i^1 - W_{ij}^0\right) = 0 \tag{3.23}
$$

where e_{ij}^* is the unique solution of this maximization problem and $s_{ij}^* \equiv s(e_{ij}^*)$ is the corresponding optimal search intensity.

Let us give the intuition of (3.23). When choosing e_{ij}^*, there is a fundamental trade-off between short-run and long-run benefits for an unemployed j located in area i. On the one hand, increasing search effort e_i is costly in the short run (more phone calls, more interviews, etc.) and it decreases instantaneous utility ($\partial U^0(e_{ij}^*, h_i)/\partial e_{ij} < 0$), but, on the other, it increases the long-run prospects of employment ($M(1, \theta)\, s'(e_{ij}^*) \left(W_i^1 - W_{ij}^0\right)$ is the marginal return of employment). We have the following result.

[7]Appendix 1 in Patacchini and Zenou (2003) shows that there is a unique solution to this maximization problem.

Proposition 16

(*i*) *The higher the cost of living* h_i *in a given area* i*, the higher the search intensity* $s_{ij}^* \equiv s(e_{ij}^*)$ *of an unemployed worker* j *living in area* i*;*

(*ii*) *The higher the labor market tightness* θ_i *in area* i*, i.e. the higher the number of vacancies* v_i *or the lower the unemployment level* u_i *in area* i*, the higher the search intensity* $s_{ij}^* \equiv s(e_{ij}^*)$ *for each unemployed worker* j *in this area.*

Proof. See Patacchini and Zenou (2003).

As stated above, when deciding the optimal level of search effort, each unemployed worker trades off the short run losses of increasing effort (higher cost of search effort $C(e_{ij})$ and thus lower instantaneous utility $U^0(e_{ij}, h_i)$) with the long-run gains (higher chance to get a job and to enjoy an intertemporal utility difference between employment and unemployment). Proposition 16 analyzes the effect of living costs h_i (short-run effect) and the one of the labor market tightness θ_i (long-run effect) on search effort e_{ij}^*.

When living costs increase, it becomes more costly to stay unemployed (see (3.20)), which reduces instantaneous utility $U^0(e_{ij}, h_i)$. As a result, the unemployed worker increases his/her search effort to raise his/her chance to obtain a job and thus be able to afford this new cost of living. The key relationship is in fact

$$\frac{\partial^2 U^0(e_{ij}, h_i)}{\partial e_{ij} \partial h_i} > 0$$

which is shown in Patacchini and Zenou (2003), Appendix 1 and states that the effect of e_{ij} on U^0 is even more negative when the living cost h_i increases.

Furthermore, when the labor market tightness rises, it becomes easier to find a job (there are relatively more jobs available compared to the unemployed) and thus the returns to search are higher. As a result, workers put more effort in search activities.

Let us now define the optimal average search intensity s_i^* of an area i as

$$s_i^* = \frac{1}{u_i} \sum_{j=1}^{j=u_i} s(e_{ij}^*(h_i, \theta_i)) \tag{3.24}$$

We have:

Proposition 17

(*i*) *The higher the cost of living h_i in a given area i, the higher the average search intensity s_i^* of this area;*

(*ii*) *The higher the labor market tightness θ_i in an area i, i.e. the higher the number of vacancies v_i or the lower the unemployment level u_i, the higher the average search intensity s_i^* of this area.*

Proof. See Patacchini and Zenou (2003).

These two results are a straightforward extension of Proposition 16 since, when we aggregate the search behavior of the unemployed, both h_i and θ_i do not change (i.e. h_i and θ_i are respectively equal to the average cost of living and labor market tightness in area i) so that if each individual searches more when h_i or θ_i increases, then, the average search intensity is also positively related to h_i and θ_i.

More generally, the basic message of this model is as follows. If we compare two areas (counties, cities, regions), the unemployed workers living either in the more expensive area and/or in the area with the higher labor market tightness, search more on average.

3.6.2 Data

We test the implications of the theoretical model using county-level data in England for the period 1991-2000. A key variable of the theoretical model is the average search intensity s_i. We consider as a measure of average search intensity in county i, hereafter the local search rate, the ratio between the number of unemployed that are actively looking for a job over the total number of unemployed in county i.[8] The other key variable in the theoretical model, the average cost of living in the area h_i, is measured by a county semidetached house price index. We are aware that the interactions between the labor market and the housing market are far more complicated (see e.g. Hughes and McCormick, 2000, Cameron and Muellbauer, 2001). However, because there is no complete set of sub-regional price indices for the UK, the main (and possibly the only) source of variation in prices within regions is differences in house prices. Furthermore, we concentrate our analysis only on young people (age 18-25) so that it is plausible to assume that, at least for the large majority of them, they are not home-owners and thus we rule out the possibility they can consider houses as assets. Because these empirical variables are not the straightforward observable counterpart of search intensity and living cost respectively, we treat them as variables measured with systematic errors. The discrepancies between the variables of interest and the observable ones are not supposed to be random. They might be also due to unobservable time-invariant county-specific effects such as unmeasured locational factors, and/or to county-invariant time-specific effects related for instance to some temporary effects of national policies. Very likely, the resulting measurement errors would follow a systematic rather than a random structure. Finally, the last variable of theoretical interest is the local labor market tightness θ_i. The

[8] In the questionnaire of our data base, the British Household Panel Survey (BHPS), people are asked their current labor force status and subsequently if they have been looking for any kind of paid job in the last four weeks. More precisely, regarding their labor force status, individuals can choose between "self-employed", "in paid employed", "unemployed", "retired", "family care", "full-time student", "long-term sick/disabled", "on maternity leave", "government training scheme", "something else". In our sample, we have only included individuals that have responded "unemployed" to this question. Among them, there is surprisingly a high number who state that they have not looked for a job during the last four weeks. Our search intensity variable is thus the ratio between individuals that declare themselves as "unemployed" and "have looked actively for a job during the last four weeks" *and* all individuals that have responded "unemployed" to the question above. For robustness check, we have also used another measure of search intensity: a derived variable from the Labour Force Survey (LFS), based on the standard (ILO) definition of economic activity. The analysis with this other measure is discussed at the end of Section 3.6.2. We are aware that these are largely imperfect measures of search intensity (they will be treated as variables contaminated by systematic measurement errors in the econometric analysis). The ideal variable to measure search effort would have been, at the individual level, the number of hours spent looking for a job. Unfortunately, this variable is not available in any British survey. This is why we resort to our aggregate indicators of search intensity and, as a result, all our empirical analysis will be conducted at an aggregate level (i.e. county level).

National On-line Manpower Information Service (NOMIS) provides information of the labor market tightness at the county level. Although, this is not an error-free measure of the true county labor market tightness, we assume the presence of a random rather than systematic measurement error.

A longitudinal data set of English counties observed yearly for the period 1991-2000 has been constructed. Three different data sources have been used. The estimated local search rates has been constructed using the waves of the BHPS, that are available also on line in the ESRC Data Archive. The information about the features of the counties housing market comes from the semidetached Halifax House Price Index.[9] The remaining indicators of the local labor markets have been derived using data available from the National On-line Manpower Information Service (NOMIS) located at the University of Durham. Appendix 2.3 contains the list of the 45 English counties considered in the sample used for estimation.[10]. Appendix 1.3 contains precise definitions for all variables. The related descriptive statistics are collected in the following table.

Table 1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min.	Max.
s_{it}	450	0.6899	0.2984	0	1
$\bar{s}_{it}$	450	0.6762	0.2011	0	1
θ_{it}	450	0.1707	0.1686	0.0193	1.0878
$\bar{\theta}_{it}$	450	0.1794	0.1580	0.038	0.8052
h_{it}	450	63.406	19.356	40.648	188.263
$\bar{h}_{it}$	450	66.39	23.777	34.456	166.651
d_{it}	450	6.187	8.382	0.6	45.245

[9] The index numbers are constructed using a Laspeyres type price index methodology. The weighted average prices in each current period is compared with the weighted average price in the base period. For the Halifax House Price Index this has been chosen as 1983. For further details see http://www.hbosplc.com/view/housepriceindex/indexmethodology02.asp

[10] Out of the 47 counties designated in England, we had to eliminate the Isle of Wight because the figures of the Halifax price index are not available for the years 1991-1997 and 2000 and we needed to merge North and South Humberside because these two counties are unite into Humberside in the Nomis database. Thus we ended up with 45 cross sectional units.

This table shows some interesting features. First, in our sample, the average search intensity in a county s_{it} is around 70%, which means that there are on average 30% young workers entering the labor market not actively looking for a job (even if they declare themselves as unemployed). Second, the labor market tightness θ_{it} is on average equal to 17%. This means that, on average, there is almost 1 vacancy for every 5 unemployed workers in a county. Finally, the features related to houses prices (h_{it} and $\overline{h}_{it}$) show a large variation between different counties in England.

3.6.3 Statistical Model and Estimation Results

Our empirical strategy is to test the results of Proposition 17, namely the positive relationships between s_i and h_i and between s_i and θ_i. As already noted above, there are *systematic measurement errors* on s_i and h_i that need to be taken into account in the econometric specification of the empirical model. Moreover, since search intensity s_i in county i may be affected by search intensities $\overline{s}_i$ in neighboring counties (for example individuals may live in county i but search in a neighboring county if the latter offers better labor outcomes), we consider *cross-sectional dependence* in our analysis. Also, since there is *unobserved heterogeneity* among counties, we undertake a panel data analysis to control for individual unobservable effects. Finally, s_{it}, search intensity in county i in period t may also be affected by $s_{i,t-1}$ the search intensity in the same county but in the previous period (for example because of the presence of long term unemployed). As a result, a *dynamic analysis* should be considered.

To take into account these four features, we estimate a spatio-temporal model specified as a typical dynamic panel data model where a spatially lagged dependent variable has been included. The advantage in using panel data models is not only the possibility to control for unobserved heterogeneity but also to allow for measurement errors in observed variables. The advantage in using spatial econometric techniques is to control for spatial effects (spatial heterogeneity as well as spatial correlations) between counties. Indeed, a feature often neglected in empirical studies using dynamic panel data models when the units have a spatial connotation, is the possible cross-sectional dependence of the residuals. The degree of interdependence between markets in regional studies, for instance, is usually very high and studies lacking to control for it lead to unreliable estimation results. We test for and find high and positive spatial autocorrelation among levels of search intensity between counties, meaning that high (low) values of search intensity in a location tend to be associated with high (low) values at nearby locations. Spatial correlation may arise for a number of reasons. Instead of trying to correctly specify these channels of interdependence, we incorporate in the model a spatially lagged dependent variable and fixed effects in order to explain the spatial correlation and spatial heterogeneity respectively in the data.

The inclusion of a spatially lagged dependent variable in addition to other explanatory variables can be interpreted in two different ways. If the main empirical interest is the spatial effects, one can consider the inclusion of a spatially lagged dependent variable in addition to other explanatory variables as a way to assess the degree of spatial dependence, while controlling for the effects of these other variables. Alternatively, the inclusion of a spatially lagged dependent variable allows us to control for spatial dependence and, having done so, to assess the significance of the other (non-spatial) variables. This latter strategy is the one pursued in our analysis. Our aim is to estimate the impact of the (county) cost of living and the (county) labor market tightness on the (county) search intensity once spatial effects have been filtered out. The formulation of the model is such that specification tests on the model in deviations cannot reject the null hypothesis of no serial or spatial correlation in the errors. Appendix 3.3 presents the tests for spatial correlation both on the observations (Table A.1) and on the residuals (Table A.2). It also contains three quantiles maps (Figure 3.1) that illustrate the geographical distribution of the local search rate, the tightness of the local labor markets and our proxy for costs of living, i.e. house prices, in England. It appears evident that most of the areas with high (low) levels of local search rate are the areas with high (low) levels of local labor market and cost of living.

Let us now write the econometric specification of the model that incorporates all the four features mentioned above. For that, define as s_{it}^* and h_{it}^* the true local search rate and the true local cost of living respectively, and as s_{it} and h_{it} their empirical counterparts. We assume that the process of the measurement error, m_{it}, has the additive structure specified in model (3.13)-(3.15) (same notation applies). In order to take into account cross-sectional dependence, we also define for county $i = 1, ..., n$ the variable

$$\bar{s}_{it}^* = \sum_{j=1}^{n} w_{ij} s_{it}^*, \tag{3.25}$$

which indicates the average value of the search rate over the counties adjacent to i, i.e. the counties that share a common boundary with i. The weights w_{ij} are set equal to 0 if $i = j$ or if i and j are not adjacent, and are equal to a constant otherwise (defined by imposing the normalization $\sum_{j=1}^{n} w_{ij} = 1$ for each i).[11]

Finally, in order to capture some determinants of unemployed people behavior when they do not declare themselves actively looking for a job, we include in the model population density. The reasoning is that in denser areas (cities, metropolitan areas) there are more job opportunities and unemployed people do not need much effort in searching for a job. Furthermore, in cities there are higher opportunities to work in the black economy so that possibly unemployed workers do not invest much effort in looking for a regular job. If the job search related question in the BHPS questionnaire is perceived as referring to an intense and time consuming process in the search for jobs, unemployed people living in denser areas may be less likely to declare themselves actively looking for a job. As a result the density variable is meant to capture this agglomeration effect and it should be inversely correlated with our measure of search intensity.

[11]The $N \times N$ matrix $W = \{w_{ij}\}$ is sometimes called the contiguity matrix in the spatial statistics literature. It describes the geographical arrangement of the spatial units.

We are now able to write the empirical model (referred to as model 1). It is given by:

$$
\begin{aligned}
s_{it}^{*} &= \alpha s_{i,t-1}^{*} + \beta \bar{s}_{it}^{*} + \gamma \theta_{it} + \delta h_{it}^{*} + \varphi d_{it} + \eta_i + \varepsilon_{it}, \\
|\alpha| &< 1, \qquad i = 1, ..., N;\ t = 2, ..., T, \\
s_{it}^{*} &= s_{it} + m_{sit}, \\
h_{it}^{*} &= h_{it} + m_{hit}, \\
m_{xit} &= \phi_{xt} + \mu_{xi} + \xi_{xit}, x = s, h
\end{aligned} \tag{3.26}
$$

where θ_{it} denotes the local labor market tightness in county i at time t, d_{it} is the population density in county i at time t, η_i is a county-specific constant capturing also spatial effects due for instance to different county size (spatial heterogeneity) and ε_{it} is a white noise disturbance term. m_{sit} and m_{hit} are measurement errors with time effects ϕ_{st}, ϕ_{ht}, group effects μ_{si}, μ_{hi}, and random components ξ_{sit}, ξ_{hit} which are mutually independent, and i.i.d, $t = 1, .., T, i = 1, .., N$. Observe that the empirical model does not include any measure of the average human capital characteristics of the different counties, nor other features of the local structure of the population. The reason is that we assume that the impact of these characteristics on the local search rate in each county is captured through the inclusion of (time) lagged values of the local search rate.

The first order space-time autoregressive model 1 is estimated using an instrumental variables approach within a Generalized Method of Moments (GMM) estimation procedure. After controlling for spatial dependence in the data by choosing an appropriate order in the spatial process, the literature on dynamic panel data models can be used. Distributional assumptions are not needed. Measurement errors in observed variables are taken into account by using sufficiently lagged variables as instruments. Technical details on the estimation procedure of the resulting dynamic panel data model, which presents some non-standard properties due to the error structure, are described in Section 3.5. The estimation results of model 1 are contained in the first two columns of Table 2.[12]

[12] All the reported results are two-step GMM estimators, obtained using the DPD98 package for Ox

They are short run effects.

The first column reports the results for the Arellano and Bond (1991) GMM estimator (GMM-DIF), which consists in taking first differences over time to get rid of the unit specific error terms and in using appropriate instruments for the lagged (spatially and temporally) dependent variable and for all the others endogenous variables. Both living cost and the tightness of the local labor markets are treated as potentially endogenous variables. Therefore, the instrumental set contains observations on the tightness of the local labor markets dated $(t-2)$ and earlier and observations on local cost of living, local search rate and search rate in neighbor counties dated $(t-3)$ and earlier. Note that the use of three-times periods lagged variables instead of the standard two-times periods lagged ones for the variables indicating the cost of living and the search rate is due to the additional endogeneity problem caused by the presumed presence of measurement errors. Under the specified assumptions for the composed error structure, valid instruments are only obtained by using variables that are at least three-times periods lagged, as shown in Section 3.5. We do not use the whole history of the variables as instruments. We truncated the history after $(t-5)$. Although the number of overidentifying restrictions is still rather large compared to the sample size, we do not find any evidence of a possible overfitting bias. Table 2 also reports the Sargan tests of the overidentifying restrictions (Sargan, 1958; Hansen, 1982) implied by the instruments matrix and the tests for autocorrelation. The Sargan test is asymptotically distributed as chi-squared under the null of instruments validity, with degrees of freedom (df) reported in parentheses. AR(1) and AR(2) are tests for first-order and second-order serial correlation in the first-differenced residuals asymptotically distributed as $N(0,1)$ under the null of no serial correlation (Arellano and Bond, 1991). The consistency of the GMM estimators requires the absence of serial correlation in the original error term. In turn, this requires

(Arellano, Bond and Doornik, 1999). One step results are not considered because we deal only with 45 cross section units and the estimated standard errors are severely downward biased in small sample. Windmeijer (2000) derived a small-sample correction which is implemented in the two-step estimation routine.

negative first-order, but no second-order correlation in the differenced error term. No evidence of misspecification is revealed in Table 2.

Let us now interpret the results of the first column of Table 2 (GMM-DIF). As predicted by the theoretical model, *both the (local) cost of living h and the (local) labor market tightness θ are found to have a positive and significant effect on unemployed search intensity.* To be more precise, a unit increase in the cost of living in a county implies a 0.14 increase in average search intensity in the county.[13] Furthermore, a unit increase in the level of labor market tightness θ increases search effort by 0.20.[14]

Although a spatially lagged dependent variable ($\bar{s}$) has been included in the model only to control for spatial correlation, and it is not a target variable, the estimated coefficient, α, is significant, it presents an interesting positive sign and it is of a large magnitude. Indeed, counties where people have an active search behavior in the labor market appear to be clustered together (because of the positive spatial correlation that we found) and are also strongly interrelated. In other words, *counties that have high (low) search intensity tend to be geographical adjacent to counties are also characterized by high (low) search intensity, with important spatial spillover effects.*

The coefficient of the population density d, has the expected negative sign but it is not significantly different from zero.

[13]Observe that, because the proxy used for (local) living costs is (local) house prices and not an index of all the consumption goods, the effect of the cost of living on search intensity should be smaller. However, since housing constitutes an important part of the household expenses, the difference should not be very big.

[14]To be sure that our estimates are not affected by reverse causality between local search rate and local houses prices, model 1 has also been estimated instrumenting the cost of living by taking the historical prices. The Halifax price index at the beginning of 1988, the first period of the available series, has been used as instruments. The reasoning underlying is that today's house prices are correlated with the historical prices but probably no determinants of today's local search rate in a county are affected by local house prices in 1988. The qualitative estimation results remain qualitatively unchanged. Therefore they are not reported here, but are available upon request.

Table 2. Estimation Results ($N = 45,\ T = 10$)

Dependent variable: Local search rate, s, at time t

	GMM-DIF$^{\diamond}$ (model 1)	GMM-SYS$^{\diamond\diamond}$ (model 1)	GMM-DIF$^{\bullet}$ (model 2)	GMM-SYS$^{\bullet\bullet}$ (model 2)
s_{t-1}	-0.2625**	-0.2136***	-0.1404**	-0.1674***
	(0.1221)	(0.0649)	(0.0621)	(0.0528)
$\bar{s}$	0.5932**	0.6287***	0.4234***	0.4542***
	(0.2864)	(0.2291)	(0.1567)	(0.1232)
θ	0.1999***	0.2342***	0.0224	0.0698
	(0.0751)	(0.0557)	(0.0780)	(0.0552)
$\bar{\theta}$	-	-	0.2841**	0.3146***
			(0.1235)	(0.1101)
h	0.1356**	0.1510***	0.0686	0.0026
	(0.0671)	(0.0459)	(0.1998)	(0.0724)
$\bar{h}$	-	-	0.2011**	0.2292***
			(0.0987)	(0.0653)
d	-0.0140	-0.0125*	-0.1207	-0.0645
	(0.0091)	(0.0067)	(0.1005)	(0.0865)
AR(1)	-3.299	-3.926	-3.594	-4.033
AR(2)	-1.310	0.3586	1.302	0.401
Sargan	26.52	33.03	15.18	18.14
(df)	(118)	(147)	(182)	(226)

Notes:

1. Year dummies are included in all specifications.

2. Asymptotic standard errors, using the small sample Windmeijer (2000) correction, are reported in parentheses.

3. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

4. Instruments used in each equation:

$\diamond$: $s_{i,t-3}, s_{i(t-4)}, \ldots s_{i1}; \bar{s}_{i,t-3}, \bar{s}_{i(t-4)}, \ldots \bar{s}_{i1}; h_{i,t-3}, h_{i(t-4)}, \ldots h_{i1}; \theta_{i,t-2}, \theta_{i,t-3}, \ldots \theta_{i1}$.

$\diamond\diamond$: $s_{i,t-3}, s_{i(t-4)}, \ldots s_{i1}; \bar{s}_{i,t-3}, \bar{s}_{i(t-4)}, \ldots \bar{s}_{i1}; h_{i,t-3}, h_{i(t-4)}, \ldots h_{i1}; \theta_{i,t-2}, \theta_{i,t-3}, \ldots \theta_{i1}; \Delta s_{i,t-2}; \Delta \bar{s}_{i,t-2}; \Delta h_{i,t-2}; \Delta \theta_{i,t-1}$.

$\bullet$: $s_{i,t-3}, s_{i(t-4)}, \ldots s_{i1}; \bar{s}_{i,t-3}, \bar{s}_{i(t-4)}, \ldots \bar{s}_{i1}; h_{i,t-3}, h_{i(t-4)}, \ldots h_{i1}; \bar{h}_{i,t-3}, \bar{h}_{i(t-4)}, \ldots \bar{h}_{i1}; \theta_{i,t-2}, \theta_{i,t-3}, \ldots \theta_{i1}; \bar{\theta}_{i,t-2}, \bar{\theta}_{i,t-3}, \ldots \bar{\theta}_{i1}$.

$\bullet\bullet$: $s_{i,t-3}, s_{i(t-4)}, \ldots s_{i1}; \bar{s}_{i,t-3}, \bar{s}_{i(t-4)}, \ldots \bar{s}_{i1}; h_{i,t-3}, h_{i(t-4)}, \ldots h_{i1}; \theta_{i,t-2}, \theta_{i,t-3}, \ldots \theta_{i1}; \Delta s_{i,t-2}; \Delta \bar{s}_{i,t-2}; \Delta h_{i,t-2}; \Delta \bar{h}_{i,t-2}; \Delta \theta_{i,t-1}; \Delta \bar{\theta}_{i,t-1}$.

Let us now focus on the second column of Table 2. A more precise GMM estimator can be obtained by combining the set of moment conditions relating to the equations in levels with a set of moment conditions relating to the equations in first differences. However, the validity of the extra instruments for the equations in levels, meaning that they are uncorrelated with the area-specific effects, has to be tested. Full details and references are given in Section 3.5. In our analysis, we find that the extended set of moment restrictions is not rejected by the Sargan test of over-identifying restrictions. No evidence of serial correlation in the original errors is provided.

It is easy to see that the estimated values of the coefficients of the second column of Table 2 (GMM-SYS) are very similar to the ones of the first column (GMM-DIF). This is consistent with the underlying econometric theory since a dramatic improvement in performance of the combined GMM (GMM-SYS) compared to the usual first-differences GMM (GMM-DIF) usually occurs with very short sample periods and persistent series or if the variance of the county effects η_i exceeds the variance of the residuals ε_{it}. Because these features are not present in our case (moderate number of points in time, small autoregressive parameter and $Var(\eta_i) < Var(\varepsilon_{it})$), the similarity in the figures of the two columns is not unexpected. The gain in precision resulting in smaller standard errors in the second column are due to the use of valid additional moment restrictions. The important implication for our analysis is that the strong and positive association between average search intensity, costs of living and labor market tightness appears to be confirmed and reliable.[15] Finally, the coefficient of the population density, d, in this second column (i.e. when the coefficient are more precisely estimated) still retains the expected negative sign but it is now also significant. This may be interpreted as evidence of the important role of agglomeration effects on unemployed behavior when searching for a job.

[15]In order to test the sensitivity of the results to the measure of the tightness of the (local) labor market θ, we have also estimated an alternative specification of model (3.26) using a measure of θ based on flows rather than stocks, i.e. θ is measured by the yearly ratio between monthly notified vacancies and unemployed on-flows. The estimation results obtained are similar to the ones of the first two columns of Table 2 and are thus not reported here.

Because the coefficient of the search rate in neighboring counties is positive and significant, we investigate further the presence of spatial effects using the following formulation (referred to as model 2):

$$s_{it}^* = \alpha s_{i,t-1}^* + \beta \bar{s}_{it}^* + \gamma \theta_{it} + \gamma_1 \bar{\theta}_{it} + \delta h_{it}^* + \delta_1 \bar{h}_{it}^* + \varphi d_{it} + \eta_i + \varepsilon_{it}, \tag{3.27}$$

$$|\alpha| < 1, \qquad i = 1, ..., N; \ T = 2, ..., T,$$

where the variables that were already in model 1 are defined in exactly the same way and where the spatial averages $\bar{\theta}_{it}$ and $\bar{h}_{it}^*$ are defined in a similar way as in (3.25).

The last two columns of Table 2 contain the estimation results for model 2. Even though not in the theoretical model, these results have been reported in order to confirm the strong spatial interdependence between local labor markets in England, as suggested by the preliminary tests for spatial correlation. The diagnostic tests (AR(1), AR(2), Sargan) suggest that the model is well specified, the instruments appear to be valid and the errors are white noise. All the coefficients show the expected sign suggesting that the theoretical predictions are confirmed, but, once the values of a variable in neighboring locations are introduced into the model, the coefficients of the key variables, h and θ, retain their positive signs but lose their significance. This indicates possible multicollinearity between local and neighboring values, suggesting a high degree of correlation between them and thus strong spatial spillovers. In this second specification of the model, the estimated coefficients of population density, d, although with the expected sign, are again not significant.[16]

Finally, one may object that our measure of unemployed in the definition of s_{it} from the BHPS (see Appendix 1.3) is based on self-reporting behavior (see our discussion in footnote 7). For robustness check, we also estimate the empirical model 1 using data from

[16]The estimation results of both model 1 and 2 without the inclusion of the density variable appear qualitatively unchanged (that is the coefficients of the target variables, h and θ, remain positive and significant in both models). However, given the significancy of the density variable in model 1, we report the estimation results with the inclusion of this variable.

the Labour Force Survey (LFS)[17] aggregated yearly at county level for a comparable time period (1992-2000). In this data set, unemployment is now defined according to the standard ILO definition. Our dependent variable is not anymore s^*_{it} but na^*_{it}, the inactivity rate in county i at time t (for a precise definition of our measure of inactivity rate, na_{it}, see Appendix 1.3). On the right hand side of equation (3.26), $s_{i,t-1}$ and $\overline{s}^*_{it}$ have also been replaced by $na_{i,t-1}$ and $\overline{na}^*_{it}$, that are similarly defined. We believe that the inactivity rate is an inverse measure of the search intensity rate since when it has a high value in a county this implies that individuals are not actively searching for a job. As a result, we expect to obtain reverse signs for h and θ since counties with larger cost of living and/or higher labor market tightness should have lower inactivity rate. Our results (that are not reported here but are available upon request) show indeed the estimated coefficients of the (local) cost of living and the (local) labor market tightness are now negative and significant for both variables.[18]

3.7 Conclusions

This chapter has discussed estimation procedures for an autoregressive panel data model in presence of measurement errors in the observed variables and unobservable heterogeneity among cross-sectional units. The analysis of a concrete case of study illustrates that this model can be useful to investigate relationships among economic variables. In the empirical context considered, most of the target variables have no straightforward observable counterpart and the effects of other kinds of unobservable factors, resulting in spatial correlation, are also assumed to be important. It is shown to what extent the model and the relative estimation methodology can be adjusted and applied in order to get reliable results in the context analyzed.

[17] All the LFS data are also available on line in the ESRC data archive. Observe that another advantage of using the LFS compared to the BHPS is that the sub-sample relevant for our analysis has a larger number of observations.

[18] All the empirical analysis presented in this chapter has also been performed excluding London. The estimation results are qualitatively unchanged and thus not reported here.

Specifically, using an English panel of county level data, the empirical model finds evidence that search intensity is higher in areas characterized by larger cost of living and higher labor market tightness. These findings are consistent with the predictions of the simple model sketched in Section 3.6.1. They imply that the market-level average search effort increases as labor market tightness or regional commodity prices increase. However, we are aware that the econometric specification departs from the model on many aspects. In particular, the econometric specification is dynamic whereas the theoretical model is written at steady state and does not include any dynamic process for search intensity; the empirical model uses the average search intensity of local neighbors as an explanatory variable whereas the theoretical model implicitly states that the local labor markets are segmented and does not propose any interactions between local markets; the country-specific effects and the disturbance terms do not have a clear economic meaning in terms of unobserved heterogeneity in the theoretical framework; agglomeration processes are not modelled; search intensity is binary in the data at the micro level and the theoretical model may be rewritten with only two discrete search intensities (0 and 1). The next step of this research is to extend the theoretical model to be more consistent with the empirical work. In particular, a potentially interesting empirical result is spatial correlation. The first aim in our future work is to model explicitly the interactions between local labor markets (e.g. by defining search intensity in terms of number of local markets visited and/or by defining job matches (hirings) in one county in terms of both local market tightness and the one in adjacent counties as well as by considering transport costs).

3.8 Appendix 1.3

Description of Variables

s_{it}: Ratio between unemployed persons aged between 18 and 25 actively searching for a job and unemployed between 18 and 25 in county i at time t. An active job seeker is a person who was neither "at work" nor "with a job but not at work" during the week before the reference day and that has taken active steps to find a work (applied directly to employer, replied to adverts, used job centre or employment agency, asked friends or contacts, taken steps to start own business). Source: BHPS, waves 1-10, 1991-2000.

$\bar{s}_{it}$: Average s_{it} in the counties that share a boundary with county i. Source: BHPS, waves 1-10, 1991-2000.

θ_{it}: Ratio between the stocks of unfilled vacancies and unemployed in county i at time t. Source: NOMIS.

$\bar{\theta}_{it}$: Average θ_{it} in the counties that share a boundary with county i. Source: NOMIS.

h_{it}: Average yearly semidetached Halifax price index for county i at time t. The index is the arithmetic average prices of houses on which an offer of mortgage has been granted, constructed on a quarterly base. The yearly average has been calculated on the available quarterly values. Source: Group Economics, Halifax plc.

$\bar{h}_{it}$: Average h_{it} in the counties that share a boundary with county i. Source: Group Economics, Halifax plc

d_{it}: Ratio of residents over squared hectometers in county i at time t. Variable taken from the 1991 Census database and subsequently updated using the Midyear Population Estimates. Source: NOMIS. Years: 1992-2000.

na_{it}: Ratio between inactive persons (between 18 and 25 years old) -not seeking job but willing to work- and inactive and unemployed persons (between 18 and 25 years old) in county i at time t. Source: LFS-INECA variable. It is a derived variable which classifies the individual economic activity according to the ILO standard definitions.

$\overline{na}_{it}$: Average na_{it} in the counties that share a boundary with county i. Source: LFS-INECA variable.

3.9 Appendix 2.3

List of English counties

Avon

Bedfordshire

Berkshire

Buckinghamshire

Cambridgeshire

Cheshire

Cleveland

Cornwall & Isles of Scilly

Cumbria

Derbyshire

Devon

Dorset

Durham

East Sussex

Essex

Gloucestershire

Greater London

Greater Manchester

Hampshire

Hereford and Worcester

Hertfordshire

Humberside

Kent

Lancashire

Leicestershire

Lincolnshire

Merseyside

Norfolk

North Yorkshire

Northamptonshire

Northumberland

Nottinghamshire

Oxfordshire

Shropshire

Somerset

South Yorkshire

Staffordshire

Suffolk

Surrey

Tyne and Wear

Warwickshire

West Midlands

West Sussex

West Yorkshire

Wiltshire

3.10 Appendix 3.3

Measures of global spatial autocorrelation

When the variable under investigation is measured on a continuous scale, the measurement of global spatial autocorrelation is usually based on Moran's I and Geary's c statistics (Cliff and Ord, 1973, 1981; Upton and Fingleton 1985). They measure the deviation from spatial randomness, or the lack of any pattern. Under this assumption, any grouping of high or low values in a particular area would be totally spurious. The existence of a spatial structure is detected by the presence of spatial correlation, that can be defined as the coincidence of value similarity with locational similarity (Anselin, 2001). There is positive spatial autocorrelation when high or low values of a random variable tend to cluster in space (spatial clustering) and there is negative spatial autocorrelation when geographical areas tend to be surrounded by neighbors with very dissimilar values (spatial outliers). Moran's I is defined as

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{3.28}$$

where n is the number of observations, z_i are variables in deviations from the mean, w_{ij} are elements of a spatial weights matrix, that indicates the way area i is spatially connected to area j, and S_0 is a scaling factor equal to the sum of all the elements in the weight matrix.

Geary's c is defined as

$$c = \frac{n-1}{2S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2}{\sum_{i=1}^{n} z_i^2}$$

where the $x_i s$ are the original variables and the other notation is as above (formula 3.28).

Moran's I is a cross product coefficient scaled to be less than one. Positive values for Moran's I indicate positive spatial correlation, while negative values indicate negative spatial correlation. In contrast, Geary's c coefficient is based on squared deviations. Values of Geary's c less than one indicate positive spatial correlation, while values larger than one suggest negative spatial correlation.

Table A.1 reports Moran' s I statistic and Geary's c statistic of the search rate for each year of the period 1991-2000, for the counties in England (column two). Inference is based on a conventional normality approach. The third column in Table A.1 reports the standardized z-value for I and c, computed by subtracting the expected value and dividing by the standard deviation assuming an approximation of the (asymptotic) distributions of I and c by the normal distribution. The associated significance level, p_1, is reported in column four. Table A.1 shows clearly that (local) search rates are positively spatially autocorrelated in every year. Both I and c statistics are highly significant (the indicators of significance, p_1, are always almost 0) and display clear evidence of positive spatial autocorrelation of the variable under analysis (positive value for the standardized Moran's I, $z(I)$, and negative values for the standardized Geary's c statistic, $z(c)$). Table A.2 reports Moran's I statistic and Geary's c statistic calculated on the residuals of model 1. It has the same structure of Table A.1. It shows that both I and c statistics are no longer significant (at 5% significance level) in any year confirming that the spatial dependence has been adequately dealt with by incorporating the spatial lag term.[19]

[19]Note that the asymptotic validity of these spatial correlation tests when applied to residuals has not been formally established.

Table A.1: Measures of Global Spatial Correlation

Search Rate

Moran's I test for spatial autocorrelation

Year	I	$z(I)$	p_1
1991	0.1963	8.9542	0.0000
1992	0.2396	10.8744	0.0000
1993	0.2235	6.6758	0.0000
1994	0.2678	7.9698	0.0000
1995	0.1309	6.1089	0.0000
1996	0.1510	6.9777	0.0000
1997	0.1949	8.9158	0.0000
1998	0.2290	10.4729	0.0000
1999	0.1353	6.2787	0.0000
2000	0.1447	6.6827	0.0000

Geary's c test for spatial autocorrelation

Year	c	$z(c)$	p_1
1991	0.7571	−7.8190	0.0000
1992	0.7208	−8.9887	0.0000
1993	0.7188	−6.2914	0.0000
1994	0.6926	−6.8789	0.0000
1995	0.8312	−5.4334	0.0000
1996	0.7958	−6.5707	0.0000
1997	0.7490	−8.0815	0.0000
1998	0.7263	−8.8105	0.0000
1999	0.8435	−5.0361	0.0000
2000	0.8124	−6.0389	0.0000

Table A.2: Measures of Global Spatial Correlation

Residuals from model 1

Moran's I test for spatial autocorrelation

Year	I	$z(I)$	p_1
1991	0.0289	1.0755	0.2821
1992	0.0432	1.4871	0.1370
1993	0.0059	0.4136	0.6792
1994	−0.0164	−0.229	0.8188
1995	−0.0051	0.0961	0.9234
1996	−0.0397	−0.8989	0.3687
1997	−0.0583	−1.4352	0.1512
1998	0.1517	0.9487	0.3428
1999	−0.0019	0.1891	0.8500
2000	0.0485	0.5480	0.5837

Geary's c test for spatial autocorrelation

Year	c	$z(c)$	p_1
1991	0.9303	−1.5596	0.1188
1992	0.9618	−0.8555	0.3922
1993	0.9264	−1.6473	0.0995
1994	1.0882	1.1951	0.2320
1995	1.0787	1.1025	0.2702
1996	1.0866	1.1899	0.2341
1997	0.9181	−1.6557	0.0974
1998	0.9507	−0.9512	0.3415
1999	0.9988	−0.8205	0.4119
2000	0.9771	−0.8904	0.3732

The remainder of this Appendix shows in Figure 3-1 the geographical distribution of the search rate (first panel), the tightness of the local labor markets (panel on the left) and our proxy for costs of living, i.e. house prices, (panel on the right) in England at the NUTS3 level of spatial disaggregation for the year 2000. Extremely similar maps can be obtained for all the other years considered in the analysis. Therefore they are not reported here.[20]

[20]All computations and maps are carried out using Spacestat 1.90 and Arcview 3.2.

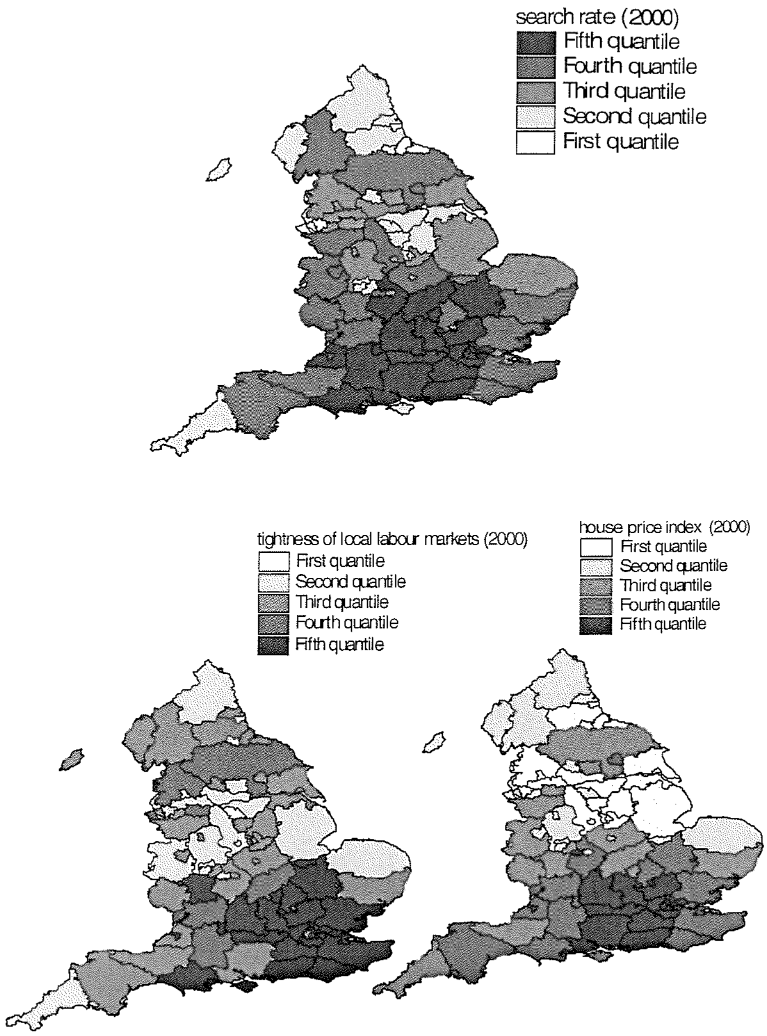


Figure 3-1: Quantile maps for England

Bibliography

[1] Ahn, S.C. and Low, S. (1996). A Reformulation of the Hausman Test for Regression Models with Pooled Cross Sectional Time Series Data, *Journal of Econometrics*, 71, 309-319.

[2] Aigner, D.J., Hsiao, C., Kapteyn, A. and Wansbeek, T.J. (1984). *Latent Variable Models in Econometrics*, Handbook of Econometrics, Vol. H, Griliches, Z. and Intriligator, M.D. (eds), North Holland, Amsterdam.

[3] Anselin, L., (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Boston.

[4] Arellano, M. (1993). On the Testing of Correlated Effects with Panel Data, *Journal of Econometrics*, 59, 87-97.

[5] Arellano, M. and Bond, S. (1991). Some Test of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.

[6] Arellano, M., Bond, S. and Doornik, J.A. (1999). Panel Data Estimation Using DPD for Ox, unpublished.

[7] Arellano, M. and Bover, O. (1995). Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

[8] Balestra, P. and Nerlove, M. (1966). Pooling Cross Section and Time Series Data, *Econometrica*, 34, 585-612.

[9] Baltagi, B.H. (1995). *Econometric Analysis of Panel Data*, New York: John Wiley and Sons.

[10] Baltagi, B.H. (1996). Testing for Random Individual and Time Effects using a Gauss-Newton Regression, *Economics Letters*, 50, 189-192.

[11] Baltagi, B.H. (1997). Problems and Solutions, *Econometric Theory*, 13, 757.

[12] Baltagi, B.H. (1998). Problems and Solutions, *Econometric Theory*, 14, 527.

[13] Biorn, E. (2000). Panel Data with Measurement Errors: Instrumental Variables and GMM procedures combining Levels and Differences, *Econometric Reviews*, 19, 391-424.

[14] Blundell, R. and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87,115-143.

[15] Bonferroni, C.E. (1936). Teoria Statistica delle classi e calcolo delle probabilità. Pubblicazioni del R. Istituto superiore di Scienze Economiche e Commerciali di Firenze, 8, 1-62.

[16] Bowden, R.J. and Turkington, D.A. (1984). *Instrumental Variables*, Cambridge University Press.

[17] Burda, M. and Profit, S. (1996). Matching across Space: Evidence on Mobility in the Czech Republic, *Labour Economics*, 3, 255-278.

[18] Burgess, S. and Profit, S. (2001). Externalities in the Matching of Workers and Firms in Britain, CEP Discussion Paper n. 490, London School of Economics.

[19] Cameron, G. and Muellbauer, J. (2001). Earnings, Unemployment and Housing in Britain, *Journal of Applied Econometrics*, 16, 203-220.

[20] Chamberlain, G. (1982). Multivariate Regression Models for Panel Data, *Journal of Econometrics,* 18, 5-46.

[21] Cliff, A.D. and Ord, J.K. (1973). *Spatial Autocorrelation*, London: Pion Limited.

[22] Cliff AD. and Ord, J.K. (1981). *Spatial Processes: Models and Applications*, London: Pion Limited.

[23] Coulson, E., Laing, D. and Wang, P. (2001). Spatial Mismatch in Search Equilibrium, *Journal of Labor Economics*, 19, 949-972.

[24] Coles, M.G. and Smith, E. (1996). Cross Sectional estimation of the Matching Function: Evidence from England and Wales, *Economica*, 63, 589-597.

[25] Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press.

[26] Durbin, J. (1954). Errors in Variables, *Review of the International Statistical Institute*, 22, 23-32.

[27] Golub, G.H. and Van Loan, C.F. (1983). *Matrix Computations*, third edition, North Oxford Academic Press.

[28] Griliches, Z. (1979). Sibling Models and Data in Economics: Beginnings of a Survey, *Journal of Political Economy,* vol. 78, n. 5, part 2.

[29] Griliches, Z. (1986). *Economic Data Issues*, Handbook of Econometrics, Vol. III, Griliches, Z. and Intriligator, M.D. (eds), North Holland, Amsterdam.

[30] Griliches, Z. and Hausman, J.A. (1986). Error in Variables in Panel Data, *Journal of Econometrics*, 31, 93-118.

[31] Hansen, L. (1982). Large Sample Properties of Generalized Method of Moment Estimators, *Econometrica*, 50, 517-533.

[32] Hausman, J.A. (1978). Specification Tests in Econometrics, *Econometrica*, 46, 1251-1271.

[33] Hausman, J.A., Newey, W.K. and Powell, J.L. (1995). Nonlinear Errors in Variables Estimation of some Engel Curves, *Journal of Econometrics*, 65, 205-233.

[34] Hausman, J.A. and Taylor, W.E. (1981). Panel Data and Unobservable Individual Effects, *Econometrica*, 49, 1377-1398.

[35] Hoch, I. (1962). Estimation of Production Function Parameters combining Time Series and Cross Section Data, *Econometrica*, 30, 34-54.

[36] Hughes, G. and McCormick, B. (2000). Housing Policy and Labour Market Performance, report for DETR, unpublished, University of Southampton.

[37] Jackman, R.and Savouri, S. (1992). Regional Migration in Britain: an Analysis of Gross Flows using NHS Central Register Data, *Economic Journal*, 102, 1433-1450.

[38] Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions*, New York: John Wiley and Sons.

[39] Judge, G. G. and Bock, M.E. (1978). *The Statistical Implications of Pre-testing and Stein-Rule Estimators in Econometrics*, Amsterdam, North-Holland.

[40] Kakwani, N.C. (1967). The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators, *Journal of the American Statistical Association*, 82, 141-2.

[41] Kuh, E. (1959). The Validity of Cross Sectionally Estimated Behavior Equations in Time Series Applications, *Econometrica*, 27, 197-214.

[42] Lee, M.J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Models*, Springer-Verlag Telos.

[43] Matyas, L. and Sevestre, P. (1996). *The Econometrics of Panel Data: A Handbook of Theory and Applications*, Kluwer Academic Publishers.

[44] Maddala, G.S. (1971). The Use of Variance Components Models in Pooling Cross Section and Time Series Data, *Econometrica*, 39, 341-358.

[45] Manning, A. (2003). The Real Thin Theory: Monopsony in Modern Labour Markets, *Labour Economics*, forthcoming.

[46] McCormick, B. and Sheppard, S. (1992). A Model of Regional Contraction and Unemployment, *Economic Journal*, 102, 366-377.

[47] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, New York: John Wiley and Sons.

[48] Nerlove, M. (1971). A Note on Error Components Models, *Econometrica*, 39, 383-396.

[49] O'Brien, R. and Patacchini, E. (2003). Testing the Exogeneity Assumption in Panel Data Models with "Non Classical" Disturbances, Discussion Papers in Economics and Econometrics n.0302, University of Southampton.

[50] Ortega, J. (2000). Pareto-improving Immigration in an Economy with Equilibrium Unemployment, *Economic Journal*, 110, 92-112.

[51] Patacchini, E. and Zenou, Y. (2003). Search Intensity, Cost of living and Local Labor Markets in Britain, CEPR Discussion Paper n. 3722, London School of Economics, and IZA Discussion Paper, n. 772.

[52] Petrongolo, B. and Pissarides, C. (2001). Looking into the Black Box: A Survey of the Matching Function, *Journal of Economic Literature,* 39, 390-431.

[53] Pissarides, C. (2000). *Equilibrium Unemployment Theory*, 2nd ed., MIT Press, Cambridge Massachusetts

[54] Rouwendal, J. (1998). Search Theory, Spatial Labor Markets, and Commuting, *Journal of Urban Economics*, 43, 1-22.

[55] Sargan, J. D. (1958). The Estimation of Economic Relationships using Instrumental Variables, *Econometrica*, 26, 393-415.

[56] Sato, Y. (2001). Labor Heterogeneity in an Urban Labor Market, *Journal of Urban Economics*, 50, 313-337.

[57] Seater, J. (1979). Job Search and Vacancy Contacts, *American Economic Review*, 69, 411-419.

[58] Simpson, W. (1992). *Urban Structure and the labor Market. Worker Mobility, Commuting, and Underemployment in Cities*, Oxford: Clarendon Press.

[59] Smith, T.E. and Zenou,Y. (2003). Spatial Mismatch, Search Effort and Urban Spatial Structure, *Journal of Urban Economics*, forthcoming.

[60] Stigler, G.J. (1961). The Economics of Information, *Journal of Political Economy*, 70, 94-104.

[61] Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment, *Review of Economic Studies*, 68, 261-295.

[62] Upton, G.J. and Fingleton, B. (1985). *Spatial Data Analysis by Example*, New York: Wiley.

[63] Vuong, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, 307-333.

[64] Wasmer, E. and Zenou, Y. (2002). Does City Structure Affect Job Search and Welfare?, *Journal of Urban Economics*, 51, 515-541.

[65] White, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press, New York.

[66] Windmeijer, F. (2000). A finite sample correction for the variance of linear two-step GMM estimators, IFS Working Paper n. W00/19.

[67] Wu, D. (1973). Alternatives Tests of Independence between Stochastic Regressors and Disturbances, *Econometrica*, 41, 733-750.