Three Essays on the Panel Data Approach to an Analysis of Economics and Financial Data

Laura Serlenga

Doctor of Philosophy

University of Edinburgh

2004

Abstract

This thesis provides an extension of panel data models on the analysis of Economics and Finance Data, discusses methods of estimation and evaluation for such models and presents empirical applications. The thesis consists in three essays. The first essay proposes three alternative approaches to test the Permanent Income Hypothesis (PIH) in the context of dynamic panels: the aggregate consumption approach, the Euler equation approach and finally Friedman (1957)'s original characteristic tests. The empirical evidence, using the British Household Panel Survey data, strongly supports the PIH. The analysis presented can be considered as supporting the view that empirical tests of PIH, based on aggregate time-series data, might suffer from misspecification or overlook some fundamental characteristics of micro data. The second essay addresses the issue of testing for factor price misspecification via a panel data approach. A theoretically coherent framework based on panel data techniques has been constructed. This allows for both the homogeneous and heterogeneous parameters when testing for anomalies in factor pricing models. The tests presented have a clear advantage over the traditional two-pass based tests because they do not suffer from errors in variable problem and have all the usual desirable asymptotic properties associated with the maximum likelihood approach. The empirical illustration shows that book to market equity and market value firm specific characteristics help explain asset returns in the UK over 1968-2002 even when all three of Fama and French's factors are present. This finding, which is in contrast to much of the literature, may be due to the improved efficiency of our estimates and power of our tests relative to those based on the two-pass method predominant in the existing literature. Lastly the third essay presents an application of Hausman-Taylor (HT) estimation in heterogeneous panels with time-specific common factors to gravity models of intra-EU trade. As an extension to the HT procedure which includes time-specific common factors and their heterogeneous individual parameters is presented. The underlying econometric techniques are developed and an alternative source of instruments is suggested. This methodology is applied to gravity models for international flows of trade using data on fifteen European countries over 42 years (1960-2001). A complete analysis of the sources of bilateral trade amongst European countries is presented using three different specifications. The empirical evidence confirms the effectiveness of the gravity model in explaining international trade flows. Results also encourage the use of our extended approach as a valid alternative to the basic time dummy specification.

Acknowledgments

I am deeply indebted and grateful to my supervisor Yongcheol Shin for his continuous and precious guidance throughout the last four years. His insights and suggestions have made an immense contribution to the development of my research ideas. His constant encouragement and enthusiasm have sustained my motivation and interest in the subject.

I am also grateful to Andy Snell for his advice and valuable guidance, to Gary Koop and Mark Steel for their helpful suggestions, to Ahmed Anwar and Fabriza Lapecorella for their continuous encouragement and support.

Finally I would like to thank my parents and my brother for always being there, my family for the continuous support, my friends and colleagues who have helped, cheered me and made my staying in Edinburgh so special. Any errors remaining in this dissertation are mine.

Declaration

In accordance with the regulations of the University of Edinburgh, it is noted that this dissertation is the result of my own work. I also declare that I have made substantial contribution in the work presented in Chapters Three and Four, which have been produced jointly with Andy Snell and Yongcheol Shin. Particularly, I have dealt with every aspect of the empirical applications and actively contributed to the discussions concerning with the econometric theoretical specifications.

Chapter 1

Introduction

Better modelling of interrelationships among economic and social processes is an important part of widening our understanding about them. Very large panels of observations on many groups of countries, firms and households over a long span of time are now commonplace. In this regard, more and more applied economists now turn to the panel data approach to enhance their understanding of the driving force behind fundamental economic relationships. What is generally referred to as panel data approach to economic research provides several major advantages over conventional cross-sectional or time-series approaches. This clearly facilitates estimation of and inference on more realistic behavioral models that could not be identified when we use only either a cross section or a single time-series approach. Recently, the main focus has been how best to model both cross sectional heterogeneity and time series dynamics, simultaneously. Another important modeling issue is how to allow for cross section dependency. Following some of recent econometric developments, this thesis aims to demonstrate the usefulness of those panel-based methodologies by analyzing three different applications in empirical finance, international trade and macroeconomics.

The conventional panel data approach allows us to control for unobservable individual heterogeneity across cross section units. Error components model is routinely used to control for these individual differences, and such a model typically assumes that stochastic error term has two components: time-invariant unobservable individual effects and the idiosyncratic random disturbances. Most of the earlier studies have primarily focused on how best model unobservable individual effects usually with short time periods and large cross section units. The conventional estimation methodologies in this vein are: the fixed effect model (FEM) and the random effect model (REM).

However, another major advantage of the panel data approach is its ability to model dynamics at an individual level as well. This type of model is called dynamic panels with short time periods, where lagged dependent variables are included as regressors. A sizeable part of the recent literature has focused on consistent estimation of the regression in dynamic panels along with unobserved individual effects. Nickel (1981) shows that the FEM leads to biased estimates in dynamic panels when the number of time periods is short irrespective of the size of cross section units. As a response to this inconsistency problem, a number of studies have recently developed alternative consistent generalised method of moments (GMM) estimation methods, which provides consistent estimation of dynamic panels, even when the number of time periods is fixed, see *e.g.* Arellano and Bond (1991), Ahn and Schmidt (1995) and Arellano and Bover (1995). However, there has been a growing concern about the generally poor performance of the GMM estimator in dynamic panels: namely, since the number of instruments increases with the time dimension, the model generates too many overidentifing restrictions even for moderate values of time periods and the quality of these instruments is often poor. In particular, Alvarez and Arellano (1998) and Bond and Blundell (1998) show that the effect of weak instruments is likely to bias the distribution of the GMM estimator towards the FEM estimates, especially when the underlying series are highly autocorrelated. It is also well-established that the OLS estimate is biased upwards in the presence of individual-specific effects [Hsiao (1986)], and that the FEM estimate will be seriously biased downwards [Nickell (1981)] in dynamic panels with short time periods. Therefore, the GMM estimate is also likely to be biased downwards [Bond, Hoeffler, and Temple (2001)]. In particular, Bond and Blundell (1998) propose the so-called "system" GMM estimator as a variant of the standard Arellano and Bond's (1991) GMM procedure. Relying on relatively mild restrictions on the initial conditions, the system GMM estimator adds further moment conditions which are derived from the model in levels and improves the performance of the GMM estimator in the dynamic panels.

The increasing availability of the macro-type panel data with the numbers of cross section units and time periods being both large also raises a number of empirically and theoretically interesting issues. First, since most economic and finance time-series data tend to be nonstationary, it would be equally important to investigate the nature of their stationary and cointegrating properties in the context of panels. In this regard, extending the time series estimation and testing procedures for integrated and cointegrated series to panels has been a natural development, see for example Shin and Snell (2001) and Im, Pesaran and Shin (2003) among the others. Next, as the increasing number of time periods increases, it would be worthwhile to consider heterogeneous panels both in static and dynamic models, where slope parameters are also allowed to differ over cross sectional units. Notice that in the conventional panel data analysis the attention focuses only on allowing for intercept variation via individual effects. In practice the extent of cross sectional heterogeneity may be so large as to preclude the use of pooling while in the case where only relatively small time periods were available, the scope for analyzing the slope heterogeneity explicitly appears limited, *e.g.* Balestra and Nelrove (1966). In particular, Pesaran and Smith (1995) show that neglecting parameter heterogeneity leads to misleading inferences about parameters estimates for large T and N. An approach that has become increasingly popular in the estimation of heterogenous panel data model is to focus estimation and inference on so called mean group quantities that are "averages" across panel units [Pesaran and Smith (1995)]. Pesaran, Shin and Smith (1999) propose the Pooled Mean Group (PMG) estimator which is considered as an intermediate estimator between traditional pooled estimators, such as fixed and random effects estimators, and the MG estimator since it involves both pooling and averaging.

Recently, particular attention has also been drawn on common factor models in the framework of panels with large T and N. While the conventional approach deals with common time specific heterogeneity simply by adding fixed time dummies, a number of recent studies attempt to model unobserved common factors explicitly, *e.g.* Ahn, Lee and Schmidt (2001), Bai and Ng (2001), Pesaran (2002) and Phillips and Sul (2002). These approaches also allow us to deal with certain degrees of cross section dependence via common time-specific factors with their individual responses being heterogeneous, in which case the conventional uncorrected estimates will be potentially biased. For instance Pesaran (2002) proposes a simple consistent estimation procedure called "Correlated Common Effect" (CCE) estimator that can be estimated by the OLS being applied to an auxiliary regression where the observed regressors are augmented by cross sectional averages of dependent variable and individual time varying regressors.

Another issue of relevance concerns consistent estimation of coefficients on individual-specific, time-invariant variables. The fixed effects estimation does not allow for estimating coefficients on time invariant variables since the within transformation wipes out them along with unobserved individual effects. Considering that the assumption of no correlation between unobserved individual effects and the regressors have been convincingly rejected in almost all empirical applications, both the Pooled OLS and the REM cannot be used. In this situation we need to use the Hausman and Taylor (1981, hereafter HT) Instrumental Variable estimator. The main advantage of this methodology is that it is able to obtain instruments internally. One of the main aims of this thesis is to develop the generalized HT estimation methodology in panels with observed and/or unobserved common time-specific factors, where individual responses to those common factors are heterogeneous across cross section units.

In the Second Chapter we apply the GMM methodology in order to re-investigate the validity of the Permanent Income Hypothesis (PIH) using micro panel data obtained from the British Household Panel Survey (BHPS). Most empirical studies based on representative-consumer's aggregate consumption function, often rejects the PIH, *e.g.* Flavin (1981), Deaton and Campbell (1989) and Galì (1991). However, it is not clear whether the statistical rejections of the theory imply a robust rejection of the theory or whether they are driven by many simplifying assumptions. In fact when the PIH is tested using micro data, empirical results often support the Hypothesis, *e.g.* Runkle (1991), Attanasio and Weber (1993) and DeJuan and Seater (1999). This suggests that analyses conducted with aggregate data might suffer from mis-specifications and vitiate the results. We aim to shed more light on the issue of empirical tests of the PIH by proposing three alternative testing procedures using a micro panel data. The first of these tests is based on Flavin's (1981) specification that has been mainly used to test aggregate implications of the PIH against the hypothesis of excess sensitivity of consumption. The second tests based on the Euler equation have been mainly conducted with micro data, and the main advantage is that the Euler equation is a coherent and flexible optimization model that includes a variety of factors and incorporates as much information about individual behavior as is available. Finally, we propose to use characteristic tests advanced originally by Friedman (1957). The characteristic tests will turn out to be a very appealing alternative testing procedure because there is no need to construct a series for permanent income or any form of expectation. Those tests have been rarely used in the literature, but mainly conducted using either time series or cross sectional, see Friedman (1957) and DeJuan and Seater (1999). We estimate those models by Pooled OLS, the FEM and GMM. In particular, Flavin's model has a dynamic specification therefore consistent estimates are provided only by the system GMM. The Euler equation specification is consistently estimated by GMM due to nonzero correlations between some of explanatory variables and error terms and due to measurement errors. Finally, characteristic tests are mainly conducted using the Pooled OLS estimation. Our estimation and test results using micro panel data unanimously provide evidence in favor of the PIH. Those findings are in conformity with the studies that use micro data to test the PIH, *e.g.* Attanasio (1998). Our analysis therefore confirms the previous evidence in favor of the PIH that an analysis based on aggregate data might suffer from mis-specifications or overlook some fundamental characteristics of micro data and therefore vitiate the results that lead to rejection of the PIH, see *e.g.* Seater (1998).

The main aim of the Third Chapter is to reexamine the testing of anomalies on factor pricing models. The central prediction of the asset pricing models is that the market portfolio of invested wealth is mean-variance efficient, see Sharpe (1964) and Black (1972). But, there have been several empirical findings which contradict the prediction of these models. The most prominent is the size effect of Banz (1981), who finds that the market value of equity adds to the explanation of the cross-section of average returns provided by market betas. More recently, there has been a large anomaly literature where firm specific characteristics such as leverage, past returns, dividend-yield, earnings-to-price ratios and book-to-market ratios as well as size help explain cross sectional returns [*e.g.* Fama and French (1992)]. These anomalies have been attributed to market inefficiency but could be the result of a misspecification of the underlying factor pricing model. The most popular approach to detect these anomaly effects has been the two pass (TP) cross-sectional regression methods, advanced by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973). In the first stage, the asset betas are estimated by time series linear regression of the asset's return on a set of common factors. Then, the cross sectional regression of mean returns on betas and characteristics is estimated, and the significance of asset specific regressors is evaluated along with factor risk premia. However, it is well-established that the TP method suffers from the errors in variables problem, because estimated betas are used in place of true betas in the second stage cross sectional regression. We propose a panel data approach to test for factor price misspecification using partially heterogenous panels. It is a salient fact that the benefits of using panel data techniques have been completely ignored. If our interest lies solely in testing the significance of these characteristics, we can show how to construct a panel data regression model with one set of variables varying over time such as common factors and another set of variables varying both over time and over asset portfolios. Our investigation provides a theoretically coherent example to which panel data techniques dealing with both homogeneous and heterogeneous parameters can be applied. Our suggested panel-based anomaly tests have one clear advantage over TP-based tests: they are based on full information maximum likelihood estimates so that they do not suffer from the errors in variables problem and have all the usual asymptotic properties associated with likelihood tests. We apply these approaches to a large data set of UK stock returns between 1968 and 2002. The empirical results from the TP and the panel data regressions show the importance of book to market equity and market value in helping explain asset returns. When such terms are added to the simple CAPM version of the model their significance is enormous. This confirms results from similar studies done on both US and UK data. Moreover, the three factor model is still mis-specified although, in terms of fits, it is an improvement over the single factor model. Perhaps even more important however are our findings from the panel data analysis. Here, contrary to the results of Fama and French (1996), we find that (i) adding size and book-to-market macro factors does not drive out the significance of a standard CAPM market factor, (ii) a firm specific book-to-market variable remains significant even after the basic CAPM factor is augmented by Fama French SMB and HML factors and (iii) a firm specific size variate remains likewise significant but generally only in subsamples drawn from the 1980's. We tentatively argue that the first and the second finding could be a result of the greater efficiency of our estimates and power of our ML testing procedure. We argue that the third finding supports Berk' s (1995) argument that the significance of firm size may be due to spurious coefficient bias rather than the existence of an asset pricing anomaly.

In the Fourth Chapter we analyse gravity models of international in panels. The gravity model of international trade has been successfully estimated with a number of panel estimation techniques such as the pooled OLS, the Fixed Effects Model, the Random Effects Model. In our study we focus on the FEM and HT estimation for two main reasons. First, the REM assumption that there is no correlation between explanatory variables and unobserved individual effects is convincingly rejected in all cases considered. Second, one of our main interests lies in consistently estimating the coefficient on time invariant variables. We also explicitly allow for the common time specific effects in order to capture business cycle effects or to deal with the generic globalization issues. While the conventional approach extends the model simply by incorporating the fixed $T-1$ time dummies in the regressions, in our study we follow recent developments of panel studies surrounding the common time effects, *e.g.* Pesaran (2002), and advance an alternative estimation framework in which we explicitly allow for the existence of observed and/or unobserved common time-specific factors and individual responses to those common factors are heterogeneous across country pairs. We also provide an underlying econometric theory for this extend HT methodology and propose a new source of instruments. We apply our proposed (extended) HT estimation technique along with the conventional approaches to a comprehensive analysis of the sources of bilateral trade amongst the 15 European countries over 1960-2001. In selecting the basic empirical specification we first consider the impacts of core explanatory variables: measures of economic size of trading partners such as GDP and population, and the distance. We then augment the basic specification by adding various variables such as common language, common border, free trade area and currency union membership dummies. Finally, we follow recent theoretical developments [*e.g.* Helpman (1987) and Egger (2002)] and include variables measuring both similarity in relative size of trading countries and differences in relative factor endowments. Our empirical findings clearly suggests the potential advantage of our proposed approach over the conventional one based on the fixed time dummies. Furthermore, comparing the estimation results for the benchmark case without allowing for time effects and our proposed model with unobserved common time factors, we may conclude that the estimation results obtained using our proposed extended model seems to be more sensible. This may reflect that it is also important to allow for a certain degree of cross section dependence via unobserved common time factors, otherwise the resulting estimates would be severely biased.

Finally, Chapter Five provides a summary of the results obtained in the thesis and discusses new directions of research.

Chapter 2

Three Alternative Approaches to Testing the Permanent Income Hypothesis in Panels

2.1 Introduction

The Permanent Income Hypothesis (PIH) focuses on the behavior of a representative agent with an infinite time horizon. According to this theory, consumers plan their expenditures on the basis of their lifetime income expectations rather than on the basis of income received period by period. Some recent attempts to test the PIH have used representative-consumer models on aggregate data, *e.g.* Hall (1978), Flavin (1981), Mankiw and Shapiro (1985), West (1988), Deaton and Campbell (1989), Campbell and Mankiw (1990) and Galì (1991). This type of model does not seem to fit the data very well and often rejects the PIH. However, it is not clear whether the statistical rejections of the theory imply a robust rejection of the theory or whether they are driven by many simplifying assumptions. Cochrane (1989) raises a very interesting question: "Suppose a consumer sets consumption equal to income each period, rather than follow the optimal permanent income decision rule. How much does she lose?" He calculates the utility cost to consumers under the following alternative decision rules such as intertemporal maximization (Hall, 1981), excess of sensitivity of consumption (Flavin, 1981), and excess smoothness of consumption (Mankiw and Shapiro, 1985 and Campbell and Deaton, 1987). The losses are small and about 0.01 percent of wealth. The utility costs are small because cyclical changes in consumption are small and because the utility costs of deviations from an optimum are an order of magnitude smaller than the deviation itself. In this context, suboptimal decision rules that cost a trivial amount of utility or profit are called near-rational. Near-rational behavior can be easily interpreted as a small mistake: people do not literally maximize and they follow a heuristic decision process that economists model by maximization. Their actual decision may deviate from the optimal decision rules if the utility costs of doing so are trivial. In a second interpretation, Cochrane asserts that the small mistakes are made by the economists in modelling the world rather than by the agents they study. Empirically useful forms of economic theory gloss over many complexities of the decision problems that consumers actually face such as transaction cost, information acquisition and decision-making. We cannot know precisely what effect including these small corrections would have on predictions of the theory until we work out a theory that includes them, which seems a hopeless task. A statistical rejection of the PIH might therefore be driven by modeling simplifications rather than by a failure of the basic theory: the tests are able to statistically distinguish alternatives that are not well-distinguished economically. The alternative behaviors that might cause the tests to reject can be generated by small costs of information acquisition or processing, transactions and so on. The main point is that finding those alternatives is not strong evidence against the basic theory that consumers intertemporally optimize but this also implies that the theory, as it stands, provides few predictions about aggregate consumption that are robust to one dollar mistakes or mis-specifications.

Cochrane (1989) confirms the theoretical validity of the Permanent Income Hypothesis in explaining intertemporal choice of consumption but also calls for testing the PIH more accurately. To a certain extent some of recent empirical findings validate this assessment. Indeed, when the PIH is tested on an micro data, empirical results generally provide evidence in favor of the PIH and the general conclusion is that the PIH is rejected in the aggregate probably because of problems of aggregation bias and insufficient allowance made for the dependence of consumption on individual characteristics, see *e.g* Zeldes (1989), Runkle (1991), Attanasio and Weber (1993) and DeJuan and Seater (1999).

In this analysis we aim to shed more light on the issue of empirical tests of the PIH by proposing three alternative testing procedures using a micro panel data. The use of panel data is expected to increase the efficiency of econometric estimates and allows for heterogeneity among households to be modelled, see also Runkle (1991), Attanasio and Weber (1993) and DeJuan and Seater (1999). The first of these tests is based on Flavin's (1981) specification that has been mainly used to test aggregate implications of the PIH against the hypothesis of excess sensitivity of consumption. The second tests based on the Euler equation have been mainly conducted with micro data and consist in verifying directly whether the first order condition is continually satisfied. The main advantage of this approach is that the Euler equation is a coherent and flexible optimization model that includes a variety of factors and incorporates as much information about individual behavior as is available. Finally, we propose some of the characteristic tests proposed originally by Friedman (1957) that focus on the key predictions of the PIH. In particular, the characteristic tests will turn out to be a very appealing alternative testing procedure because neither a consumption function (an Euler equation) specification nor assumptions on the time series properties of income are necessary and there is no need to construct a series for permanent income or any form of expectation.

As emphasized by Attanasio (1998), consumption cannot be studied in isolation: consumption and saving choices are determined together with a number of other choices, ranging from labor supply to household formation and fertility decision to planned bequest. In this regard, we therefore use the British Household Panel Survey (BHPS) data which is a microeconomic survey data that provides information on 8,167 individuals from 1991 to 1999. The BHPS contains data on household consumption of non-durable goods, income and several socio-economic household and individual characteristics. We extract a balanced panel with information on 2,976 households for seven consecutive years (1991-1997) and estimate by Pooled OLS, the Fixed Effect Model, Arellano and Bond' s (1991) GMM and System GMM [Blundell and Bond (1998), and Blundell, Bond and Windmeijer (2000)]. Under the aggregate consumption approach the results of the GMM estimates all support the PIH. Also when the Euler equation is consistently estimated empirical results are in favor of the PIH. Moreover, the results of the characteristic tests generally support the Hypothesis. We may therefore conclude that our empirical evidence strongly supports the PIH, when models are consistently estimated empirical results unanimously provide evidence in favor of the PIH. Those findings are in conformity with the recent studies that use micro data to test the PIH and sharply contrast the results of the analyses conducted on aggregate data [see Flavin (1981), Deaton (1982), West (1988), Deaton and Campbell (1989), Attanasio and Weber (1993), Attanasio and Browing (1995), Attanasio (1998)]. The most relevant result is that testing a model suitable for aggregate consumption with panel data provides evidence in favor of the PIH. Our analysis can be considered as a piece of evidence in favor of the thesis that empirical tests of the PIH, based on aggregate data, might suffer from mis-specifications or overlook some fundamental characteristics of micro data and therefore vitiate the results that lead to rejection of the PIH [see Attanasio and Weber (1993), Attanasio and Browing (1995) and Seater (1998)].

This work is organized as follows. In Section 2, we propose an overview of the theory and look at the most representative models of intertemporal choice of consumption. In Section 3, we outline the underlying panel data econometric methodology in details. Section 4 presents our empirical results and Section 5 concludes.

2.2 Overview on Testing the Permanent Income Hypothesis

2.2.1 Testing the PIH with aggregate consumption function

Early empirical studies using aggregate consumption expenditure are mainly inspired by Keynes (1936). In the Absolute Income Hypothesis (AIH) Keynes stresses the dominant role of real disposal income, y, in determining current real consumption, c. He suggests the consumption function should be approximated to a linear relationship

$$c = a + by, \tag{2.1}$$

where a is autonomous consumption and b is the marginal propensity to consume (MPC). In short, the main implications of the AIH model are the following: (i) autonomous consumption is greater than zero, (ii) MPC is less or close to unity, (iii) MPC is less than average propensity to consume (APC), (iv) the average propensity to save

(APS) increases as income rises. Finally, the AIH also predicts that when government spending falls, as happened after WWII, the economy would move towards recession, and consumption would decrease. Kuznets (1946) uses time series data dating back to the Civil War to test the AIH. He finds that the MPC is less than APC in budget data and short-run time series data but is equal to APC in the long run. Moreover, APS and APC did not rise secularly, whereas private demand increased sharply and APS was lower than during the interwar period. These results among others motivated various economists to find a plausible alternative relationship with consistent short-run and long-run implications [see Duesenberry (1949), and Brown (1952), Friedman (1957), Ando and Modigliani (1963), Johnson (1971)].

The Permanent Income Hypothesis (PIH) by Friedman (1957) and the Life Cycle Hypothesis (LCH) by Ando and Modigliani (1963) are the most remarkable examples in this innovative field of research. Both approaches adopt a precise microeconomic framework to analyze the optimal behavior of a forward looking rational consumer. The main difference between the LCH and the PIH lies in the time horizon considered. The PIH focuses on the behavior of a representative agent with infinite life. The LCH refers to the aggregation of finitely-lived overlapping generations and introduces different behavior of consumers with respect to their age. Common assumptions of the models are that, at any time t, the representative consumer has full information about future real disposal labor income y and can issue or redeem a risk-free bond at a constant after tax real rate r against future income.

According to the PIH, income is defined as the amount that a consumer can consume while wealth remains unchanged [see Friedman and Kuznets (1945)]:

$$y^p = rW, \tag{2.2}$$

where y^p is a permanent income and W is a wealth defined as discounted income receipts.

Under the LCH, in any period t, the total income of one person of age T will be proportional to the present value of the total resources accruing to her for the rest of her life

$$c_t^T = \omega_t^T v_t^T, \tag{2.3}$$

here ω is a proportionality factor which depends on the form of the utility function, on the rate of return on assets, on the age of the person, but not on the total resources v.

However, the theoretical definitions (2.2) and (2.3) themselves are not testable. In the time series studies, Friedman (1957) provides a formal representation arguing that a weighted average of past and current income is a plausible estimate of the permanent income (so called "income approach"). The PIH is defined as a relation between consumption and expected income, emphasizing the dynamic nature of the consumption-income relationship. The permanent income denoted y_p, can be expressed as:

$$y_p = \int_{-\infty}^{T} w(t-T)\, y_m(t)\, dt, \tag{2.4}$$

where y_m is measured income and $w(t-T)$ is a weighting function such that:

$$\int_{-\infty}^{T} w(t-T) = 1. \tag{2.5}$$

Friedman (1957) defines the weighting pattern as an exponential one:

$$w(t-T) = \beta e^{\beta(t-T)}, \tag{2.6}$$

where β is the subjective discount rate. This form makes the weighting pattern equivalent to the form of adaptive expectations that Cagan (1955) uses to estimate the expected rate of price changes in the hyperinflation era. The adaptive expectation hypothesis states that the consumer learns from her past and suggests that expected income is a proxy for permanent income in the explanation of current consumption.

On the other hand, Modigliani and Ando (1963) introduce a number of rather drastic and simplifying assumptions on the life pattern of earnings in order to provide an empirical specification for the LCH. They also notice that under adaptive expectations and assuming that aggregate income follows an exponential growth process, the distinction between PIH and LCH blurs. As nicely summarized by Ferber (1973) the resulting empirical relationships, estimated using data sets from the 1950s and 1960s, are structurally stable and have a successful forecasting record. However, their performance gradually deteriorates as the economic disturbances in the early 1970s begin to reflect themselves in the corresponding data. This experience, together with other advances in theoretical and econometric modelling, had a considerable impact on the work that followed. A theory of consumption based on the rational expectation hypothesis (REH) was firstly explored by Lucas (1976) and then formalized by Hall (1978). The empirical models that test the PIH using a proxy for expected income under REH are often called "life cycle-permanent income models" in order to emphasize that in this case no distinction can be made between the two models in their aggregate implications.

Under the REH current consumption depends on permanent income

$$c_t = y_t^p \equiv \frac{r}{1+r}\left[A_t + \sum_{i=0}^{\infty}(1+r)^{-i} E_t y_{t+i}\right], \tag{2.7}$$

where y_t^p is permanent income, r is the (constant) return on nonhuman wealth, A is nonhuman wealth, and y is labor income. E_t is the expectations operator conditional on all the information available to the representative consumer at time t. The evolution of assets over time is governed by

$$A_{t+1} = (1+r)\, A_t + y_t - c_t. \tag{2.8}$$

The first difference of equation (2.7) can be written as

$$\Delta c_t = r\sum_{i=1}^{\infty}(1+r)^{-i}\left(E_{t+1} - E_t\right) y_{t+i}, \tag{2.9}$$

so that changes in consumption are driven by innovation in labor income. More precisely, in this infinite horizon model, the change in consumption is simply the annuity value of the present discounted value of the change in the expected value of future labor incomes.

Several studies have followed the REH approach to test PIH on aggregate data. Hall (1978) analyzes the impact of uncertainty in the intertemporal choice of consumption. Consumers maximize expected utility under uncertainty keeping the expected discounted marginal utility of consumption constant. The stochastic implication of Hall's model is that when a consumer maximizes expected future utility, her conditional expectation of future utility is a function of today's level of consumption alone and all other information is irrelevant. In other words, apart from trend, marginal utility follows a random walk. Moreover, if the marginal utility is a linear function of consumption then the consumption is also a random walk apart from trend. Previous research on consumption has suggested that lagged income might be a good predictor of current consumption but this hypothesis is inconsistent with the intelligent, forward-looking behavior of consumers that forms the basis of the PIH.

Flavin (1981) tests the validity of the PIH in equation (2.9) estimating the following system of equations:

$$\Delta c_t = \gamma + \beta_0 \Delta y_t + \beta_1 \Delta y_{t-1} + \ldots + \beta_7 \Delta y_{t-7} + \theta \varepsilon_t + u_t, \tag{2.10}$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_8 y_{t-8} + \varepsilon_t, \tag{2.11}$$

where γ is the productivity growth term, β_i $(i = 1, .., 7)$ represent the excess sensitivity of changes in consumption with respect to changes in income, θ is the warranted change in consumption, and ε_t is the innovation in the income process. The term u_t represents the measurement error in consumption change together with the effects of the information that the consumer may have on permanent income that is not captured by the autoregressive specification of income in (2.11). If the PIH is valid, β_i, $i = 1, .., 7$, should be zero. If β_i are non-zero, say positive, then according to the excess sensitivity hypothesis, consumption responds even to predictable changes in income. Flavin uses US quarterly data from 1949 to 1979 to estimate the reduced form of system (2.10)-(2.11) after detrending the variables. More specifically, Flavin (1981) suggests by fitting exponential time-trends to both consumption and income, and replacing consumption and income in the regressions by their residuals, and finds that some of the coefficients β_i are significantly positive, contradicting the PIH. Deaton (1992) raises a strong criticism of Flavin' s model, arguing that the econometric methodology applied by Flavin biased the results toward the rejection of the PIH. According to Deaton, the first difference of labor income is stationary, so labor income itself is difference stationary and the source of the problem is the presence of the integrated regressor in the reduced form of (2.11). Mankiw and Shapiro (1985) clarify this point. In their analysis, they assume that disposable personal income follows a non-stationary process and use quarterly time-series data from 1959 to 1983 to test the hypothesis that consumption is a function of income, consider various specifications for the income generating process and demonstrate that Flavin' s detrending procedure generates spurious findings of excess sensitivity.[1]

Further evidence against the PIH is given by West (1988) and Deaton and Campbell (1989). They provide an alternative explanation to the reason why consumption is smooth, which differs from the PIH. According to the PIH, consumption is smooth because permanent income is smoother than measured income. On the other hand, the "Deaton paradox" shows that permanent income is in fact less smooth than measured income. Deaton and Campbell (1989) use a VAR analysis and find a positive correlation between the change in consumption and the lagged change in income, a correlation that would be zero if the PIH were true. They conclude that the consumer is "excessively sensitive" to anticipated changes in income but is "excessively insensitive" to unanticipated changes in income. Consumption is slow to adjust to innovations in income, thus changes in consumption are related to averages of previous innovations and this explains both smoothness and correlation.

Galì (1991) tests PIH with US aggregate data developing a procedure based on a long-run restriction implied by the consumer's intertemporal budget constraint. The relevance of his approach is that it does not require any assumption on the stochastic proprieties of labor income. Starting from model (2.7), he defines the variability of consumption relative to the variability implied by the PIH model as

$$\psi = \left[\frac{Var\,(\Delta c)}{Var\,(\xi)}\right]^{\frac{1}{2}}, \tag{2.12}$$

where ξ is the innovation in permanent income defined by

$$\xi \equiv y^p - E_{t-1}y_t^p = r\sum_{i=1}^{\infty}(1+r)^{-i}\left(E_{t+1} - E_t\right)y_{t+i}. \tag{2.13}$$

The standard PIH model implies $\Delta c_t = \Delta y_t^p = \xi_t$ so that ψ is equal to one. Galì shows how the restriction implied by the budget constraint on the consumption time-series allows us to identify the variance ratio ψ and derive its consistent estimator.

[1]The detrending procedure proposed by Flavin eliminates the deterministic trend and not the stochastic trend such as a random walk plus drift process.

His empirical results support the finding of "excess of smoothness" providing empirical evidence in favor of the "Deaton paradox". On the other hand, Quah (1989) decomposes labor income into permanent and temporary components and shows that when agents distinguish permanent and transitory movements in their labor income, the PIH correctly predicts the observed smoothness in consumption. Although Quah (1989) resolves the "Deaton paradox" in favor of the PIH, we can conclude that the empirical results, using aggregate data, generally do not provide evidence in favor of the PIH.

2.2.2 Testing the PIH using an Euler equation with micro data

Attanasio and Weber (1993) and Attanasio and Browing (1995) among others emphasize the importance of testing PIH and LCH with micro data. Simple permanent income and life cycle models assume intertemporally additive preferences, perfect capital markets and rational expectations. As observed in the previous section, models based on the assumption of the existence of a representative consumer are often rejected when tested on aggregate time series. Attanasio and Browing (1995) claim that the main reasons for these rejections are aggregation bias and the insufficient allowance for the dependence of consumption on demographic characteristics. As a solution they suggest using micro data and to condition the model on household-specific factors that may affect consumption decisions. Consumption cannot be studied in isolation: consumption and saving choices are determined together with a number of other choices, ranging from labor supply to household formation and fertility decision to planned bequest (Attanasio, 1998). This leads to the necessity of a coherent and flexible optimization model that includes a variety of factors and incorporates as much information about individual behavior as is available. This new approach tests directly whether the first order condition (Euler equation) is continually satisfied. Without losing empirical tractability the Euler equation allows for lots of factors in the analysis, *i.e.* labor and demographic factors, and to study their effects on the marginal utility of consumption. An example of models based on the Euler equation is DeJuan and Seater (1999). First, consider the intertemporal choice of a consumer i who chooses the path of consumption in order to maximize her expected utility function:

$$E_{it} \sum_{t=t_0}^{T} V\left(C_{it}, H_{it}, t\right), \tag{2.14}$$

subject to the budget constraint

$$A_{it+1} = \left(1 + r_{it}\right) A_{it} + Y_{it} - C_{it}, \tag{2.15}$$

where C_{it} is consumer i's consumption in period t, H_{it} is the vector of household characteristics, E_{it} the expectations conditional on the information available at time t and $V(.)$ is the utility function. In budget constraint (2.15), r_{it} is the real after-tax interest rate, A_{it} is the household non-human wealth and Y_{it} is the real disposal income. Also, H_{it}, includes three components: those that cause transitory consumption denoted T_{it}, those that affect the household's intertemporal rate of substitution, R_i, and those that affect the household choice in other ways, X_{it}. Assuming no liquidity constraint and an isoelastic utility function, DeJuan and Seater formulate the following explicit specification of the Euler equation

$$\ln\left(\frac{C_{it+1}}{C_{it}}\right) = \beta_0 + \beta_1 r_{it+1} + \beta_2 \ln\left(\frac{F_{it+1}}{F_{it}}\right) + \beta_3 R_i + e_{it+1}, \tag{2.16}$$

where F_{it} represents a vector of household characteristics that change over time and across cross sectional units and e_{it+1} is a compound error term that includes a time invariant individual effect. In order to be able to test the validity of PIH against the validity of the AIH, DeJuan and Seater provide an alternative model. Following Campbell and Mankiw (1990), they assume that consumers can be divided in two groups: consumers in the first group receive share $(1-\lambda)$ of total disposal income and behave according to the PIH (2.16), in the second group they simply consume their current income (share λ of total disposal income, so called "rule-of-thumb" consumers) and estimate the following model

$$\ln\left(\frac{C_{it+1}}{C_{it}}\right) = B_0 + B_1 r_{it+1} + B_2 \ln\left(\frac{F_{it+1}}{F_{it}}\right) + B_3 R_i + B_4 \ln\left(\frac{Y_{it+1}}{Yit}\right) + e^{*}_{it+1} \tag{2.17}$$

where $B_j = (1 - \lambda)\beta_j$ for $j = 0, ...3$, and $e^*_{it+1} = (1 - \lambda) e_{it+1}$.[2] If the PIH is to be valid, B_4 should be zero. DeJuan and Seater estimate model (2.17) using a panel data set of CEX (Consumer Expenditure Survey) from 1986 to 1991, and their findings support the PIH.

The increasing availability of micro data also allows for testing of the presence of liquidity constraints. If agents are liquidity constrained they consume their entire disposable income and the consumption function will be an extreme Keynesian one. Zeldes (1989), Runkle (1991) and Mariger and Shaw (1993) test for the validity of PIH against the alternative of prevalent liquidity constraints. Runkle (1991) uses data from the Panel Study Income Dynamic (PSID), from 1968 to 1982, and tests for the validity of the PIH using an Euler equation specification very similar to (2.17). In order to test for liquidity constraint he splits the sample following two criteria: whether a household owns or rents its residence and whether the annualized value of the household's asset income is greater or less than two months income. Assuming that homeowners and people with liquid wealth probably would not be liquidity constrained, past income should not have much power in predicting their consumption growth. The income variable never appears to be statistically significant in explaining consumption growth. Thus, Runkle' s empirical results strongly support the PIH.

Summarizing, we can conclude that empirical tests of PIH with panel data provide general support for the theory. However, allowing for portfolio allocation opens other avenues of further research. A critique to the Euler equation approach comes from Miller and Sieg (1997). They notice that the PIH does not typically impose sufficiently strong identification conditions on the budget constraint necessary to achieve consistent estimations with panel data. In particular, they notice that the rationality of the economic agents is assumed irrespective of markets although it is very difficult to characterize equilibrium allocations when markets are incomplete. In order to overcome the under-identification problem, Miller and Sieg (1997) propose a model that imposes more structure on the markets. Their specification is based on assumptions of complete and competitive markets (CCM) and incorporates uncertainty in a sufficiently simple way to yield a tractable econometric model.

[2] B_4 is not a fraction λ of the rule-of-thumb consumer, but equation can be viewed as a log-linear approximation of the true model, see Campbell and Mankiw (1990) and DeJuan and Seater (1999).

2.2.3 Testing the PIH using characteristic tests

Friedman (1957) states a formal complete model for PIH as follows:

$$c^p = k\,(i, w, u)\, y^p, \tag{2.18}$$

$$y = y^p + y^t, \tag{2.19}$$

$$c = c^p + c^t, \tag{2.20}$$

$$\rho_{y^p y^t} = \rho_{c^p c^t} = \rho_{y^t c^t} = 0, \tag{2.21}$$

$$\mu_{y^t} = \mu_{c^t} = 0, \tag{2.22}$$

where y represents current income, and y^p and y^t the permanent and transitory component of current income. Also, c represents current consumption, c^p its permanent component and c^t the transitory one. The ratio of non-human wealth to income is given by w and u is a variable which determines the consumer tastes and preferences versus additions to wealth. (2.18)-(2.20) represent the theoretical model whereas (2.21)-(2.22) are needed to make the theoretical model operational and empirically testable through characteristics tests. In (2.21) ρ is the correlation coefficient between the variables designed by its subscript. Zero correlations of the first two relations imply that the average transitory component is the same for all the values of permanent income, whereas the absolute value is directly proportional to the permanent component. The third relation indicates that consumption is determined by rather long-term considerations, *i.e.* any transitory change in income leads primarily to addition to assets or to the use of previously accumulated balances rather than to corresponding changes in consumption. The latter is a crucial postulate because empirical findings do not always support this assumption. Finally, in (2.22) μ is the mean of the variable designed by its subscript. The mean of the transitory components of consumption and income are equal to zero. This condition turns out to be plausible as long as the probability distribution in question is sufficiently comprehensive.[3]

The characteristic tests are based on a complete model, (2.18)-(2.22) and test the

[3] This assumption denies systematic shocks at time t within a characteristic group of individuals.

key proprieties of PIH, such as proportional hypothesis, difference in income elasticities between permanent and transitory components, zero income elasticity of consumption for transitory income against the alternative of validity of AIH. Unlike most of the empirical applications of the PIH those tests are performed with neither a consumption function nor an Euler equation, no assumptions are necessary on the time series proprieties for income and there is no need to construct a series for permanent income or any form of expectation.

Since permanent and transitory components of income and consumption are not measurable the characteristic tests use qualitative external information in order to proxy transitory and permanent components by qualitative instrumental variables. Friedman performs sixteen characteristic tests using both time series and cross-sectional data. The sample of households is divided according to criteria that identify whatever aspect of permanent income is relevant to the test in question. For instance, if we assume *a priori* that education is positively correlated with the level of permanent income, classifying individuals by level of education is a way to classify them by permanent income.

The characteristic test refers to a type of test that targets a specific aspect of the empirical model. In other words a characteristic test is to match the characteristics of the data with those of the empirical model. Although the conventional view of testing in economics is about rejecting candidate hypotheses, there are other categories of testing in Economics. Confirmationalist tests secure a basis for belief, look for satisfactoriness of empirical models and confirm the characteristics of empirical models, using Econometrics as a measuring device to reassure the theorist in her belief, see Kim, De Marchi, and Morgan (1995). It seems that characteristic tests are proposed to secure belief that consumption is determined by permanent income. Indeed, Friedman (1957) concludes that the consistency of the PIH with data supports the belief that PIH is a useful tool to explain the major apparent anomalies that arise if the observed regression between measured consumption and measured income is interpreted as a stable relation between permanent components. Characteristic tests have been described by Mayer (1972) as tests of direction rather than rigorous tests of the full theory: the tests' results in Friedman's empirical analysis follow the direction predicted by the PIH, but not necessarily by the amount predicted by the theory. More recently, De-

Juan and Seater (1999) revive Friedman's characteristic tests. Using two data sets of the Consumer Expenditure Survey (CEX) from 1980 to 1981 and from 1986 to 1991 they provide a rigorous specification of the consumer choice model, and find that their empirical results generally support the main implications of the PIH.

2.2.4 The life cycle hypothesis versus the permanent income hypothesis

In this section we aim to outline the main differences between the LCH and the PIH and to briefly review further developments on the empirical procedures that specifically test for the validity of the LCH. Both LCH and PIH give less weight to the effect of temporary changes in income on consumption as compared to AIH, and have similar implications with respect to the effects of shocks in income on individual or group behavior, though they are based on different assumptions. The main difference between the two theories is that the LCH assumes life is finite and structured, whereas the PIH states life is infinite. In terms of LCH this means that income and consumption exhibit, in addition to random shocks, systematic variations arising from the life cycle of income and consumption needs. Income and consumption reflect the succession of preworking, working and retired phase. Hence, the LCH explains systematic life cycle variations of saving and wealth that have many implications on aggregate. On the other hand, the PIH does not say much about aggregate, unless the effect on saving of random shock to income drives away from permanent income. However, the variations of transitory income are not too significant since they tend to cancel out under aggregation.

A number of tests which reject PIH, *e.g.* Flavin(1978), Hall (1981) and Campbell and Mankiw(1985), do not imply a rejection of LCH. This is because they impose very restrictive assumptions on preferences which are not required by LCH, *i.e.* infinite and uniform life and additive utility function. Hence, the PIH can be interpreted as a special case of the LCH based primarily on a distinction between "measured" and "permanent" income.

In general, the LCH has attracted considerably less attention than the PIH in the empirical literature. This is mainly because originally the LCH is not accompanied by suggestions for empirical tests. However, the LCH provides two main testable implications: the individual consumption depends on life resources rather than current income, cross-country differences in the aggregate saving rates result from differences in the rate of growth of incomes. The LCH has been tested empirically with both micro and macro data. When the theory is tested at micro level the most common approach is to test the LCH by estimating an Euler equation. Panel data are often used for this kind of analysis. Very often individuals are grouped by the year of birth such that the sample is divided in cohorts. This procedure follows a group of individuals who were born in the same year and thus come of age at the same time over time. In this way, one can track consumption and income in different periods of a sample of individuals that belong to the same generation. Attanasio and Weber (1995) and Attanasio and Browning (1995) among the others test the null hypothesis of validity of the LCH against the AIH using this approach. Attanasio and Browning (1995) build six cohorts and regress consumption and income against age. Here, they find strong correlation between consumption and income patterns and show evidence against the LCH. They also test the LCH against the hypothesis of excess sensitivity of consumption by estimating the Euler equation. In this case, when characteristics like household type, size and age are included in the regression, the coefficient on income growth becomes very low and insignificantly different from zero. This result is interpreted as evidence that the aggregation process creates a bias in the estimation and leads to wrongly rejecting the LCH when aggregate data are used.

Tests of the LCH using aggregate data focus on the following implications of the basic model: (i) the saving rate of a country is independent of its per capita income, two countries may have different rates of saving even though individuals have the same life path of saving and wealth, (ii) between economies where individuals have the same life path of saving and wealth, the one with the fastest growth can be expected to save most, (iii) the rate of saving depends on the rate of growth, with zero growth saving will be zero (regardless of income and habits, there might be saving only where there is growth).

When growth is due to productivity, if people expect a higher level of income in the future, and they choose a pattern of consumption dependent on the overall life resources and not on the actual growth path of income, then they would choose to consume more than their current income in the early stages of their life, and thus dissave. The cumulated negative wealth of these cohorts would be larger the faster the growth. In light of this the difference between LCH and PIH becomes clearer. The PIH assumes that life is infinite, thus there is no reason to build up resources for later dissaving. Once we take into account a finite life, retirement span is realistic therefore Friedman's conclusion does not necessary hold. It can be shown that the saving rate must necessarily rise at least for moderate rate of growth. For large growth, saving could reach a peak and start declining with more growth and might even become negative.

As already observed in Mankiw and Shapiro (1985) and Campbell and Deaton (1989), Clarida (1991) finds out that in the US data changes in consumption are much less volatile than are changes in observed income. He considers an overlapping generation model and shows that the change in per capita aggregate consumption is (approximately) a time-invariant function of the innovation in per capita labor income and that the standard deviation of the consumption change is significantly smaller than that of the income innovation. This evidence supports the LCH more than PIH, since workers save in order to smooth expected consumption during the remaining years of work and retirement. According to the LCH the propensity to consume out of any increment of labor income that is sustained during the working years will be less than one, as workers increase saving to finance higher consumption during retirement. Also, the propensity to consume out of permanent shifts in labor income declines monotonically with age. The economy's aggregate propensity to consume out of permanent changes in labor income is an average of the marginal propensity to consume of all working age cohorts. Even if individual consumption is by assumption unforecastable, the variance of changes in per capita consumption predicted by a properly aggregate life cycle model has to be substantially less than is implied by the representative agent PIH when shocks to per capita labor income are permanent.

Two extensions to the original formulation of the LCH involve liquidity constraints and the bequest motive. Adding those provides other empirically testable implications of the LCH such that the aggregate saving ratio should increase and the wealth-income ratio should decrease in line with to the rate of growth of income and saving is not only determined by growth but also by the duration of the retirement span, social security, precautionary motive and inheritance. Modigliani and Jappelli (1998) state that liquidity constraints can be included without changing very much the implication of the LCH. Liquidity constraints, precautionary savings and life uncertainty affect the age when you should start to observe negative saving and wealth decumulation, but not the main implication of the theory that individual wealth must eventually fall with age. Moreover, according to the first version of the theory wealth must be clearly declining after retirement, and at a sufficiently fast pace to reach exhaustion at the end of life. The actual behavior of wealth by age seems quite different. Empirical studies find that saving tends to decline after age 50 or 60 but does not turn negative, see Fisher (1950), Ando and Kennickell (1985). Other studies on a number of different countries tend to find that net worth reaches a peak around age 60 and then declines fairly steadily [Shorrocks (1975), King and Dicks-Mireaux (1982), Diamond and Hausman (1985), Hurd (1986)]. The decline does not appear as rapid as one might expect from the basic model but the differences may not be irreconcilable when one takes into account the fact that since the 1930s retirement ages have fallen and life expectancy has risen, and also the social securities revolution (social securities tend to reduce saving and offsetting the rise that should result from longer retirement span) and the bequest motive. An important issue here is which definition of bequest has to be involved: precautionary motive, response to uncertainty as to the time of death (agents leave some positive bequest when they die or invest in annuity values), implicit contract (whereby parents secure services from their children in exchange for some inheritance). When the bequest is generated by the true bequest motive, that is the utility of leaving an inheritance, we need to add further assumptions to the original model such as that the flow of bequest is proportional to income.[4] However, it can be shown that under certain conditions the implications of LCH are preserved. As a result, bequest does not change the nature of aggregate saving and wealth but adds several implications because bequest adds a layer to life cycle wealth.

Summarizing, we notice the PIH can be interpreted as a special case of the LCH as it imposes more restrictive assumptions on preferences which are not required by LCH. When those assumptions hold, tests of the PIH might also report on the validity of the

[4] Similarly to the Relative Income Hypothesis (RIH) it can be assumed that the proportion left should depend on the relative and not absolute life cycle resources. The size distribution of life resources should be reasonably stable over time.

LCH. Many authors that test the PIH by estimating the Euler equation refer to these tests as LCPIH tests where the null hypothesis of validity of both the LCH and the PIH is tested against the alternative of validity of the AIH, see Zeldes (1989), Runkle (1991), and DeJuan and Seater (1999). On the other hand, there are also important differences between the two theories. Most of them are related to the assumptions made on the time horizon considered. Under the LCH retirement, accumulation of resources and dissaving are important factors in the intertemporal choice of consumption that lead to more testable implications that have so far been empirically tested mainly at macro level.

2.3 Econometric Methodology

In this section we set up the framework for the empirical application and discuss various panel data estimation methods. We consider the three different specifications for testing the PIH, see *e.g.* Friedman (1957), Flavin (1981), DeJuan and Seater (1999). These models are estimated by Pooled OLS (POLS), Fixed Effect Model (FEM) and Generalized Method of Moments (GMM). The POLS estimation is likely to gain in efficiency due to the increased number of observations but estimation results might be biased due to neglected heterogeneity. The FEM explicitly takes into account the individual heterogeneity by specifying that unobserved fixed individual effects are correlated with the regressors. Since the fixed effect transformation wipes individual effects out, the FEM eliminates the source of possible correlation and leads to consistent estimations. But, the FEM leads to biased estimates in dynamic panels where the lagged dependent variables are also used as the explanatory variables and when the time series is short relative cross-sectional size, see Nelrove (1967, 1971), and Nickel (1981). In this case the within transformation induces correlation between the lagged dependent variable and the error term. In order to obtain consistent estimates in dynamic panels, we therefore turn to the Arellano and Bond's (1991) GMM estimator. It is obtained by differencing the equation in order to remove the effect, then estimating by instrumental variable using as instruments values of the dependent variable lagged two or more period. This treatment leads to consistent estimates even when the time dimension is fixed. However, more recently many authors have expressed their concern about the poor performance of the above mentioned GMM estimator in dynamic panels, see *e.g.* Ahn and Schmidt (1995), Kiviet (1995), Bond and Blundell (1998) and Alonso-Borrego and Arellano (1999). Their argument is that since the number of instruments increases with the time dimension (T), the model generates too many overidentifing restrictions even for moderate values of T although the quality of these instruments is often poor. In particular, Alvarez and Arellano (1998), Bond and Blundell (1998), Alonso-Borrego and Arellano (1999) and Blundell, Bond and Windmeijer (2000) show that the effect of weak instruments is to bias the distribution of the GMM estimator towards the FEM coefficient especially when the underlying series are highly autocorrelated. Also in an AR(1) regression, it is well known that the OLS estimate of the lagged variable in levels is biased upwards in the presence of individual-specific effects [Hsiao (1986)], and that the FEM estimate will be seriously biased downwards in short T panels [Nickell (1981)]. Therefore, it is also likely that the GMM estimate is biased downwards whereas a consistent estimate can be expected to lie in between the OLS and FEM estimates [Bond, Hoeffler, and Temple (2001)].

Alvarez and Arellano (1998) acknowledge that applied econometricians have tended to use less than the total number of instruments available in practice. Alonso-Borrego and Arellano (1999) propose a symmetrically normalized instrumental GMM of the Limited Information Maximum Likelihood type, which exhibits less bias than the conventional GMM via simulation studies. Another response to the weak instruments problem is to use further moment conditions that will improve the performance of the corresponding estimators, see Ahn and Schmidt (1995), Arellano and Bover (1995), Bond and Blundell (1998) and Blundell, Bond and Windmeijer (2000). Here we follow Bond and Blundell (1998) and Blundell, Bond and Windmeijer (2000) and will implement the so-called "system" GMM estimator that is a variant of the standard Arellano and Bond (1991) GMM procedure. Relying on relatively mild restrictions on the initial conditions, the system GMM estimator adds further moment conditions which are derived from the model in levels and improves the performance of the GMM estimator in the dynamic panels.

Finally, we also notice that we apply the Arellano and Bond (1991) GMM even when the lagged dependent variable is not included in the regressors in order to deal with problems of possible endogenous regressors and errors in measurement.

2.3.1 Aggregate consumption function approach

Recalling and simplifying Flavin' s specifications (2.10) and (2.11), we will test the null of validity of the PIH against the alternative hypothesis of excess sensitivity of changes in consumption using

$$\Delta c_{it} = \gamma + \phi \Delta y_{it} + \theta \hat{\varepsilon}_{it} + u_{it}, \tag{2.23}$$

$$y_{it} = \lambda + \delta y_{it-1} + \varepsilon_{it}, \tag{2.24}$$

where ϕ is the excess sensitivity parameter and

$$\varepsilon_{it} = \alpha_i + \eta_{it}. \tag{2.25}$$

Distinguishing between anticipated and unanticipated errors, we first estimate (2.24), construct the estimates of unanticipated error terms ε_{it}, then estimate (2.23) by POLS, FEM and standard and system GMM. In particular, since in (2.24)

$$E(y_{it-1}, \alpha_i) \neq 0, \tag{2.26}$$

the use of the system GMM is preferable. From (2.24), the following standard moment conditions are derived:[5]

$$E\left(\Delta \varepsilon_{it} y_{it-s}\right) = 0 \text{ for } i = 1, ...N,\ t = 3, ..T \text{ and } 2 \leq s \leq t-1. \tag{2.27}$$

The values of income variables lagged three and more times can be used as instruments to estimate the differenced model by the Arellano and Bond (1991) GMM procedure. Arellano and Bover (1995) and Blundell and Bond (1998) also consider the following moment conditions:

$$E\left(\alpha_i \Delta y_{i2}\right) = 0 \text{ for } i = 1, ...N, \tag{2.28}$$

[5] These conditions depend only on the assumed absence of the serial correlation in the time varying disturbance η_{it} in (2.25) and on the following conditions imposed on the initial value $E\left(y_{i1}\eta_{it}\right) = 0$ for $i = 1, ...N$ and $t = 2, ...T$, see Bond and Blundell (1998).

which requires a stationarity restriction on the initial conditions y_{i1}.[6] Combining (2.24) and (2.28) together, this yields the following $T-2$ linear moment conditions:

$$E\left(\varepsilon_{it}\Delta y_{it-1}\right)=0 \text{ for } t=3,...,T. \tag{2.29}$$

This allows us to use lagged differences of the dependent variable as possible instruments for equations in levels, as suggested by Arellano and Bover (1995). Blundell and Bond (1998) show that the calculation of the GMM estimator using the full set of moment conditions (2.27) and (2.28) can be based on a stacked system comprising all the $T-2$ equations in first-differences and the $T-2$ equations in levels corresponding to periods $3,...,T$, for which instruments are observed. This procedure is known as system GMM.

After obtaining $\hat{\varepsilon}_{it}$, we estimate equation (2.23) again by POLS, FEM and GMM.[7] As noticed above, the u_{it} in (2.23) represents both the effects of the information that the consumer may have about permanent income, which are not captured by the autoregressive specification of income, and the measurement error in the consumption changes. Attanasio and Weber (1993) also observe that uncertain current variables are to be treated as endogenous since the error term may contain an expectational component. Since income is a decision variable and u_{it} includes the forecast error, which arises from new information that affects both the consumption and the income choice [Zeldes (1989), Margier and Shaw (1993), Attanasio and Weber (1993), and De Juan and Seater (1999)], the income variables might be correlated with the error term even in equation (2.23) such that

$$E(\Delta y_{it},u_{it})\neq 0. \tag{2.30}$$

In this case we notice that only the Arellano and Bond (1991) GMM will provide a

[6] A sufficient condition is that the initial conditions y_{i1} satisfy the mean stationarity restriction $E(y_{i1}|\alpha_i)=\frac{\alpha_i}{(1-\phi)}$ for each $i=1,...,N$. Condition (2.28) holds if the means of the y_{it} series, whilst differing across individuals, are constant through time for periods $1,2,...,T$ for each $i=1,...,N$.

[7] We also acknowledge the problems caused by the "generated regressor" $\hat{\varepsilon}_{it}$ in (2.23), and thus the standard errors reported for the coefficient ϕ are likely to be incorrect. In the time series analysis, when a generated regressor is included in the model, the standard errors are underestimated and consequentially the values of the t-statistic are overestimated [Pagan (1984)]. The correct way to calculate standard errors in the case of generated regressor in panel data will be the interesting subject of future studies.

consistent estimation of (2.23), where the lagged values of income and consumption variables in levels are used as valid instruments.

2.3.2 Euler equation approach

For convenience we rewrite equation (2.17) as

$$\ln\left(\frac{C_{it+1}}{C_{it}}\right) = B_0 + B_1 r_{it+1} + B_2 \ln\left(\frac{F_{it+1}}{F_{it}}\right) + B_3 R_i + B_4 \ln\left(\frac{Y_{it+1}}{Y_{it}}\right) + e^*_{it+1}, \quad (2.31)$$

and we estimating (2.31) by OLS, FEM and GMM, respectively and test the null hypothesis of the PIH[8] ($H_0 : B_4 = 0$) against the alternative of AIH. Since the compound error term, e^*_{it}, is likely to comprehend the forecast errors, both income and interest rates may be correlated with the error term in (2.31) such that

$$E\left[\ln\left(\frac{Y_{it+1}}{Y_{it}}\right), e^*_{it}\right] \neq 0 \text{ and } E(r_{it}, e^*_{it}) \neq 0.$$

Also notice here that the household's tax rate is a choice variable because it depends on the household's income. Since the compound error term includes transitory consumption, the composition of consumption and thus the tax rate depends on the compound error term as well, see Zeldes (1989), and De Juan and Seater (1999). To overcome this problem, we need to use the GMM estimator, see also Zeldes (1989), Runkle (1991), Attanasio and Browning (1995), Attanasio and Weber (1995), and De Juan and Seater (1999).

Here we also suggest to use the Arellano and Bond's (1991) GMM estimation for consistent estimation of (2.31), where the choice of the instruments is internal to the model and instrument set includes lagged values of income and consumption in levels. We also extend (2.31) by incorporating the fixed time dummies in the compound error term and provide the Wald test results for joint significance of all the time dummies as a group. Such a component could arise from unanticipated macroeconomic shocks that lead all the households to make common mistakes in forecasting future economic variables.

[8] As mentioned before many authors refer to this procedure as a test of the LCPIH versus the AIH.

2.3.3 Characteristic tests

All the characteristic tests in Friedman's original work are derived in the following general scenario. Supposing that there are G groups of individuals for each group $g = 1, 2, ..G$ we have

$$C_{it} = \alpha + \beta Y_{it} + u_{it}, \tag{2.32}$$

where C_{it} is current consumption and Y_{it} is current income. As discussed earlier, measured income and consumption can be partitioned into systematic (permanent) and temporary components:

$$Y_{it} = Y_{it}^p + Y_{it}^t, \tag{2.33}$$

$$C_{it} = C_{it}^p + C_{it}^t. \tag{2.34}$$

The PIH can be seen as a model with errors in variables. If the sample is sufficiently large such that the sampling error can be neglected, this model assumes strict proportionality between the systematic components per household inside the group (unlike other linear models of "errors in variables"):

$$C_{it}^p = kY_{it}^p. \tag{2.35}$$

According to model (2.32), the OLS regression of consumption on income yields:

$$\beta = \frac{Cov\left(C_{it}, Y_{it}\right)}{Var\left(Y_{it}\right)}. \tag{2.36}$$

The regression coefficient measures the difference in consumption associated with a one dollar difference in measured income. Under the PIH, the size of this difference in consumption depends on two things:

$$\beta = \frac{Cov\left(C_{it}, Y_{it}\right)}{Var\left(Y_{it}\right)} = \frac{Cov\left(C_{it}, Y_{it}\right)}{Var\left(Y_{it}^p\right)} \times \frac{Var\left(Y_{it}^p\right)}{Var\left(Y_{it}\right)} = kP_Y, \tag{2.37}$$

where

$$P_Y = \frac{Var\left(Y_{it}^p\right)}{Var\left(Y_{it}\right)} \quad \text{and} \quad k = \frac{Cov\left(C_{it}^p, Y_{it}^p\right)}{Var\left(Y_{it}^p\right)} = \frac{Cov\left(C_{it}, Y_{it}^p\right)}{Var\left(Y_{it}^p\right)}. \tag{2.38}$$

P_Y measures how much of the difference in measured income is due to a difference in permanent income (since only differences in permanent income are regarded as affecting consumption systematically) and k measures how much of permanent income is devoted to consumption. If P_Y is equal to one, transient factors are either entirely absent or affect the income of all the members of the group by the same amount, *i.e.* $\beta = k$. If P_Y is equal to zero, there are no differences in permanent income, and the difference in measured income is associated with no systematic difference in consumption, *i.e.* $\beta = 0$.

An estimate of P_Y can be obtained from data on income of identical consumer units in different years. Since permanent income is not directly observable, Friedman (1957) proposes two alternative statistical estimates of P_Y derived under the mean and the variability assumption. First, under the mean assumption, the permanent component of each household's income changes in the same proportion as the average income of the group over two different time periods, t and s:

$$Y_{it}^{p} = mY_{is}^{p}, \tag{2.39}$$

where $m = \frac{\bar{Y}_t}{\bar{Y}_s}$ and $\bar{Y}_t = \frac{1}{N_g}\sum_{i=1}^{N_g} Y_{it}$ is the mean of measured income and N_g is the number of individuals in group g. Notice that if we assume that there is no correlation between transitory income in two different time periods

$$Cov\left(Y_{it}^{t}, Y_{is}^{t}\right) = 0, \tag{2.40}$$

the regression coefficient of measured income at time s on measured income at time t, denoted by B_{st}, can be written as:

$$\begin{aligned} B_{st} &= \frac{Cov\left(Y_t, Y_s\right)}{Var\left(Y_t\right)} = \frac{Cov\left(Y_t^{p} + Y_t^{t}, Y_s^{p} + Y_s^{t}\right)}{Var\left(Y_t\right)} = \frac{Cov\left(Y_t^{p} + Y_t^{t}, mY_t^{p} + Y_s^{t}\right)}{Var\left(Y_t\right)} \\ &= m\frac{Cov\left(Y_t^{p}, Y_t^{p}\right)}{Var\left(Y_t\right)} = mP_Y. \end{aligned} \tag{2.41}$$

Furthermore, we also define P_Y as

$$P_{Y_t} = B_{st}\frac{\bar{Y}_s}{\bar{Y}_t} = r_{ts}\frac{\sigma_t\bar{Y}_s}{\sigma_s\bar{Y}_t} \quad \text{and} \quad P_{Y_s} = B_{ts}\frac{\bar{Y}_t}{\bar{Y}_s} = r_{st}\frac{\sigma_s\bar{Y}_t}{\sigma_t\bar{Y}_s}, \tag{2.42}$$

where σ_t is the standard deviation of measured income at time t and r_{st} is the correlation coefficient between measured income at time t and s. Under the mean assumption, P_Y can be estimated by

$$P_{y_t} = b_{st} \quad \text{and} \quad P_{y_s} = b_{ts}, \tag{2.43}$$

where b_{st} is the regression coefficient in the regression of the logarithm of income at time s on the logarithm of income at time t and vice versa.

Under the variability assumption we assume that the fraction of the total variability contributed by the permanent component P_Y is the same in years t and s,

$$P_Y = \sqrt{P_s P_t} = r_{st}. \tag{2.44}$$

Then, P_Y is estimated simply by the correlation coefficient between measured incomes in two different years.

In sum, under the mean assumption we estimate P_Y by running the regression of current income at time t on income at time $t-1$ for each group[9]

$$y_{it}^g = \delta_0 + \delta_1 y_{t-1}^g + \varepsilon_{it}, \; g = 1, ..., G. \tag{2.45}$$

We denote, the estimates of P_y^g for each group g by $\eta_{yy_{-1}}^g$ $g = 1, .., G$. Next, under the variability assumption, we estimate P_y^g by the correlation coefficient between two adjacent years for each group denoted by $\rho_{yy_{-1}}^g$. Here we simply calculate the average of the correlation coefficients between adjacent periods (counting six in our sample of seven time observations). This choice is motivated by empirical evidence, also confirmed in our sample, that the correlation coefficient between two adjacent years is greater than the correlation coefficient between two non-adjacent years (Friedman, 1957, pp.186-187). Actually, the difference in the numerical results reflects an implicit difference in the definition of the permanent income component.[10] Nevertheless, the

[9] The size of correlation between two successive years provides some evidence of importance of the permanent component in producing differences in measured income (Friedman, 1957).

[10] Suppose that we have data on income for three consecutive years and denote P_3 as the fraction of variance contributed by permanent income in year 3. If we estimate by P_3 the correlation coefficient between income at time 3 and income at time 2, we implicitly define the permanent component as the component that is attributable to common factors affecting income in two or more successive years and the transitory component is attributable to factors affecting income in one and only one year.

decline of coefficients is numerically moderate; hence the results are not likely to be affected by the definition of the permanent component adopted.

In what follows we provide the four different operational versions of characteristic tests, see also De Juan and Seater (1999).

The first test

The first prediction of the PIH is that if transitory factors are either entirely absent or affect all members of the group by the same amount, the value of P_y is equal to the income elasticity of consumption and is close to one: $P_y = \eta_{cy} \leq 1$. Hence, the first test is based on comparing the estimates of the income elasticity and P_y. Income elasticity is computed on the basis of the following approximations. If transitory income and transitory consumption average out to zero over all households within a group, the average propensity to consume should be equal to k:

$$\frac{\overline{C}_{NT}}{\overline{Y}_{NT}} = k, \tag{2.46}$$

where $\overline{C}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} C_{it}$ and $\overline{Y}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}$. It follows that the elasticity of current consumption with respect to current income evaluated at the point of the sample means, can be written as:

$$\eta_{cy} = \beta\frac{\overline{C}_{NT}}{\overline{Y}_{NT}} = P_Y.$$

In order to perform this test the sample has to be divided according to some characteristic variables for which the relative variance of permanent income and transitory income are likely to differ. Here we select the 6 different groups: occupation, region, education, job status, economic status and marital status. For each group g, the estimates of P_y and η_{cy} are obtained. The values of η_{cy} are then regressed on P_y[11]:

$$\eta^g_{cy} = \alpha_0 + \alpha_1 \eta^g_{yy_{-1}} + \varepsilon, \tag{2.47}$$

$$\eta^g_{cy} = \alpha_0 + \alpha_1 \rho^g_{yy_{-1}} + \varepsilon. \tag{2.48}$$

According to the PIH, the estimated coefficient α_1 must be equal to one. Notice that because of the lack of data on the imputed rent of durable goods, we propose a weak form of the test where any positive relationship between P_y and η_{cy} will be interpreted as evidence in favor of the PIH. Finally, we calculate the Spearman and Pearson statistics as an alternative way to figure out whether there is any positive relation between η^g_{cy} and $\eta^g_{yy_{-1}}$ or $\rho^g_{yy_{-1}}$. The Spearman and Pearson correlation coefficients measure the strength of the linear relationship between two variables. The latter is a non parametric test, *i.e.* the data is believed not to follow the normal distribution, while the former is parametric; their statistical significance is tested using a t-test.

The second test

The second test is based on a common-sense intuition that the annual income of the self-employed is more volatile than that of the employees. It is also well established that income elasticities of the self-employed, η^s_{cy}, are empirically smaller than those of the employees, η^e_{cy}. The significance of this difference would then provide evidence in favor of the PIH. Households with more transitory income should have lower income elasticity of consumption than households with less transitory income. Hence, we divide the sample in two groups, employed and self-employed, and compare whether the self-employed have a lower income elasticity of consumption than employees. Assuming that η^s_{cy} and η^e_{cy} are independent, we also carry out a t-test under the null hypothesis that those income elasticities are equal against the one-sided alternative hypothesis that the elasticity for the self employed is less than the elasticity for employed, namely, $t_1 = \left(\eta^s_{cy} - \eta^e_{cy}\right) / \sqrt{\sigma^2_{\eta^s_{cy}} + \sigma^2_{\eta^e_{cy}}}$.

[11]For example in the BHPS households are divided in twenty groups depending on the occupation of the head of the household, i.e. $g = 1, 2, 3..., 20$ and $G = 20$ for occupation characteristic. Hence, in order to compute (2.47) or (2.48), first we use data within each of those twenty groups to calculate P^g_y and η^g_{cy} for $g = 1, 2, 3..., 20$. Then, the values of η_{cy} are regressed on P_y using twenty observations. For detailed information on the definitions of the classification variables and the frequency of the observations within each group see the Data Appendix.

The third and fourth tests

According to the proportionality hypothesis, the PIH predicts that the elasticity of permanent consumption with respect to permanent income is equal to one. The value of the elasticity of consumption with respect to current income is either equal to or less than one (strictly less than one where there is some transitory income). In light of this, two new relationships can be tested:$\eta_{cy} < \eta_{c^p y^p}$ and $\eta_{c^p y^p} < 1$.

In order to avoid the difficulty of measuring permanent income, the "test by group-means" method proposed by Ando and Modigliani (1960) is performed. If the mean of the transitory components of consumption and income average out at zero for each group then the differences between mean consumption and mean income should reveal a difference between permanent income and permanent consumption. The proxy for the elasticity of consumption with respect to permanent income is the mean group elasticity of consumption evaluated at the sample mean.

Following earlier studies we classify the whole sample by occupation, education and region, see Ando and Modigliani (1960), Mayers (1972) and DeJuan and Seater (1999). Then, for each characteristic variable we calculate the group-mean income elasticity of consumption by

$$\overline{c}_{NT} = \gamma \overline{y}_{NT} + \varepsilon. \tag{2.49}$$

$\eta_{\bar{c}\bar{y}}$ is estimated by the estimate coefficient of γ whereas the overall income elasticity η_{cy}, is estimated regressing measured consumption against measured income for the whole sample. We carry out the third test testing the null hypothesis that $\eta_{\bar{c}\bar{y}} = \eta_{cy}$ against the alternative of the proportionality hypothesis that $\eta_{cy} < \eta_{\bar{c}\bar{y}}$. Assuming that η_{cy} and $\eta_{\bar{c}\bar{y}}$ are independent of each other, we obtain the corresponding t-statistic by $t_3 = \left(\eta_{cy} - \eta_{\bar{c}\bar{y}}\right) / \sqrt{\sigma^2_{\eta_{cy}} + \sigma^2_{\eta_{\bar{c}\bar{y}}}}$.

In the forth test we test the null hypothesis that $\eta_{\bar{c}\bar{y}}$ is equal to one against the alternative that it is different from unity. The corresponding t-statistic is obtained by $t_4 = \left(\eta_{\bar{c}\bar{y}} - 1\right) / \sigma_{\eta_{\bar{c}\bar{y}}}$.

2.4 Empirical results

2.4.1 Data

In this study, we use data of the British Household Panel Survey (BHPS). The BHPS is a microeconomic survey that provides information on 8,167 individuals from 1991 to 1999 and data on household consumption of non-durable goods such as food, heating and fuel. The monthly mortgage or housing rent cost is the only information available on imputed rent of durable goods. Unfortunately, the expenditures on durable goods are not recorded continuously but only discrete values are available and therefore not comprehended in our definition of consumption. Also the official survey provides detailed information on family income but not on taxation. Thus, we refer to an unofficial survey "British Panel Survey Derived Current and Annual Net Household Income Variables" that collects data on annual disposal income from 1991 to 1997. The total household annual net income variable includes information about net labor income, investment income, benefit and pension income and transfer income. The BHPS also provides information about household characteristics like size and type of the household (which vary over time and across household) and the sex and race of the head of the household (that vary only across household). Data on the real after-tax interest rate for each household are derived using the following formula: $r_{it} = i_t (1 - \tau_{it}) - \pi_t$, where i_t is the nominal interest rate, τ_{it} is the average tax rate for household i and π_t is the inflation rate. For the nominal interest rate we use the one-year LIBOR (London InterBank Offered Rate) index[12] whereas the inflation rate is given by the Retail Price Index. All the data are deflated by the 1987 base-year Retail Price Index. In order to obtain a balanced panel we consider only the households responding to all the waves and also exclude the very few cases of households with annual net income less than £100, in order to avoid outliers in the sample. We therefore end up with a sample of 2,976 households that respond to all the seven waves and have an annual income greater than £100. A more detailed description of the variables used is provided in the Data Appendix.

[12] In the empirical literature the one-year Treasury Security index is commonly used and the UK Treasury Securities are quoted at three months, five, ten and twenty years. The one-year LIBOR index compares most closely to the one-year Treasury Security index.

2.4.2 Aggregate consumption approach

Table 2.1 presents the estimation results for the model given by (2.23) and (2.24). We provide the estimates of the OLS, FEM and GMM in Table 2.1, the GMM is the only consistent estimator as discussed earlier. The estimates of the intercept λ are summarized in the first column. Due to the data transformations, here the only available estimate is the OLS, which appears to be positive and significant. The estimates of δ are presented in the second column and they are all significantly positive. As expected, the OLS estimate seems to be biased upwards whereas the FEM estimate appears severely biased downwards. In addition the standard GMM estimate is very close to the FEM, which may reflect the problem of weak instruments. In this regard, the system GMM estimates are the preferred ones as explained earlier, which lie between the OLS and FEM estimates, a consistent finding with previous studies [see Bond, Hoeffler, and Temple (2001)].

The remaining columns of Table 2.1 show the alternative estimates of the parameters of equation (2.23). For instance, the first row shows the estimation results when unanticipated errors are residuals obtained from the GMM regression whereas in the second, the third and the fourth rows unanticipated errors are residuals of the system GMM, OLS and FEM regressions, respectively. First, we notice that the method of estimating the unanticipated errors, $\hat{\varepsilon}$, seems to make no difference on the estimates of the other parameters. The OLS and FEM estimates of ϕ (excess of sensitivity parameter) are positive and significant whereas the GMM estimates of ϕ all turn out to be insignificant. Also the estimates of θ (warranted change in consumption) are all significant when estimated by OLS and GMM and insignificant when estimated by FEM.[13] Recalling that the GMM is the only consistent estimator here, our findings provide a strong support for the PIH as change in consumption only depends on the innovation in the income process. This result differs from the evidence presented by earlier studies based on the aggregate time series consumption function approach, see Flavin (1981), Mankiw and Shapiro (1985), West (1988), Deaton and Campbell (1989), Quah (1989),

[13] Note that equation (2.23) has the same form as a Hausman test. For a simple regression model with only a single regressor, it can be shown that the Hausman test is equivalent to applying a standard test of significance to the coefficient on the vector of residuals obtained by regressing the regressor on the instrumental variable [see Greene (1997) for a detailed explanation]. Thus, here the significance of ε_t also confirms the appropriate use of the GMM estimator to estimate model (2.23).

Campbell and Mankiw (1990) and Galì (1991) and Deaton (1992). Here, we test the PIH with micro panel data using a specification commonly used for aggregate data and find empirical evidence in favor of the PIH. This result might give support to the thesis that the process of aggregation may vitiate the results of the previous studies conducted on aggregate data and often leads to the rejection of the PIH [see Attanasio (1998)].

Table 2.1. The PIH test results based on aggregate consumption function

			ϕ			θ		
	λ	δ	POLS	FEM	GMM	POLS	FEM	GMM
POLS	1.57* (0.04)	0.83* (.004)	0.13* (.01)	0.12* (.01)	0.24 (.15)	−.0005* (0.0002)	0.002 (0.007)	−.001* (0.0004)
FEM		0.17* (.008)	0.13* (.01)	0.13* (.01)	0.24 (.15)	−.0004* (0.0002)	0.002 (0.005)	−.001* (0.0003)
GMM		0.16* (.14)	0.13* (.01)	0.12* (.01)	0.2 (.18)	−.0004* (0.0001)	0.003 (0.007)	−.001* (0.0003)
SGMM		0.73* (.007)	0.14* (.01)	0.13* (.01)	0.2 (.19)	−.001* (0.0004)	0.003 (0.007)	−.003* (0.001)

Notes: λ and δ are estimated parameters of regression (2.24) and ϕ excess sensitivity parameter in (2.23). POLS, FEM GMM and SGMM stand for pooled OLS estimator, fixed effects model, Arellano and Bond's (1991) generalized method of moments, System GMM estimators, respectively. Standard errors are in parenthesis, and coefficients highlighted with * are significant at the 5% level of significance.

2.4.3 Euler equation specification

Table 2.2 presents the estimation results for (2.17) by POLS, FEM and GMM. Given the problems of correlation of some of the explanatory variables with the error term and of measurement error, above mentioned, the GMM is our preferred estimator. In the first column we present the results of the regression of the change in consumption against the change in net income and levels of the real after-tax interest rate variable. The coefficient of income is not significantly different from zero whereas the coefficient of the interest rate is significant. The value of the coefficient on interest rate is similar in size to the one obtained by DeJuan and Seater (1999) but rather low with respect to previous findings [Zeldes (1989), Runkle (1991) and Attanasio and Weber (1995)]. Here, a low coefficient may reflect the lack of enough intertemporal variation in the data, due to the short time dimension of our data set. We also extend by incorporating the $T-1$ fixed time dummies in order to capture the effects of aggregate shocks, and report such estimation results in the second column. The coefficient of income remains insignificant whereas the value of the coefficient of the interest rate increases greatly. The Wald test result for the joint significance of the year dummies as a group is statistically significant, implying that aggregate shocks are not negligible and influence the choices of consumption mainly through the interest rate.

Columns 3 and 4 report the estimation results when also including the variation in the size and in the type of household as explanatory variables. In the model without time dummies (column 3) all the coefficients are significant with the exception of the income variable. Variables of change in size and type of family are both significant, a consistent finding with the previous studies, see Attanasio and Browing (1995), Miller and Sieg (1997) and DeJuan and Seater (1999). As expected, changes in consumption are positively related to the change in the size of the household, it is a straightforward implication that household consumption increases as new components join. On the other hand, the estimates of the coefficient on the variable that represents the change in the type of household have no economic interpretation because of the way this variable has been inserted in the data set.[14] When included, the time dummies are not significant either as a group or singularly. Also in this case the only significant coefficients are those of the change in type and in size of household (see column 4).

Finally, in the last two columns 5 and 6 we also include the time invariant characteristics such as sex and race of the head of the household. In column 5 the only significant variables turn out to be the interest rate and the two time invariant characteristics. The sign on sex of the head of the household is positive, meaning that households with a female as head of the household consume more. The sign of the coefficient on race of the head of the household is negative but economic interpretation is unclear. Consistently with the previous cases, the magnitude of the coefficient on the interest rate in column 5 is low while it increases once the time dummies are included in the regression (column 6). Here, the variables that represent the change in the type and size of a household become significant whereas the individual time invariant characteristics lose their significance. The time dummies as a group appear to be significant. We also briefly mention that in the OLS and FEM estimations all the

[14] Type of the household is a qualitative variable that has not been converted into dummy variables, see Data Appendix for details.

coefficients appear to be mostly significant although those estimates are not consistent.

Summarizing, when (2.17) is consistently estimated the income variable coefficient in all the regressions appear not to be significantly different from zero. These findings provide further evidence in favor of the PIH and comply with the previous studies that test the PIH by Euler equation with panel data [Runkle (1991), Attanasio and Weber (1993), Attanasio and Browing (1995), and Attanasio (1998), DeJuan and Seater (1999)]. In particular, our findings outline the importance of testing PIH both with micro data and with the household characteristics in intertemporal choice of consumption.

Table 2.2. The PIH test results based on the Euler equation

	GMM						OLS	FEM
	1	2	3	4	5	6		
r	0.01* (.002)	1.26* (.75)	0.01* (.002)	0.56 (.62)	0.02* (.008)	1.28* (.61)	0.06* (.02)	0.01* (.002)
$\Delta \ln Y$	0.04 (.03)	0.03 (.04)	−.007 (.04)	−.002 (.12)	−.34 (.25)	−.28 (.49)	0.06* (.01)	0.05* (.007)
Size			0.88* (.48)	2.15* (1.33)	1.64* (1.07)	3.54 (1.72)	0.39* (.03)	0.38* (.02)
Type			−0.98* (.49)	−1.84* (.92)	−1.21* (.76)	−1.91 (.91)	−.04* (.02)	−.04* (.01)
Sex					0.84 (.27)	0.21* (.14)	0.006* (.001)	
Race					−1.05 (.35)	−.62* (.45)	−.003 (.016)	
Wald		32.1 [0.000]		4.2 [0.648]		8.3 [0.217]		

Notes: Column 1 shows the estimation results for the regression (2.17) where $\Delta \ln Y_{it}$ and r_{it} are explanatory variables and in column 2 we add the fixed $T-1$ time dummies. In column 3 we use $\Delta \ln Y_{it}$, r_{it}, $\Delta \ln \text{Size}_{it}$ and $\Delta \ln \text{Type}_{it}$ as explanatory variables and in column 4 we include the fixed $T-1$ time dummies. In column 5 we use $\Delta \ln Y_{it}$, r_{it}, $\Delta \ln \text{Size}_{it}$,$\Delta \ln \text{Type}_{it}$, Sex_i and Race_i as explanatory variables and in column 6 we add the fixed $T-1$ time dummies. Wald test is for the joint significance of the fixed $T-1$ time dummies, and p-values are in brackets. See also notes to Table 2.1.

2.4.4 Characteristic tests

Table 2.3(a) presents the estimation results for (2.47) and (2.48) when P_y is estimated under the mean assumption. There is evidence of a positive relation between η_{cy} and $\eta_{yy_{-1}}$in all cases except for marital status. Excluding this case, all the values of α_1 are significantly greater than zero and the Spearman and Pearson correlation coefficients are significantly different from zero. The strong version of the test generally supports the PIH, in all cases α_1 is not significantly different from one apart from the case of job status. Table 2.3(b) presents the results of the first test estimating P_y under the variability assumption. All cases, except for education and marital status, comply with the weak version of the test, whereas the cases of education and job status do not satisfy the strong version. Interestingly, in the case of marital status both under the mean and under the variability assumption α_1 turns out to be statistically neither greater than zero nor statistically different from one. This result can be explained by the low frequency of groups inside the marital status classification (see the Data Appendix for details) and therefore, marital status cannot be considered as a valid proxy for permanent income.

Table 2.3(c) presents the results of the second test. First notice that the values of sample average and sample standard deviation of current income confirm the intuition that income for self-employed is more volatile than income for employees. The fifth and sixth column show the estimates of P_y under the variability assumption, where ρ_{yy_1} is the average of correlation coefficients between two adjacent periods and ρ_{yy_3} is the correlation coefficient between the first and the fourth year. Column 7 gives the estimates of P_y obtained under the mean assumption. The numerical value of P_y is always greater for the case of employees. This indicates that both under the mean assumption and under the variability assumption a large proportion of income variation among self-employed is accounted for by the transitory factor. This confirms that classifying people as employees and self-employed is a valid way of identifying permanent and transitory income. We also find that the estimate of income elasticity for self-employed is significantly lower than for employees, as expected.

Table 2.3(d) presents the results of the third and the fourth test. The overall elasticity of consumption is numerically lower than the mean-group elasticity. This result is also supported by the t-tests. The values of t_3 leads to rejection of the null hypothesis of equality of elasticities in favor of the hypothesis that the overall income elasticity is lower than the mean-group elasticity. The last column of the table presents the results of the fourth test. Here, the mean-group elasticities are insignificantly different from unity regardless of what characteristic is used. These findings uniformly support the proportionality hypothesis.

Overall results of the characteristic tests give support to the PIH. This evidence is also in accordance with results reported in DeJuan and Seater (1998) where they split the sample in seven categories and find general support of the PIH.

Table 2.3. The PIH test results based on characteristic tests.

Table 2.3(a). The first test using the estimate of P_y under the mean assumption

	α_0	α_1	$\alpha_1 > 0$	$\alpha_1 = 1$	Spearman	Pearson
Occupation	$-.18$ (.29)	0.92 (.35)	y	y	0.476^*	0.546^*
Region	$-.25$ (.34)	0.98 (.4)	y	y	0.353^*	0.545^*
Education	$-.36$ (.36)	1.11 (.45)	y	y	0.771^*	0.776^*
Job status	-1.64 (.06)	2.63 (.08)	y	n	0.768^*	0.4^*
Economic status	$-.71$ (.54)	1.36 (.66)	y	y	0.771^*	0.718^*
Marital status	$-.440$ (.67)	1.07 (.82)	n	y	0.6	0.675

Notes: α_0 and α_1 are intercept and slope coefficients of regressions (2.47) and (2.48). Columns headed $\alpha_1 = 1$ and $\alpha_1 > 0$ report the outcomes of one-sided test at 5% significance. Spearman and Pearson denote the Spearman and Pearson correlation coefficients. See also notes to Table 2.1.

Table 2.3(b). The first test using the estimate of P_y under the variability assumption

	α_0^1	α_1	$\alpha_1 > 0$	$\alpha_1 = 1$	Spearman	Pearson
Occupation	$-.35$ $(.28)^2$	1.12 (.33)	y	y	0.58^*	0.64^*
Region	$-.26$ (.394)	0.97 (.46)	y	y	0.35^*	0.49^*
Education	0.63 (.12)	0.33 (.23)	n	n	0.43	0.47
Job status	-2.58 (.07)	3.73 (.08)	y	n	0.44^*	0.69^*
Economic status	-1.32 (.76)	2.06 (.91)	y	y	0.83^*	0.75^*
Marital status	-0.32 (.94)	0.92 (1.17)	n	y	0.8	0.49

Notes: See notes to Table 2.3(a).

Table 2.3(c). The second test

	N	$\bar{Y}$	S	ρ_{yy_1}	ρ_{yy_3}	$\eta_{yy_{t-1}}$	η_{cy}
Employees	1047	13864	6600	0.83	0.68	0.83 (.01)	0.55 (.02)
Self-employed	144	14840	10045	0.75	0.54	0.71 (.04)	0.24 (.05)
t_1	-4.19						

Notes: N denotes the sample size, $\bar{Y}$ is the mean income and S the sample standard deviation, ρ_{yy_1} is the average of the correlation coefficients calculated between two adjacent years, ρ_{yy_3} is the correlation coefficient between the first and the fourth year, $\eta_{yy_{-1}}$ is the estimation of P_Y under the mean assumption, η_{cy} denotes the income elasticity of consumption. t_1 is the value of the t-test of the null hypothesis that the income elasticities of the self employed and employed people are equal against the alternative hypothesis that the elasticity for self employed is less than the elasticity for employed. See also notes to Table 2.1.

Table 2.3(d). The third and fourth tests

Whole Sample	η_{cy}	Variables	$\eta_{\bar{c}\bar{y}}$	t_3	t_4
	0.59 $(.01)^2$	Education	0.87 (.005)	-19.1	-.13
		Occupation	0.87 (.005)	-19	-.128
		Region	0.88 (.001)	-26.4	-.123

Notes: η_{cy} denotes overall elasticity and $\eta_{\bar{c}\bar{y}}$ represents the mean-group elasticity. t_3 tests the null that the overall elasticity is equal to the mean-group elasticity against the alternative that the overall elasticity is less than the mean group elasticity, whilst t_4 tests the null that the mean-group elasticity is equal to one against the alternative that it is different from unity. See also notes to Table 2.1.

2.5 Conclusions

Regardless of the evident theoretical importance of the PIH in explaining intertemporal choice of consumption [see Cochrane (1989)] some of the empirical tests, mainly conducted on aggregate data, do not provide support for the theory. On the other hand, tests performed on micro data generally provide evidence in favor of the Hypothesis. In this analysis we present three alternative specifications to test the PIH using the same data set. First, we test for the validity of the PIH against the hypothesis of "sensitivity of consumption" using a model for aggregate consumption. Second, we test for the validity of the PIH against the validity of the AIH using the Euler equation specification. Finally, we use characteristic tests for testing some of the most important implications of the PIH against the validity of the AIH.

We estimate by Pooled OLS, the Fixed Effect Model, Arellano and Bond' s (1991) GMM and System GMM [Blundell and Bond (1998), and Blundell, Bond and Windmeijer (2000)]. The GMM procedures are mostly preferred mainly because it allows to deal with problems of possible endogenous regressors and errors in measurement. When models are consistently estimated the PIH receives general support from our data. This result is in conformity with the recent studies that use micro data to test the PIH and sharply contrasts the results of the analyses conducted on aggregate data [see Flavin (1981), Deaton (1982), West (1988), Deaton and Campbell (1989), Attanasio and Weber (1993), Attanasio and Browing (1995), Attanasio (1998)]. The most relevant result is that testing a model suitable for aggregate consumption with panel data provides evidence in favor of the PIH. Our analysis can be considered as a piece of evidence in favor of the thesis that empirical tests of the PIH, based on aggregate data, might suffer from mis-specifications or overlook some fundamental characteristics of micro data and therefore vitiate the results that lead to rejection of the PIH.

Possible extensions of this study might concern with the analysis of the impacts of a number of different variables on the choice of intertemporal consumption as the BHPS is a rather comprehensive source of information on individual and household. In particular, it might be interesting to better evaluate the role of the expectations and the role of intergenerational transfers (the BHPS contains information on expectation of future income as well as on pensions or benefits). Along this line of logic the differences between the PIH and the LCH might be better analysed and possibly tested.

2.6 Data Appendix

Variables definitions

Consumption: Consumption is defined by the aggregation of expenditure on total food and grocery bills, the expenditures on oil, gas and electricity and the expenditure due to mortgage or rent housing costs.

Race of the head of the household: White with value 1; Black-Caribbean with value 2; Black-African with value 3; Black-Other with value 4; Indian with value 5; Pakistani with value 6; Bangladeshi with value 7; Chinese with value 8; Other ethnic groups with value 9;

Sex of the head of the household: Male 0; Female 1.

Size of household: Contains information about the number of persons in the household, goes from 1 to 11.

Total household annual net income: Total household annual net income is a variable recorded in "British Panel Survey Derived Current and Annual Net Household Income Variables". It is the sum of total household annual net labor income (total annual household labor income minus household annual national insurance contributions, total household annual occupational pension contributions and minus total household annual income tax after credits[15]), total household annual investment income, total household annual benefit, total household annual pension income and total household annual transfer income.

Type of household: Households are divided in: single non elderly with value 1; single elderly with value 2; couple with no children with value 3; couple with dependent children with value 4; couple with non-dependent children with value 5; lone partner with dependent children with value 6; lone partner with non-dependent children with value 7; couple plus unrelated adults with value 8; other household with value 9.

Characteristic groups

The panels for each group within the classification variables are derived from a sample of households of the BHPS. We select 2,976 households responding to all the seven waves of the Survey. Each group is derived by dividing the whole sample ac-

[15]The total household annual income tax after credits is equal to total household annual income tax before credits minus total household annual credits on income tax.

cording to classification variables. Each group contains at least 50 households which belong to that particular group for all the survey time span.

Economic status: Economic status records information about each household. In dividing the sample with respect to this variable we select six groups with the following frequencies: self-employed 118; single or couple, all in full-time work 350; couple, one in full-time work, one part-time 70; couple, one in full-time work, one not working 59; one or more in part-time work 32; head of the household or spouse aged 60 or over 626.

Education: Education is an individual variable that provides information about the highest present academic qualification. In dividing the sample we refer to information regarding the head of the household. We select seven groups with the following frequencies: higher degree 50; 1st degree 225; HND, HNC, teaching 172; A level 400; O level 625; CSE 126; none of these 166.

Job status: Job status is an individual variable about the current labor force status. We divide the sample in five groups referring to information about the head of the household, with the following frequencies: employed 1047; self employed 144; in paid employed 999; retired 489.

Marital status: Marital status is an individual variable. We divide the sample in five groups referring to information about the head of the household, with the following frequencies: married 1369; living as couple 50; widowed 323; divorced 136; never married 303.

Occupation: Occupation is an individual variable that records the present job according to socio economic class, we divide the sample referring to the occupation of the head of the household. We select twenty groups with the following frequencies: high service class 213; low service class 302; routine non-manual workers 168; personal service workers 97; small proprietors with employee and without employee 89; foreman and technicians 112; routine manual workers 357; managers of large business 112; managers of small business 50; professional self-employed and professional employee 67; intermediate non-manual workers 160; intermediate non-manual foreman 54; junior non-manual workers 263; personal service workers 69; foreman manual 73; skilled manual workers 193; semi-skilled manual 144; unskilled manual workers 112; own account workers 74; farmers (employers), farmers (own account), smallholder and agricultural workers 50.

Region or Metropolitan Area: Region or Metropolitan Area provides information about the residence of household. We select eighteen groups with the following frequencies: Inner London 70; Outer London 146; Regions of South East 502; Regions of South West 248 East Anglia 131; East Midlands 254; West Midlands Conurb 98; Regions of West Midlands 152; Greater Manchester 104; Merseyside 58; Regions of North West 126; South Yorkshire 84; West Yorkshire 91; Regions of Yorkshire and Humber shire 87; Tyne and Wear 74; Regions of North 120; Wales 148; Scotland 259.

Chapter 3

A Panel Data Approach to Testing Anomaly Effects in Factor Pricing Models

3.1 Introduction

The central prediction of the asset pricing models of Sharpe (1964), Lintner (1965), and Black (1972) is that the market portfolio of invested wealth is mean-variance efficient. This implies that expected returns on securities are an exact positive linear function of their market betas. But, there have been several empirical findings which contradict the prediction of these models. The most prominent is the size effect of Banz (1981), who finds that the market value of equity adds to the explanation of the cross-section of average returns provided by market betas. More recently, there has been a large anomaly literature where firm specific characteristics such as leverage, past returns, dividend-yield, earnings-to-price ratios and book-to-market ratios as well as size help explain cross sectional returns. See for example Keim (1983), Fama and French (1992, 1996), Berk (1995) and Gauer (1999).

To accommodate these anomaly effects, a general procedure pursued in the literature is as follows. First, find characteristics that may prospectively be associated with average returns. Then sort portfolios based on those characteristics, compute betas for the portfolios and check whether differences in average return are accounted for only by the differences in the betas. Fama and French's (1993, 1996) model successfully explains the average returns of the 25 size and book-to-market sorted portfolios with three factors, namely, returns on the market, returns on a small minus big (SMB) portfolio and returns on a high minus low (HML) portfolio. Even if the choice of factors is motivated mostly by empirical experience and thus somewhat arbitrary, their three factor model has been widely used in evaluating various expected return puzzles.

One practically important issue is to check whether the factor pricing models need to be augmented by asset specific characteristics. For example, momentum effects, where a portfolio consists of short-term winners, have been found to be important [see Jagadeesh and Titman (1993)], violating the Fama and French three factor model, [see Fama and French (1996)]. In a similar vein, Daniel and Titman (1997), and Daniel, Titman and Wei (2001) have advanced ways to distinguish between factor models and characteristic models.

These anomalies have been attributed to market inefficiency but could be the result of a mis-specification of the underlying factor pricing model. The most popular approach to detect these anomaly effects has been the two pass (TP) cross-sectional regression methods, advanced by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), which have been widely used to evaluate linear factor pricing models, including the capital asset pricing model (CAPM), the arbitrage pricing theory (APT) and their variants [see Cochrane (2001) for an excellent survey]. In the first stage, the asset betas are estimated by time series linear regression of the asset's return on a set of common factors. Then, the cross sectional regression of mean returns on betas and characteristics is estimated, and the significance of asset specific regressors are evaluated along with factor risk premia estimation. The same approach could be applied to evaluating momentum anomaly effects using asset specific proxy variables for the past short term performance of portfolios (such as lagged portfolio returns).

However, it is well-established that the TP method suffers from the errors in variables (EIV) problem, because estimated betas are used in place of true betas in the second stage cross sectional regression. In this regard, many econometricians have suggested several ways to derive the EIV corrected standard errors of the TP estimators under a different set of assumptions. A detailed treatment of TP estimation and associated asymptotic theories can be found in Shanken (1985, 1992), Jagannathan and Wang (1998), and Ahn and Gadarowski (2001).

In this analysis we address the issue of testing for factor price mis-specification and apply the traditional two pass regression method, the Fama and French (1993) time series procedures and a panel data approach. The first two approaches have been broadly used in the literature and are presented here as preliminary to the panel data analysis. It is a salient fact that the benefits of using panel data techniques have been completely ignored. Perhaps one of the main reasons for this neglect is that in factor pricing models, all betas are heterogeneous in the first pass time series regression. As a result there is no room for exploiting the panel dimension since there are no homogeneous coefficients to estimate. Instead, the validity of the null hypothesis that the time series factor pricing model is correctly specified is in fact tested in the second pass cross sectional regression, for example, of pricing errors on characteristics. If our interest lies solely in testing the significance of these characteristics, we can show how to construct a panel data regression model with one set of variables varying over time such as common factors and another set of variables varying both over time and over asset portfolios. A statistical model where the parameters on factors are heterogenous and the parameters on characteristics are homogeneous is required to analyse the existence of anomalies in factor pricing models such as the CAPM or APT.

The current investigation provides a theoretically coherent example to which panel data techniques dealing with both homogeneous and heterogeneous parameters can be applied. This partially heterogeneous panel data model shares common features with the econometric framework recently proposed by Pesaran, Shin and Smith (1999), who develop dynamic heterogeneous panel estimation techniques that allow the simultaneous investigation of both homogenous long-run relationships and heterogenous short-run dynamic adjustment towards that long run relationship. Though similar, in spirit the exact econometric methodology developed and used in this study is different from that of Pesaran, Shin and Smith (1999), and is therefore developed separately here.

Our suggested panel-based anomaly tests have one clear advantage over TP-based tests; they are based on full information maximum likelihood estimates so that they do not suffer from the EIV problem and have all the usual asymptotic properties associated with likelihood tests. In addition the panel technique adopted here yields parameter estimates of firm specific effects that (under the alternative) are fully efficient.

In Section 4 we apply all three approaches to a large data set of UK stock returns between 1968 and 2002. The empirical results from the TP and the panel data regressions show the importance of book to market equity and market value in helping to explain asset returns. When such terms are added to the simple CAPM version of the model their significance is enormous. This confirms results from similar studies done on both US and UK data. Moreover, the three factor model is still mis-specified although, in terms of fits, it is an improvement over the single factor model. Perhaps even more important however are our findings from the panel data analysis. Here, contrary to the results of Fama and French (1996), we find that (i) adding size and book-to-market macro factors does not drive out the significance of a standard CAPM market factor, (ii) a firm specific book-to-market variable remains significant even after the basic CAPM factor is augmented by Fama French SMB and HML factors and (iii) a firm specific size variate remains likewise significant but generally only in subsamples drawn from the 1980's. We tentatively argue that the first and the second finding could be a result of the greater efficiency of our estimates and power of our ML testing procedure. We argue that the third finding supports Berk' s (1995) argument that the significance of firm size may be due to spurious coefficient bias rather than the existence of an asset pricing anomaly.

The next section presents a review on anomaly effect in factor pricing models. Section 3 presents an overview on modeling issue, outlines a heterogenous panel model within which factor pricing anomalies can be analysed and derives the econometric theory required for the panel data analysis. Section 4 gives an empirical illustration of the techniques applied to the UK excess stock returns. Section 5 concludes.

3.2 Review on anomaly effects in factor pricing models

The central prediction of the asset pricing models of Sharpe (1964), Lintner (1965), and Black (1972) is that the market portfolio of invested wealth is mean-variance efficient. This efficiency of the market portfolio implies that expected returns on securities are a positive linear function of their market βs, and only market βs suffice to describe the cross section of expected returns. But, there have been several empirical findings which contradict the prediction of these models. The most prominent is the size effect of Banz (1981). Using a sample including all common stocks quoted on the NYSE for at least five years between 1926 and 1975, Banz finds that market equity adds to the explanation of the cross-section of average returns provided by market βs. Following empirical research by Banz (1981), Reinganum (1981) and Keim (1983), Levis (1985) investigates on the presence of size and January effect in UK data. The rate of return data is drawn from the London Business School monthly returns file. The sample covers the period from January 1958 to December 1982. Portfolios are constructed by ranking all firms in a particular annual sample according to their market value at the beginning of each year and placing them in one of the 10 portfolios depending on their relative market position. Levis calculates the median market value and monthly average rates of return for each of the 10 portfolios over the entire period and five - year subperiods. He also calculates the differential return between the two extreme size portfolios (SML). Table 3.1 shows the results of the analysis. Over the entire 25 year period the smallest portfolio seems to outperform its largest counterpart by about 5% per annum. Furthermore the average portfolio returns decline quite uniformly as firm size increases. The differential return over the entire period, however, is not statistically significant at the conventional levels. The subperiod results clearly indicate that neither the portfolio returns nor the size premium have been stable over time. In fact size premium is evident in only 16 out of 25 years. Small firms, however, outperformed their larger counterpart throughout the seventies with the sole exception of 1975. This constitutes a dramatic reversal of performance in comparison to the early sixties and eighties when larger firms appear to have performed better than the smaller firms. Levis calculates OLS βs by regressing monthly portfolio returns against the monthly returns of the FTA (FT All share index) value-weighted market index for each period. The OLS βs increase monotonically with the size of the portfolio. They range from 0.0 for the smallest portfolio to around 1.0 for the portfolio of the largest firms. Such evidence is surprising as it contravenes both the main premise of modern financial theory about the positive trade-off between risk and return and other empirical evidence from other capital markets. Levis points out that the lower βs for smaller firms are in line with estimates provided by the London Business School Risk Measurement in spite of the difference in the respective methodologies. Levis also computes the month-by-month rates of return for each of the 10 portfolios, the FTA value-weighted and the differential return between the two extreme portfolios (SML). This analysis indicates the presence of more than one seasonal in the data. Not only are rate of returns not time invariant, but the monthly variations are not of the same pattern across the ten portfolios. For example, the mean returns for January and April are significantly different from zero for all the portfolios but the largest and the FTA index. The FTA index exhibits a similar pattern suggesting that these two seasonals, found across all portfolios, are predominantly due to overall market behavior. Given the British tax regime, the April seasonal could be taken as evidence of the tax-loss selling hypothesis. Beyond these two basic seasonalities the results reveal the small firm premium is not equally distributed across the year. Finally it is interesting to note that almost 50% of the total average size premium is earned during one single month, May.

Levis (1989) analyses stock market anomalies related to size earning or dividend yields in the UK market. The main aim of his analysis is whether such additional anomalies are independent of or related to market size. The evidence on this issue is rather controversial. While Reinganum (1981) and Banz and Breen (1986) argue that size effect subsumes the E/P effect, Basu (1983) asserts quite the opposite (i.e. size related anomalies disappear when one controls for the E/P effect). As the common factor between the three variables (market size, E/P, dividend yields and price) is the share price, it is not inconceivable that these three effects may be attributed to some underlying relationship between share price and stock returns. Levis uses data from the London Share Price Database from 1956 to 1985 constructs portfolios for size, E/P ratio, dividend yields and price. In order to control for the interaction between the four attributes , combined portfolios are constructed. All the firms are ranked first by a chosen criterion and quintiles are formed. Then within each quintile firms are re-ranked on a second variable and quintiles are formed for each combination of two attributes.

Table 3.1. Mean monthly returns for excess returns, FTA index and differential return and OLS β estimates. January 1958 to December 1982. Levis (1985)

		MV^1	Average retuns	OLS β Estimates^2
		1958-82	1958-82	1958-82
	Small	<500	1.32 (0.21)	0.31 (10.56)
	2	750	1.9 (0.24)	0.47 (17.23)
	3	1,250	1.22 (0.26)	0.56 (21.43)
	4	1,750	1.16 (0.27)	0.62 (23.86)
Size	5	2,750	1.13 (0.29)	0.73 (30.40)
	6	4,500	1.08 (0.31)	0.78 (33.45)
	7	6,750	1.08 (0.32)	0.83 (37.95)
	8	12,000	1.07 (0.34)	0.89 (44.13)
	9	27,000	1.09 (0.36)	0.98 (60.37)
	Big	91,000	0.93 (0.32)	1.01 (89.46)
	FTA		1.07 (0.35)	
	(SML)		0.39 (0.31)	

Notes: ^1MV is measured in thousands of pounds. $^2\beta$ is obtained by regressing monthly portfolio returns against the monthly returns of the FTA market index for each period. The values inside (.) indicate the standard errors.

Abnormal returns (u_{pt}) are estimated by subtracting from the actual portfolio return (R_{pt}) the returns predicted by the model used, given the market return (R_{mt}), the risk free (R_{ft}) and the parameter estimates (b_p). Two main models are employed:

$$u_{pt} = (R_{pt} - R_{ft}) - (R_{mt} - R_{ft}) \tag{3.1}$$

$$u_{pt} = (R_{pt} - R_{ft}) - b_p (R_{mt} - R_{ft}) \tag{3.2}$$

Model (3.1) can be regarded as a limit case of model (3.2), the simple CAPM framework, where all βs are assumed to be unity and the abnormal return is estimated by subtracting the market from the portfolio return. Model (3.2) involves a two-stage estimation procedure. The first stage consists of the estimation of the respective β coefficients. In the second stage the βs are used to obtain abnormal returns. The summary statistics of the full sample indicates that a firm size effect is also evident in the UK data and that there is a positive relation between dividend yield and returns. In short, the monthly average returns of the portfolios based on the four ranking procedures indicate that during the period April 1961 to March 1985, an investment strategy based on divided yields would have outperformed a E/P ratio strategy by 2.40 % per annum and a market size strategy by 1.68 %. The OLS β coefficients are rather surprising: they range from 0.7 for smaller size portfolio to 1.21 for the bigger size portfolios; furthermore the β estimates for all the other portfolios of the three other ranking procedures are essentially equal to unity. As Levis points out this evidence is obviously in sharp contrast with the US findings but not entirely surprising for the UK, see Dimson (1979) and Marsh (1979), Levis (1985). Levis also performs a t-test for testing whether the individual portfolio abnormal return is equal to zero and an F-test to test the hypothesis that the abnormal returns vector for a particular ranking procedure is equal to zero. The results suggest the presence of significant abnormal returns for all four ranking procedures. These findings are consistent with the hypothesis that the dividend yield and E/P ratio have significant impact on the risk-adjusted returns of UK firms and are largely independent from each other and other confounding effects. The market size anomaly resembles in many respects the share price effect but ceases to be pervasive when those portfolios are controlled for the differences in dividend yield and share price effects. Abnormal returns using (3.1) are calculated for all the pairs of primary and secondary portfolio grouping. This methodology is used to analyse to what extent the individual effects depend on the particular quintile of the portfolio formation. The emphasis is on the search of anomalies within the same quintile of the primary ranking variable. The size effect is not entirely independent of the other three portfolio formation procedures. The significant market size, for example, is markedly reduced when control over differences in dividend yield is exercised. Closer examination reveals that the market size effect is not consistent across all dividend yields, E/P or share price quintiles; from the configuration of abnormal returns across the various portfolio formation procedures it is often hard to distinguish between the size and share price effects. Levis concludes that these two variables are either proxies for each other or both are just proxies for more fundamental determinants of expected returns for common stocks. The dividends yield or the E/P ratio for example, appear as possible candidates for such a proxy. Their individual effects are still maintained even when control for their reciprocal differences is exercised, in spite of the fact that either one is consistent at every single level of the other. The evidence presented in this study raises questions about the strength of firm size as an independent determinant of the stock returns generating process. Its strong dependence with the other firm attributes suggests that it cannot be viewed as either an independent anomaly or a profitable investment strategy in its own right.

More recently, there has been a large anomaly literature, concerning the importance of the asset specific characteristics such that the cross-sectional pattern of stock returns can also be explained by size, leverage, past returns, dividend-yield, earnings-to-price ratios and book-to-market ratios. See for example Keim (1983) Fama and French (1992) and Berk (1995). Very often the choice of characteristics is motivated mostly by empirical experiences, and thus somewhat arbitrary and based on empirical evidence.

Fama and French (1992) use the cross-sectional approach of Fama and MacBeth (1973). Each month they regress cross-section of stock returns on size, β, leverage, earning price ratio and book to market equity. The time series means of the monthly regression slopes then provide standard tests of whether different explanatory variables are on average priced. Testing the null hypothesis that the coefficients of those variables are equal to zero means testing for efficiency in market portfolio under validity of the CAPM. They use all the non financial firms in the intersection of NYSE, AMEX and NASDAQ returns files from CRSP and merged with COMPUSTAT from 1963 to 1990. Table 3.2 shows the results of the regression. It is evident that size helps explain the cross-section of average stock returns. The average slope on size alone is -0.15% with a t-statistic of -2.58. This reliable negative relationship persists no matter which other explanatory variables are in the regressions. The size effect is thus robust in the 1963-1990 returns on NYSE, AMEX and NASDAQ stocks. In contrast β does not help explain average stock returns for the 1963-1990, it shows no power to explain average returns in Fama and MacBeth regressions that use various combinations of β with size, book-to-market, leverage and earning price ratio.

Notice that Fama and French use two leverage variables, the ratio of book assets to market equity, A/MV and the ratio of book asset to book equity, A/BE. A/MV is interpreted as a measure of market leverage whereas A/BE is a measure of book leverage. The regressions of the returns on the leverage variables show that the two leverage ratios are related to returns but with opposite signs. The average slopes are opposite in sign but close in absolute value. Fama and French conclude that it is the difference between market and book leverage, that is book-to-equity, that helps explain average returns. Also, the close link between the leverage and book-to-market results suggests two equivalent ways to interpret the book-to-market effect in average returns. A high ratio of book-to-market says that the market judges the prospects of a firm to be poor relative to firms with low BE/MV. Thus BE/MV may capture the relative distress effect postulated by Chan and Chen (1988). A high Book-to-Market ratio also says that firm's market leverage is high relative to its book leverage; the firm has large amount of market imposed leverage because the market judges that its prospects are poor and discounts its stock price relative to book value. The relative distress effect captured by BE/MV, can also be interpreted as an involuntary leverage effect, which is captured by the difference between A/MV and A/BE. The inclusion of E/P as an explanatory variable is justified by previous empirical findings. Ball (1978) claims that E/P is a catch-all for omitted risk factors in expected returns. If current earnings proxy for expected future earnings, high-risk stocks with high expected returns will have a low price relative to earnings. Thus, E/P should be related to expected returns, whatever the omitted sources of risk. This arguments only makes sense for firms with positive earnings. Thus, the slope for E/P in the Fama and MacBeth regression is based only on positive values. Fama and French use a dummy variable for E/P when earnings are negative. In the results the average slope on E/P dummy variable confirms that firms with negative earnings have higher average returns. The average slope for stocks with positive E/P shows that average returns increase with E/P when it is positive. Adding size to the regression kills the explanatory power of the E/P when it is positive. In general these results suggest that most of the relation between (positive) E/P and average return is due to the positive correlation between E/P and ln(BE/MV); firms with high E/P tend to have high Book-to-Market equity ratio. The results obtained by Fama and French show that the opposite rules of market leverage and book leverage in average returns are captured well by Book-to-Market. On the other hand, the relationship between E/P and average returns seems to be absorbed by the combination of size and book-to-market equity. Size and Book-to-Market equity capture all the cross sectional variation in average returns that is related to leverage and E/P.

Table 3.2. Average slopes from Month-by-Month Regressions of Stock Returns on β, Size, Book-to-Market Equity, Leverage and E/P. July 1963 to December 1990. Fama and French (1992).

β	$\ln(MV)$	$\ln(BTM)$	$\ln(A/MV)$	$\ln(A/BE)$	$E/PDummy$	$E(+)/P$
0.15 (0.46)						
	−0.15 (−2.58)					
−0.37 (−1.21)	−0.17 (−3.41)					
		0.50 (5.71)				
			0.50 (5.69)	−0.57 (5.34)		
					0.57 (2.28)	4.72 (4.57)
	−0.11 (−1.99)	0.35 (4.44)				
	−0.11 (−2.06)		0.35 (4.32)	−0.50 (−4.56)		
	−0.16 (−3.06)				0.06 (0.38)	2.99 (3.04)
	−0.13 (2.47)	0.33 (4.46)			−0.14 (−0.90)	0.87 (1.23)
	−0.13 (−2.47)		0.32 (4.28)	−0.46 (−4.45)	−0.08 (−0.56)	1.15 (1.57)

Notes: The values inside (.) indicate the t-ratio for the significance of an individual coefficient.

Strong and Xu (1997) examine the cross-section of the expected returns for UK equities. For the period 1973-1992, they test the relationship between expected returns and market value, book-to-market equity, leverage, earning price-ratio and β. As Table 3.3 shows average returns are significantly positively related to β, Book-to-Market and market leverage, and significantly negatively related to market value and book leverage. However, when the regression is augmented by either market value or any accounting based variables along with β, the latter becomes insignificant. Either Book-to-Market or leverage cause market value to become insignificant. The explanatory power of any combination of variables for average returns is low. In isolation, β has a positive coefficient. Market value has a significant negative coefficient, consistent with the small-firm effect. When they include β and market value together, β becomes insignificant while market value stays significantly negative. Of the accounting variables Book-to-Market has a consistent explanatory power for average returns t-statistic in the range 3.9 to 5.4. The leverage variables, included either alone or with market value cause the latter to be insignificant. Generally these results support Fama and French (1992). They also confirm that since Book-to-Market is the difference between on market and book leverage and the coefficients on market and book leverage are opposite in sign but close in absolute value, it is Book-to-Market equity that explains average returns. E/P enters in the regression as a dummy variable for the stocks with negative earnings and as actual E/P value for positive earnings.

Table 3.3. Average slopes from Month-by-Month Regressions of Stock Returns on β, Size, Book-to-Market Equity, Leverage and E/P. July 1973 to December 1992. Strong and Xu (1997)

α	β	$\ln(MV)$	$\ln(BE/MV)$	$\ln(A/MV)$	$\ln(A/BE)$	$E/PDummy$	$E(+)/P$
0.06 (0.08)	1.70 (2.77)						
2.22 (5.18)		−0.19 (−2.56)					
2.37 (4.64)	−0.06 (−0.14)	−0.19 (−2.26)					
1.88 (4.49)			0.51 (5.41)				
1.75 (4.64)				0.49 (5.02)	−0.34 (−2.35)		
1.68 (3.87)						0.75 (2.48)	0.96 (1.85)
2.14 (5.10)		−0.13 (−1.81)	0.34 (3.92)				
2.07 (5.63)		−0.13 (−1.76)		0.32 (3.60)	−0.25 (−1.71)		
2.06 (4.98)		−0.16 (−2.29)				0.41 (1.59)	0.46 (1.00)
2.10 (5.22)		−0.12 (−1.68)	0.36 (4.13)			0.27 (1.06)	−0.52 (−1.07)
2.07 (5.67)		−0.12 (−1.66)		0.33 (3.56)	−0.32 (−2.34)	0.27 (1.06)	−0.37 (−0.74)

See notes to Table 3.2.

When E/P is the only explanatory variable, the earning dummy is significantly positive. However in the regression including market value or accounting variables, both earnings variables become insignificant. The explanatory power of any combination of independent variables for returns, as measured by the average monthly R^2, never exceeds 8%, this means that the ability to explain the cross-section of returns with Fama and French variables is very low.

Strong and Xu conclude that Book-to-Market equity is a significant variable for explaining the cross section average return. Their results confirm for the UK the importance of Book-to-Market equity found previously for the US. Size dominates β in explaining average returns throughout the 1955-1992 period, but becomes insignificant when Book-to-Market equity is included for the 1973-1992 period.

To accommodate these anomaly effects, a general procedure pursued in the literature is to find characteristics that are associated with average returns, sort portfolios based on those characteristics, compute βs for portfolios and check whether the average return spread is accounted for only by the spread in βs. Fama and French (1993) use the time series regression approach of Black, Jensen and Scholes (1972). Using all NYSE stocks on CRSP from 1963 to 1991, they build six portfolios formed from sorts of stocks on MV and BE/MV meant to mimic the underlying risk factors in returns related to size and Book-to-Market equity, SMB and HML. They use excess returns on 25 portfolios, formed on size and Book-to-Market equity, as dependent variables in the time series regression. The time series regression slopes are factor loadings that have a clear interpretation as risk-factor sensitivities for stocks. The analysis proposed by Fama and French (1993) goes further than simply testing for the efficiency of the market under the CAPM. It is also convenient for studying if assets are priced rationally: variables that are related to average returns, such as size or book-to-market, must proxy for sensitivity to common (undiversifiable) risk factor in returns. In particular, the slopes and R^2 are direct evidence on whether different risk factors capture common variations in stock returns. Noticing that a well-specified asset-pricing model produces intercepts in excess return that are indistinguishable from 0, a multifactor asset-pricing model [see Merton (1973) and Ross (1976)] implies a simple test of whether the premium associated with any set of explanatory returns suffices to describe the cross-section of average return on the mimicking portfolio. In such regressions, a well specified asset-pricing model produces intercepts that are indistinguishable from zero.

In Fama and French (1993) the role of stock market factors in returns is developed in three steps. In the time series regressions stock returns are regressed on the excess market return, on SMB and HML and finally on market return together with SMB and HML. Tables 3.4(a), 3.4(b) and 3.4(c) show the results of those regressions. The R^2 values near 0.9 are for big-stock low-Book-to-Market portfolios. For small-stock and high-Book-to-Market portfolios R^2 values less than 0.8 or 0.7 are the rule. In these portfolios the market leaves much variation in stock returns that might be explained by other factors: SMB and HML turn out to have high explanatory power. When only SMB and HML are considered as explanatory variables 20 of the 25 R^2 values are above 0.2 and eight are above 0.5. Especially in the portfolios in large-size quintile, SMB and HML leave common variation in stock returns that is picked up by the market portfolio. Finally in the case where the three stock-market factors capture strong common variation in stock returns, the market βs are all more than 38 standard errors from 0. With one exception the t-statistic on the SMB slopes are greater than 4; most are greater than 10. SMB clearly captures a shared variation in stock return that is missed by the market and by HML. The slopes on SMB are related to size: in every Book-to-Market quintile, the slopes on SMB decrease monotonically from smaller- to bigger-size quintiles. Similarly the slopes on HML are systematically related to Book-to-Market: in every size quintile of stocks, the HML slopes increase monotonically from strong negative values for the lowest-Book-to-Market quintile to strong positive values for the highest Book-to-Market quintile. Except for the second Book-to-Market quintile, where the slopes pass from negative to positive, the HML slopes are more than five standard errors from 0. HML clearly captures shared variation in stock return, related to Book-to-Market equity, that is missed by the market and by SMB. Given the strong slopes on SMB and HML it is not surprising that adding the two returns to the regressions results in large increases in R^2. The market alone produces only two R^2 values greater than 0.9; in the three-factor regression, R^2 values greater than 0.9 are routine. Adding SMB and HML to the regressions has an interesting effect on the market βs. In the one-factor regressions the βs for the portfolio of stocks in the smallest-size and lowest-Book-to-Market quintiles is 1.40. At the other extreme, the univariate β for portfolio of stocks in the biggest-size and highest-Book-to-Market quintiles is 0.89. In the three-factor regressions the βs for these two portfolios are 1.40 and 1.06. In general, adding SMB and HML to the regressions collapses the βs for stocks towards 1.0; low βs move up towards 1.0 and high βs move down. The behavior is due to the correlation between the market and SMB and HML. Although SMB and HML are almost uncorrelated (-0.08), the correlation between excess market return, denoted by r^m and SMB and HML returns are 0.32 and -0.38.

Fama and French perform further tests based on the cross section of average returns.

It centers on the intercepts in the time series regressions of excess stock returns against excess market return, SMB and HML. When the excess market return is the only explanatory variable, the intercepts show size effect of Banz (1981). Except in the lowest-Book-to-Market quintile, the intercepts for the smallest-size portfolios exceed those for the biggest by 0.25% to 0.37% per month. In every size quintile, the intercepts increase with Book-to-Market. These results are parallel to Fama and French (1992) that market βs, used alone, leave the cross-sectional variation in average stock returns that is related to size and Book-to-Market. The two-factor regressions of stock returns on SMB and HML produce similar intercepts for the 25 portfolios. The two-factor intercepts are, however, large and close to or more than two standard errors from 0. Intercepts that are similar in size support the conclusion from the cross-section regressions in Fama and French (1992) that size and Book-to-Market factors explain the strong differences in average returns across stocks. But large intercepts also say that SMB and HML do not explain the average premium of stock returns. Adding the excess market return to the time-series regressions pushes the strong positive intercepts for stocks observed in the two-factor (SMB and HML) regressions to values close to 0. Intercepts close to 0 say that the regressions that use r^m and SMB and HML to absorb common time-series variation in returns do a good job explaining the cross-section of average stock returns. The smaller intercepts obtained when the excess market is included to the two-factor regression says that the size and Book-to-Market factors can explain the difference in average returns across stocks, but the market factor is needed to explain why stocks returns are on average above the one-month bill rate. Table 3.4(d) shows intercepts from excess returns regressions. Finally, Fama and French also propose a joint F-test on the regression intercepts to formally test the hypothesis that a set of explanatory variables produces regression intercepts for the stock portfolios that are all equal to 0. The F-test rejects the hypothesis that r^m suffices to explain average returns at the 0.99 level. This confirms that the excess market return cannot explain the size and Book-to-Market effect in average stock returns. The large positive intercepts for stocks observed when SMB and HML are the only explanatory variables produce an F-statistic that rejects the zero-intercepts hypothesis at the 0.98. The three stock-market factors produce the best-behaved intercepts. Nevertheless, the joint test that all intercepts are 0 rejects at about the 0.95 level. The rejection comes largely from the lowest-Book-to-Market quintile. Among stocks with lowest ratio of Book-to-Market equity, the smallest stocks have returns that are too low relative to the predictions of the three-factor model, and the biggest stocks have returns that are too high. In this case, the rejection of the three-factor model is due to the absence of a size effect in the lowest-Book-to-Market quintile. Despite its marginal rejection in the F-test, Fama and French conclude that the three-factor model does a good job on the cross section of average returns. In practical terms, only one of the 25 three-factor regression intercepts for stocks is much different from 0. The regressions produce intercepts for stocks that are close to 0, even though SMB and HML surely contain some of firm-specific noise as proxies for the risk factor in returns related to size and Book-to-Market equity.

Basically like the cross section regressions of Fama and French (1992), the time series regressions of Fama and French (1993) say that the size and book-to-market factors explain the average excess returns across stocks over one-month bill returns.

Fama and French (1996) extend the analysis of the three-factor model to portfolios formed on E/P, C/P and sales growth. Low E/P, low C/P and high sales growth are typical of strong firms that have negative slopes on HML. Since the average HML return is strongly positive, these negative loadings imply lower expected stock returns. Conversely, stocks with high E/P, high C/P or low sale growth tend to load positively on HML and have higher average returns. Following Lakonishok, Shleifer and Vishny (1994) Fama and French examine the returns on sets of deciles formed from sorts on Book-to-Market, E/P, C/P and five-year sales rank. Like Lakonishok, Shleifer and Vishny (1994) they find that past sales growth is negatively correlated to future returns. The estimates of the three-factor regression show that the three-factor model captures these patterns in average returns. As Table 3.5 shows, the regression intercepts are consistently small. Despite the strong explanatory power of the regression (most of the R^2 values greater than 0.92), the F-test never comes close to rejecting the hypothesis that the three-factor model describes average returns. In terms of both the magnitudes of the intercepts and the F-tests, the three-factor model does a better job than it does on the 25 Fama and French size-Book-to-Market portfolios. Higher-C/P portfolio produces larger slopes on SMB and especially on HML. Given the evidence that loadings on HML proxy for relative distress, Fama and French infer that low Book-to-Market, E/P and C/P are typical of strong stocks, while high Book-to-Market, E/P and C/P are typical of stocks that are relatively distressed. The patterns in the loadings of the Book-to-Market, E/P and C/P deciles on HML, and the high average values of HML largely explain how the three-factor regressions transform the strong positive relations between average return and these ratios into intercepts that are close to 0. The three-factor model performs very poorly when portfolios are formed on sales-rank portfolios. Recall that high sale-rank firms (strong past performers) have low future returns and low sales-rank firms (weak past performers) have high future returns. The three-factor model captures most of this pattern in average returns, largely because low sales-rank stocks behave like distressed stocks (they have stronger loadings on HML). Moreover except for the highest sales-rank decile the intercepts are close to 0. Although the intercepts for the sales-rank deciles produce the largest F-statistics, it is close to the median of its distribution when the true intercepts are all 0.

Table 3.4. Regressions reported in Fama and French (1993) July 1963 to December 1991

Table 3.4(a). Regression of excess returns, r, on the excess market return, r^m: $r_t = a + br_t^m + e_t$

Size quintile	Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	1.40	1.26	1.14	1.06	1.08	26.33	28.12	27.01	25.03	23.01
2	1.42	1.25	1.12	1.02	1.13	35.76	35.56	33.12	33.14	29.04
3	1.36	1.15	1.04	0.96	1.08	42.98	42.52	37.50	35.81	31.16
4	1.24	1.14	1.03	0.98	1.10	51.67	55.12	46.96	37.00	32.76
Big	1.03	0.99	0.89	0.84	0.89	51.92	61.51	43.03	35.96	27.75
			R^2					$s(e)$		
Small	0.67	0.70	0.68	0.65	0.61	4.46	3.76	3.55	3.56	3.92
2	0.79	0.79	0.76	0.76	0.71	3.34	2.96	2.85	2.56	3.25
3	0.84	0.84	0.80	0.79	0.74	2.65	2.28	2.33	2.26	2.90
4	0.89	0.90	0.87	0.80	0.76	2.01	1.73	1.84	2.21	2.83
Big	0.89	0.92	0.84	0.79	0.69	1.66	1.35	1.73	1.95	2.69

Table 3.4(b). Regression of excess returns on mimicking returns for size (SMB) and Book-to-Market (HML) factors: $r_t = a + sSMB_t + hHML_t + e_t$

Size quintile	Low	2	3	4	High	Low	2	3	4	High
			s					$t(s)$		
Small	1.93	1.73	1.63	1.56	1.67	22.52	21.38	21.88	22.30	22.16
2	1.52	1.46	1.35	1.18	1.40	17.23	17.68	17.08	15.47	16.42
3	1.28	1.12	1.05	0.93	1.16	14.43	13.89	13.42	12.13	13.45
4	0.86	0.82	0.77	0.72	0.95	10.16	9.64	9.29	8.57	10.02
Big	0.28	0.35	0.22	0.29	0.44	3.70	4.39	2.79	3.69	5.02
			h					$t(h)$		
Small	-0.95	-0.57	-0.35	-0.18	0.01	-9.72	-6.19	-4.10	-2.20	0.16
2	-1.23	-0.66	-0.38	-0.16	0.00	-12.25	-7.02	-4.20	-1.82	0.05
3	-1.09	-0.65	-0.31	-0.11	-0.01	-10.84	-7.07	-3.43	-1.23	-0.12
4	-1.11	-0.65	-0.36	-0.11	-0.01	-11.43	-6.69	-3.80	-1.12	-0.09
Big	-1.07	-0.65	-0.42	-0.06	0.08	-12.46	-7.07	-4.64	-0.06	0.81
			R^2					$s(e)$		
Small	0.65	0.60	0.60	0.60	0.59	4.57	4.31	3.98	3.79	4.01
2	0.59	0.53	0.49	0.42	0.44	4.68	4.41	4.20	4.06	4.53
3	0.51	0.43	0.37	0.31	0.35	4.71	4.31	4.19	4.10	4.60
4	0.43	0.30	0.24	0.18	0.23	4.53	4.55	4.40	4.48	5.06
Big	0.34	0.18	0.08	0.04	0.06	4.02	4.27	4.20	4.19	4.69

Table 3.4(c). Regression of excess returns on the excess market return, r^m, the mimicking returns for size (SMB) and Book-to-Market (HML) factors: $r_t = a + b\, r_t^m + sSMB_t + hHML_t + e_t$.

Size quintile	Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	1.04	1.02	0.95	0.91	0.96	39.37	51.80	60.44	59.73	57.89
2	1.11	1.06	1.00	0.97	1.09	52.49	61.18	55.88	61.54	65.52
3	1.12	1.02	0.98	0.97	1.09	56.88	53.17	50.78	54.38	52.52
4	1.07	1.08	1.04	1.05	1.18	53.94	53.51	51.21	47.09	46.10
Big	0.96	1.02	0.98	0.99	1.06	60.93	56.76	46.57	53.87	38.61
			s					$t(s)$		
Small	1.46	1.26	1.19	1.17	1.23	37.92	44.11	52.03	52.85	50.97
2	1.00	0.98	0.88	0.73	0.89	32.73	38.79	34.03	31.66	36.78
3	0.76	0.65	0.60	0.48	0.66	26.40	23.39	21.23	18.62	21.91
4	0.37	0.33	0.29	0.24	0.41	12.73	11.11	9.81	7.38	11.01
Big	-0.17	-0.12	-0.23	-0.17	-0.05	-7.18	-4.51	-7.58	-6.27	-1.18
			h					$t(h)$		
Small	-0.29	0.08	0.26	0.40	0.62	-6.47	2.35	9.66	15.53	22.24
2	-0.52	0.01	0.26	0.46	0.70	-14.57	0.41	8.56	17.24	24.80
3	-0.38	-0.00	0.32	0.51	0.68	-11.26	-0.05	9.75	16.88	19.39
4	-0.42	0.04	0.30	0.56	0.74	-12.51	1.04	8.83	14.84	17.09
Big	-0.46	0.00	0.21	0.57	0.76	-17.03	0.09	5.80	18.34	16.24
			R^2					$s(e)$		
Small	0.94	0.96	0.97	0.97	0.96	1.94	1.44	1.16	15.53	1.22
2	0.95	0.96	0.95	0.95	0.96	1.55	1.27	1.31	17.24	1.23
3	0.95	0.94	0.93	0.93	0.93	1.45	1.41	1.43	16.88	1.52
4	0.94	0.93	0.91	0.89	0.89	1.46	1.48	1.49	14.84	1.88
Big	0.94	0.92	0.88	0.90	0.83	1.16	1.32	1.55	18.34	2.02

Notes: R^2 and residual standard error, $s(e)$, are adjusted for degrees of freedom.

Table 3.4(d). Intercepts from excess returns regressions for 25 portfolios formed on size and Book-to-Market.

(i)			a					$t(a)$		
Size quintile	Low	2	3	4	High	Low	2	3	4	High
Small	-0.22	0.15	0.30	0.42	0.54	-0.90	0.73	1.54	2.19	2.53
2	-0.18	0.17	0.36	0.39	0.53	-1.00	1.05	2.35	2.79	3.01
3	-0.16	0.15	0.23	0.39	0.50	-1.12	1.25	1.82	3.20	3.19
4	-0.05	-0.14	0.12	0.35	0.57	-0.05	-1.50	1.20	2.91	3.71
Big	-0.04	-0.07	-0.07	0.20	0.21	-0.49	-0.95	-0.70	1.89	1.41
(ii)										
Small	0.24	0.46	0.49	0.53	0.55	0.97	1.92	2.24	2.52	2.49
2	0.52	0.58	0.64	0.58	0.64	2.00	2.40	2.76	2.61	2.56
3	0.52	0.61	0.52	0.60	0.66	2.00	2.58	2.25	2.66	2.61
4	0.69	0.39	0.50	0.62	0.79	2.78	1.55	2.07	2.51	2.85
Big	0.79	0.52	0.43	0.51	0.44	3.41	2.23	1.84	2.20	1.70
(iii)										
Small	-0.34	-0.12	-0.05	0.01	0.00	-3.16	-1.47	-0.73	0.22	0.14
2	-0.11	-0.01	0.08	0.03	0.02	-1.24	-0.20	1.04	0.51	0.34
3	-0.11	0.04	-0.04	0.05	0.05	-1.42	0.47	-0.47	0.71	0.56
4	0.09	-0.22	-0.08	0.03	0.13	1.07	-2.65	-0.99	0.33	1.24
Big	0.21	-0.05	-0.13	-0.05	-0.16	3.27	-0.67	-1.46	-0.69	-1.41

Notes: In model (i) excess returns are regressed on market factor; in model (ii) excess returns are regressed on SMB and HML; in model (iii) excess returns are regressed on market factor, SMB and HML.

Table 3.5. Three-Factor Time-Series Regressions for Monthly Excess Returns on the excess market return, r^m SMB and HML factors on equal weight deciles for BE/MV, E/P, C/P and Sales Growth. July 1963 to December 1991.Fama and French (1996):$r_t = a + b\, r_t^m + sSMB_t + hHML_t + e_t$

	1	2	3	4	5	6	7	8	9	10	F-stat	p
BE/MV	Low									High		
a	0.08	-0.02	-0.09	-0.11	-0.08	-0.03	0.01	-0.04	0.03	-0.00		
$t(a)$	1.19	-0.26	-1.25	-1.39	-1.16	-0.40	0.15	-0.61	0.43	-0.02	0.57	0.841
R^2	0.95	0.95	0.94	0.93	0.94	0.94	0.94	0.94	0.95	0.89		
E/P	Low									High		
a	-0.00	-0.07	-0.07	-0.04	-0.03	0.02	0.06	0.09	0.12	0.00		
$t(a)$	-0.07	-1.07	-0.94	-0.52	-0.43	0.24	1.01	1.46	1.49	0.05	0.84	0.592
R^2	0.91	0.95	0.94	0.94	0.94	0.94	0.94	0.94	0.92	0.92		
C/P	Low									High		
a	0.02	-0.08	-0.07	-0.00	-0.04	0.00	0.00	0.05	0.06	0.01		
b	1.04	1.06	1.08	1.06	1.05	1.04	0.99	1.00	0.98	1.14		
s	0.45	0.50	0.54	0.51	0.55	0.50	0.53	0.48	0.57	0.92		
h	-0.39	-0.18	0.07	0.11	0.23	0.31	0.36	0.50	0.67	0.79		
$t(a)$	0.22	-1.14	-1.00	-0.04	-0.51	0.00	0.06	0.72	0.92	0.14	0.49	0.898
$t(b)$	51.45	61.16	62.49	64.15	59.04	61.28	60.02	63.36	58.92	46.49		
$t(s)$	15.56	20.32	22.11	21.57	21.49	20.72	22.19	21.17	24.13	26.18		
$t(h)$	-12.03	-6.52	2.56	4.28	7.85	11.40	13.52	19.46	24.88	19.74		
R^2	0.93	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94	0.92		
5-Yr SR	High									High		
a	-0.21	-0.06	-0.03	-0.01	-0.04	-0.02	-0.04	0.00	0.04	0.07		
b	1.16	1.10	1.09	1.03	1.03	1.03	1.00	0.99	0.99	1.02		
s	0.72	0.56	0.52	0.49	0.52	0.51	0.50	0.57	0.67	0.95		
h	-0.09	0.09	0.21	0.20	0.24	0.33	0.33	0.36	0.47	0.50		
$t(a)$	-2.60	-0.97	-0.49	-0.20	-0.61	-0.25	-0.66	0.07	0.47	0.60	0.87	0.563
$t(b)$	59.01	70.59	67.65	65.34	56.68	68.89	62.49	54.12	50.08	34.54		
$t(s)$	25.69	25.11	22.59	21.65	20.15	23.64	21.89	21.65	23.65	22.34		
$t(h)$	-2.88	3.55	8.05	7.89	8.07	13.63	12.80	12.13	14.78	10.32		
R^2	0.95	0.96	0.95	0.95	0.93	0.95	0.94	0.93	0.92	0.87		

Notes: The F-statistic tests the null that intercepts are 0; p is the probability value.

Hussain, Diacon and Toms extend the empirical work on the three-factor model to the UK. They build 25 portfolios on size and Book-to-Market, E/P, C/P and ranking sale and compare the performance of the CAPM model against the three-factor model. As Table 3.6 (a) and Table 3.6 (b) show, the average of the 25 regressions adjusted R^2 for the CAPM and the three-factor model are 0.59 and 0.83. Also the intercepts in Table 3.6 (c) have significant t ratios in many cases and are therefore clearly non-zero. In contrast, since the smaller average absolute intercept is 0.22% for the three-factor model and 0.35% for the CAPM, the three-factor model appears to capture more variation in the average returns on the portfolios than the CAPM. They also formally test the effect of the addition of SMB and HML to the CAPM model conducting an F-test on the incremental effect of these variables for the 25 portfolios formed on size and Book-to-Market. The F-test is strongly significant in all the cases. CAMP tends to have less need for the addition of SMB and HML for large firms that are not distressed. However, even in these cases the three-factor model is significantly superior in explaining the pattern of stock market returns. For portfolios formed on Book-to-Market, E/P and C/P the average adjusted R^2 for the regressions using the CAPM are 0.74, 0.75 and 0.75 for Book-to-Market, E/P and C/P for the three-factor model increase to 0.89, 0.88 and 0.89 on average (see Table 3.6 (d)). The average absolute intercept decreases using the three-factor model and the results of the F-test confirm the incremental significance of the additional variables. Comparing the R^2, the average absolute intercept and the F-test between the single factor and the three-factor model, the three-factor model appears to be a better model than CAPM for absorbing market anomalies for portfolios formed on Book-to-Market, E/P and C/P. Considering the regression slopes on average, excluding the lowest C/P and E/P portfolios, the trend is for higher Book-to-Market, E/P and C/P to produce larger slopes on SMB and especially HML. The pattern in the loadings of the Book-to-Market, E/P and C/P deciles on HML, and the high average value of HML largely explain how the three-factor regressions transform the strong positive relations between average return and these ratios into intercepts that are closer to zero than CAPM. Among the different portfolios the three-factor model has the greatest difficulty with the returns on the sale-rank portfolios. The three-factor model appears to work better largely because low sale-rank stocks behave like distress stocks, they tend to have stronger loadings on HML relative to non-distress stocks. The average of the 10 regressions R^2 for the CAPM and three-factor model are 0.75 and 0.89. Also the t statistics on the intercepts show that they are distinguishable from zero. The average absolute intercept when using the three-factor model as opposed to CAPM decreases from 0.27 to 0.14.

Hussain, Diacon and Toms conclude that all the market anomalies stated for the US using CAPM model also seems to hold for the UK. The three-factor model only seems to capture the returns to portfolios formed on E/P whereas Book-to-Market, C/P, size and sales growth still appear to be market anomalies in the UK. However, the three-factor model seems to be an improvement on the CAPM. Like the US stocks with high Book-to-Market, high E/P, high C/P, or low sale growth tend to load positively on HML, this is because they are relatively distressed and have higher average returns. Conversely low Book-to-Market, low E/P, low C/P, or high sale growth are typical of strong firms that have negative slopes on HML, these negative slopes imply lower expected returns.

Table 3.6. Regressions reported in Hussain, Diacon and Toms (1997) July 1963 to December 1991.

Table 3.6(a). Regression of excess returns on the excess market return, r^m:$r_t = a + b\, r_t^m + e_t$

Size quintile	Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	0.50	0.50	0.52	0.48	0.51	8.76	10.16	9.11	11.39	12.16
2	0.68	0.58	0.65	0.63	0.71	16.10	15.22	17.40	16.87	17.76
3	0.70	0.68	0.72	0.77	0.82	21.64	20.14	20.06	20.89	17.88
4	0.86	0.85	0.90	0.91	0.94	29.02	30.07	29.19	31.13	23.29
Big	0.98	1.04	1.07	1.10	1.07	55.71	55.22	57.17	47.38	35.17
			R^2					Adj-R^2		
Small	0.21	0.26	0.22	0.31	0.34	0.21	0.26	0.22	0.31	0.33
2	0.47	0.44	0.51	0.49	0.52	0.47	0.44	0.51	0.49	0.52
3	0.62	0.58	0.58	0.60	0.52	0.62	0.58	0.58	0.60	0.52
4	0.74	0.76	0.75	0.77	0.65	0.74	0.76	0.74	0.77	0.65
Big	0.91	0.91	0.92	0.89	0.81	0.91	0.91	0.92	0.88	0.81

Table 3.6(b). Regression of excess returns on the excess market return, r^m, the mimicking returns for size (SMB) and Book-to-Market (HML) factors.

$r_t = a + b\, r_t^m + sSMB_t + hHML_t + e_t$

Size quintile	Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	0.86	0.90	0.91	0.80	0.85	15.41	21.71	17.92	25.08	28.29
2	1.01	0.93	0.98	0.95	1.06	26.89	34.13	44.26	42.99	50.90
3	0.98	0.98	1.04	1.07	1.17	37.68	41.95	44.95	48.21	41.85
4	1.08	1.08	1.14	1.09	1.13	39.58	47.34	46.30	41.57	31.12
Big	1.07	1.09	1.11	1.14	1.06	56.03	49.31	54.60	47.25	35.19
			s					$t(s)$		
Small	1.04	1.12	1.15	0.95	1.00	12.05	17.58	14.64	19.32	21.45
2	0.92	0.98	0.97	0.95	1.03	15.58	23.28	28.39	27.94	32.08
3	0.78	0.86	0.91	0.90	1.06	19.35	23.81	25.71	26.32	24.65
4	0.62	0.65	0.69	0.53	0.63	14.68	18.51	18.28	13.27	11.16
Big	0.23	0.15	0.14	0.16	0.07	7.89	4.41	4.47	4.25	1.52
			h					$t(h)$		
Small	0.22	0.07	0.34	0.40	0.41	1.76	0.72	2.99	5.68	6.14
2	-0.17	0.04	0.29	0.35	0.53	-2.05	0.67	6.01	7.05	11.46
3	-0.13	0.14	0.29	0.47	0.74	-2.25	2.69	5.72	9.65	12.05
4	-0.10	0.14	0.24	0.37	0.70	-1.62	2.77	4.42	6.42	8.71
Big	-0.22	0.07	0.31	0.47	0.78	-5.09	1.46	6.85	8.92	11.68
			R^2					Adj-R^2		
Small	0.49	0.65	0.58	0.73	0.77	0.49	0.65	0.57	0.73	0.77
2	0.72	0.81	0.88	0.88	0.91	0.71	0.81	0.88	0.88	0.91
3	0.83	0.87	0.88	0.90	0.88	0.83	0.86	0.88	0.90	0.88
4	0.85	0.89	0.89	0.88	0.81	0.85	0.89	0.89	0.88	0.81
Big	0.93	0.92	0.94	0.92	0.87	0.93	0.92	0.94	0.92	0.87

Table 3.6(c). Intercepts from excess returns regressions for 25 portfolios formed on size and Book-to-Market.

(i)			a					$t\,(a)$		
Size quintile	Low	2	3	4	High	Low	2	3	4	High
Small	0.23	0.63	0.96	0.97	1.22	0.78	2.52	3.31	4.54	5.65
2	-0.14	0.20	0.24	0.42	0.74	-0.63	1.01	1.27	2.20	3.59
3	-0.18	0.02	0.28	0.40	0.48	-1.08	0.13	1.53	2.15	2.04
4	-0.16	0.01	0.04	0.24	0.38	-1.07	0.08	0.27	1.62	1.85
Big	-0.24	-0.10	-0.01	0.08	0.38	-2.69	-1.02	-0.13	0.65	2.45
(ii)										
Small	-0.14	0.31	0.51	0.55	0.78	-0.58	1.76	2.31	3.96	5.96
2	-0.30	-0.08	-0.14	0.01	0.23	-1.84	-0.64	-1.51	0.14	2.56
3	-0.32	-0.26	-0.09	-0.04	-0.13	-2.87	-2.62	-0.88	-0.44	-1.09
4	-0.28	-0.22	-0.24	-0.06	-0.10	-2.37	-2.25	-2.31	-0.56	-0.62
Big	-0.21	-0.17	-0.19	-0.18	0.01	-2.50	1.77	-2.13	-1.70	0.10

Notes: In model (i) excess returns are regressed on market factor; in model (ii) excess returns are regressed on market factor, SMB and HML.

Table 3.6(d). Regression of excess returns on the excess market return, r^m and the mimicking returns for size (SMB) and Book-to-Market (HML) factors on equal weight deciles for BE/ME, E/P and Sales Growth: $r_t = a + b\, r_t^m + sSMB_t + hHML_t + e_t$

Deciles	1	2	3	4	5	6	7	8	9	10
BE/MV	Low									High
a	-0.25	-0.15	-0.28	0.11	-0.13	-0.21	-0.09	0.14	0.30	0.44
$t(a)$	-2.24	-1.46	-2.74	-1.14	-1.45	-2.03	-0.81	1.53	3.14	3.86
R^2	0.88	0.87	0.88	0.89	0.91	0.89	0.88	0.91	0.90	.086
E/P	Low									High
a	-0.03	-0.15	0.03	-0.13	-0.11	-0.03	-0.10	0.04	0.05	0.10
$t(a)$	-0.25	-1.63	0.32	-1.43	-1.25	-0.31	-1.00	0.37	0.53	0.78
R^2										
C/P	Low									High
a	-0.48	-0.29	-0.20	-0.16	-0.05	-0.02	-0.04	0.10	0.27	0.54
b	0.98	1.00	1.04	1.04	1.02	1.07	1.08	1.05	1.12	1.12
s	0.63	0.49	0.54	0.55	0.58	0.57	0.59	0.68	0.72	0.82
h	0.15	0.09	0.12	0.04	0.19	0.26	0.38	0.33	0.42	0.38
$t(a)$	-3.90	-2.99	-2.38	-1.67	-0.56	-0.21	-0.39	0.95	2.80	4.84
$t(b)$	34.70	44.26	52.48	47.56	46.00	45.08	46.02	43.10	49.84	42.99
$t(s)$	14.37	14.24	17.70	16.18	17.07	15.54	16.24	18.09	20.68	20.52
$t(h)$	2.42	1.84	2.79	0.89	3.98	4.87	7.39	6.05	8.44	6.55
R^2	0.82	0.88	0.91	0.90	0.89	0.89	0.90	0.88	0.91	0.88
5-Yr SR	High									Low
a	-0.40	-0.20	-0.14	-0.11	-0.03	0.01	0.04	0.23	0.05	0.22
b	1.20	1.12	1.08	1.06	1.07	1.02	1.00	0.96	0.98	1.03
s	0.63	0.69	0.57	0.62	0.61	0.57	0.56	0.56	0.63	0.72
h	0.11	0.18	0.20	0.26	0.20	0.26	0.29	0.32	0.28	0.24
$t(a)$	-3.29	-2.00	-1.35	-1.08	-0.33	0.08	0.39	2.44	0.51	2.36
$t(b)$	42.59	47.58	44.63	44.31	50.29	47.68	43.84	44.24	45.26	46.98
$t(s)$	14.52	18.83	15.41	16.74	18.62	17.41	16.03	16.57	19.00	21.41
$t(h)$	1.80	3.46	3.84	4.91	4.30	5.59	5.82	6.61	5.93	5.00
R^2	0.88	0.90	0.89	0.88	0.91	0.90	0.88	0.89	0.89	0.89

3.3 Overview on Modelling Issues

It is nowadays standard to assume that returns on the individual portfolio (or the individual stock returns) are linearly generated by multiple common factors,

$$r_{it} = a_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{3.3}$$

where r_{it} is the excess return of assets in the portfolio i at time t, $\mathbf{f}_t$ is the k vector of factors, a_i is the portfolio-specific intercept term, $\boldsymbol{\beta}_i$ is the k vector of betas (factor loadings) of portfolio i corresponding to $\mathbf{f}_t$, and ε_{it} is assumed to be the zero mean idiosyncratic error for portfolio i at time t. This model includes the standard CAPM and multiple factors model as special cases. The (linear) beta-pricing restrictions imposed on (3.3) is given by

$$H_0 : E(r_{it}) = \gamma_0 + \boldsymbol{\beta}_i' \gamma_1,\ i = 1, ..., N, \tag{3.4}$$

where $E(r_{it})$ is the expected return on assets in the portfolio i and expectations are taken over time, γ_0 is an unknown constant (*e.g.* zero-beta expected return), γ_1 is the k vector of associated factor risk premia. If (3.4) holds then asset markets are efficient in the sense that there are no (asymptotic) gains to arbitrage. However, as mentioned in the previous section, there is mounting empirical evidence that asset specific factors are also priced. To the extent that these asset specific factors have idiosyncratic components (*i.e.* sources of risk that are diversifiable), then their pricing is incompatible with zero (asymptotic) arbitrage.[1] Specifically, previous studies, most famously that by Banz (1981) have added asset specific regressors to (3.4) and have estimated alternative models of the form

$$H_A : E(r_{it}) = \gamma_0 + \boldsymbol{\beta}_i' \gamma_1 + \mathbf{s}_{it}' \gamma_2,\ i = 1, ..., N, \tag{3.5}$$

[1] Fama and French (1996) argue that most of the asset specific variables (particularly size and book to market) that generate anomalies in this way can be accounted for by additional pricing factors. We do not enter this debate here but focus on tests of pre-specified factor models against asset specific alternatives. Interestingly, Fama and French also admit that their factors cannot drive out the significance of own lagged returns in the cross section. The inclusion of variables such as own lagged returns makes the model a heterogenous dynamic panel but does not raise any problems for our approach as will be shown below.

where $\mathbf{s}_{it}$ is a q vector of asset specific variables such as size or book-to-market value for assets in the portfolio i at time t, γ_2 is a $q \times 1$ vector of unknown parameters of return premiums associated with $\mathbf{s}_{it}$. However, it is worth noticing that characteristics like size might be proxied by variables that are possibly endogenous, leading to spurious correlations as discussed by Berk (1995).

To test the null model against the alternative model H_A, a traditional two pass (TP) regression method has been applied to (3.5). To estimate $\gamma = (\gamma_0, \gamma_1', \gamma_2')'$, we run the second pass cross sectional regression (CSR),

$$\bar{r}_i = \gamma_0 + \hat{\boldsymbol{\beta}}_i'\gamma_1 + \bar{\mathbf{s}}_i'\gamma_2 + \eta_i, \; i = 1, ..., N, \tag{3.6}$$

where $\bar{r}_i = T^{-1}\sum_{t=1}^{T} r_{it}$, $\bar{\mathbf{s}}_i = T^{-1}\sum_{t=1}^{T} s_{it}$, and $\hat{\boldsymbol{\beta}}_i$ are the OLS estimates of $\boldsymbol{\beta}_i$ obtained from the first pass time series regression (3.3). Alternatively, Fama and MacBeth (1973) considered a rolling CSR in each time period t,

$$r_{it} = \gamma_{0t} + \hat{\boldsymbol{\beta}}_i'\gamma_{1t} + \mathbf{s}_{it}'\gamma_{2t} + \eta_{it}, \; i = 1, ..., N. \tag{3.7}$$

where $\hat{\boldsymbol{\beta}}_i$ is estimated using time series observations 1 through $t-1$. Once the consistent TP estimator of γ, denoted $\hat{\gamma}_{TP}' = \left(\hat{\gamma}_{0,TP}, \hat{\gamma}_{1,TP}', \hat{\gamma}_{2,TP}'\right)'$, is obtained, the validity of the asset price restriction (3.4) can be evaluated by testing $H_0 : \gamma_2 = 0$, using for example a Wald test statistic given by

$$Wald = \hat{\gamma}_{2,TP}' \left[Var\left(\hat{\gamma}_{2,TP}\right)\right]^{-1} \hat{\gamma}_{2,TP}, \tag{3.8}$$

which is distributed as χ_q^2 under the null.

A well-known problem with this TP-based estimation is that the use of estimated betas in the second pass regression generates an errors in variables (EIV) problem. There has been a large literature attempting to derive the EIV corrected standard errors of the TP estimators under different sets of assumptions. In particular, with an arbitrary positive definite weighting matrix, the TP estimator can be obtained by OLS, GLS, or GMM estimation. [For a treatment of TP estimation and associated asymptotic theories, see Shanken (1985, 1992) and Jagannathan and Wang (1998).]

An alternative method used to avoid the EIV problem is the ML estimation of

Gibbons (1982). These authors express the null model in (3.4) as

$$H_0^* : a_i = \lambda_0 + \boldsymbol{\beta}_i' \boldsymbol{\lambda}_1, \; i = 1, ..., N, \tag{3.9}$$

where a_i is the individual intercept in the first-pass regression (3.3), $\boldsymbol{\lambda}_1$ is an unknown $k \times 1$ vector. Notice that the following relationships hold between the γ's and λ's [see Ahn and Gadarowski (2001, p.6)]:[2] $\lambda_0 = \gamma_0$, $\boldsymbol{\lambda}_1 = \gamma_1 - E(\mathbf{f}_t)$. Similarly, the alternative model in (3.5) can be equivalently written as

$$H_A^* : a_i = \lambda_0 + \boldsymbol{\beta}_i' \boldsymbol{\lambda}_1 + \mathbf{s}_{it}' \boldsymbol{\lambda}_2, \; i = 1, ..., N, \tag{3.10}$$

where $\lambda_0 = \gamma_0$, $\boldsymbol{\lambda}_1 = \gamma_1 - E(\mathbf{f}_t)$, $\boldsymbol{\lambda}_2 = \gamma_2$. Thus, the validity of the null H_0^* can be checked now by testing the restriction $\boldsymbol{\lambda}_2 (= \gamma_2) = 0$. Applying the minimum distance approach to (3.9) and (3.10) in terms of TP estimation, Ahn and Gadarowski (2001) have developed several robust methods to estimate $\lambda = (\lambda_0, \lambda_1', \lambda_2')'$, but also provide EIV corrected standard errors of the TP estimators, such that the validity of asset pricing models can be evaluated under a general set of assumptions.

Suppose now that we are interested solely in testing the significance of the asset characteristics, as envisaged either by (3.5) or (3.10). More specifically, under (3.10), the time series linear factor pricing regression can be extended to the following panel data regression:

$$r_{it} = \alpha_i + \boldsymbol{\delta}' \mathbf{s}_{it} + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it}, \; i = 1, ..., N, \; t = 1, ..., T, \tag{3.11}$$

where $\alpha_i = \lambda_0 + \boldsymbol{\beta}_i' \boldsymbol{\lambda}_1$ and $\boldsymbol{\delta} = \boldsymbol{\lambda}_2$. If certain asset characteristics are statistically significant for explaining excess returns, then these anomaly effects can be regarded as evidence against the underlying multi-factor models. As a natural extension to the analysis so far conducted we present a panel data-based test for the null model (3.9) against the alternative model (3.10) in the context of multi-beta pricing models. We propose this be done via a simple Wald test of $\boldsymbol{\delta} = \mathbf{0}$ in (3.11). Because this does

[2] Notice here that the factor risk premia γ_1 are now decomposed into the population mean vector of the factors $E(\mathbf{f}_t)$, and the so-called lambda component $\lambda_1 = \gamma_1 - E(\mathbf{f}_t)$. This lambda component can be interpreted as the vector of factor mean adjusted risk premia, see Zhou (1998).

not require second pass cross sectional estimation, the panel-based test will not suffer from the EIV problem discussed above. Further, the fact that we use a Wald test gives the procedure all of the desirable (asymptotic) inferential properties associated with likelihood based tests. Finally, a by product of the method is that it generates full information ML estimates of all the model's parameters under the alternative and these estimates will be fully efficient.

One possible reason why previous authors have completely ignored the potential efficiency gains associated with ML panel data estimation is that under the null, all betas are heterogeneous so that there are no homogeneous coefficients to estimate and no efficiency gains (apart from those arising from imposing the null restrictions on the intercepts) to be made from system wide ML estimation. However under the alternative as (3.11) clearly shows, a panel-based analysis becomes not only natural but desirable from the point of view of efficient estimation and inference.

We close this section with some brief comments on the panel model. First, there are two different types of regressors: the asset pricing factors, which vary over time but are constant across assets/portfolios and the asset specific characteristic variables, which vary over both time and assets/portfolios. By contrast, factor loadings $\boldsymbol{\beta}_i$, are heterogenous across portfolios whilst the parameters on characteristics, $\boldsymbol{\delta}$, are homogeneous across portfolios. Hence, the panel data model (3.11) shares common features with the econometric framework recently proposed by Pesaran, Shin and Smith (1999), who develop dynamic heterogeneous panel estimation techniques that allow the simultaneous investigation of both homogenous long-run relationships and heterogenous short-run dynamic adjustment towards that long run relationship. Though similar in spirit, the exact econometric methodology developed and used in this study is different from that of Pesaran, Shin and Smith (1999). Hence we must develop the underlying econometric theory for estimation and inference using (3.11) anew. This is achieved in the next section.

3.3.1 Heterogeneous Panel Data Methodology

In this section we formally develop the econometric theory underlying the panel data model. To this end it will be convenient to generalize notation. Explicitly, we consider the heterogeneous panel regression model,

$$y_{it} = \boldsymbol{\delta}'\mathbf{x}_{it} + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{3.12}$$

with error components,

$$u_{it} = \alpha_i + \varepsilon_{it}, \tag{3.13}$$

where y_{it} is a scalar dependent variable, $\mathbf{x}_{it}$ is a q vector of explanatory variables, $\mathbf{f}_t$ is the k vector of common factors, α_i contains individual effects, and ε_{it}'s are independently distributed (over time and cross-section) with mean zero and heterogeneous variance, σ_i^2. We assume that α_i are identically and independently distributed with zero mean and variance σ_α^2, and that α_i are uncorrelated with ε_{jt} for all i, j and t.

In this panel data model, some parameters ($\boldsymbol{\beta}_i$) are allowed to be heterogenous, but others ($\boldsymbol{\delta}$) are homogenous. Under the assumption that ε_{it} are normally distributed with heterogenous variances, σ_i^2, we obtain the following (concentrated) log-likelihood function:[3]

$$\ell_T(\varphi) = -\frac{T}{2}\sum_{i=1}^{N} \ln 2\pi\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\delta}\right)'\mathbf{H}_i\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\delta}\right), \tag{3.14}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{x}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$, $\mathbf{H}_i = \mathbf{I}_T - \mathbf{W}_i(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'$, $\mathbf{I}_T$ is an identity matrix of order T, $\mathbf{W}_i = (i_T, \mathbf{f})$ with $i_T = (1, ..., 1)'$ and $\mathbf{f} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$, and $\varphi = \left(\boldsymbol{\delta}', \sigma_i^2, ..., \sigma_i^2\right)'$.

The maximum likelihood estimator of the homogeneous parameters $\boldsymbol{\delta}$ can be obtained by maximizing (3.14) with respect to $\left(\boldsymbol{\delta}, \sigma_1^2, ..., \sigma_N^2\right)$, respectively. It is then straightforward to obtain the following formula for $\hat{\boldsymbol{\delta}}$, and $\hat{\sigma}_i^2$:

$$\begin{aligned} \hat{\boldsymbol{\delta}} &= \left(\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\mathbf{x}_i'\mathbf{H}_i\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\mathbf{x}_i'\mathbf{H}_i\mathbf{y}_i\right), && (3.15) \\ \hat{\sigma}_i^2 &= T^{-1}\left(\mathbf{y}_i - \mathbf{x}_i\hat{\boldsymbol{\delta}}\right)'\mathbf{H}_i\left(\mathbf{y}_i - \mathbf{x}_i\hat{\boldsymbol{\delta}}\right),\ i = 1, ..., N. && (3.16) \end{aligned}$$

These need to be solved iteratively. Starting with an initial estimate of $\boldsymbol{\delta}$, say $\hat{\boldsymbol{\delta}}^{(0)}$, estimates of σ_i^2 can be computed using (3.16), which can then be substituted in (3.15)

[3] Normality can be relaxed in which case a quasi-ML approach would be invoked.

to obtain a new estimate of $\boldsymbol{\delta}$, say $\hat{\boldsymbol{\delta}}^{(1)}$, and so on until convergence is achieved. Alternatively, these estimators can be computed by the familiar Newton-Raphson algorithm which makes use of both first and the second derivatives. Once the converged $\hat{\delta}$ is obtained, the final OLS estimates of β_i and a_i are obtained from the following individual regression:

$$y_{it}^{*} = \boldsymbol{\beta}_i'\mathbf{f}_t + a_i + \varepsilon_{it},\ t = 1, ..., T. \tag{3.17}$$

where $y_{it}^{*} = y_{it} - \hat{\delta}' x_{it}$. Next, subject to the homogeneous restriction the estimates of σ_i^2 are obtained by

$$\tilde{\sigma}_i^{2\,(1)} = \frac{1}{T - K^{*}} \sum_{t=1}^{T} \tilde{\varepsilon}_{it}^2$$

where K^{*} is the number of parameters in regression (3.17) and

$$\tilde{\varepsilon}_{it}^2 = y_{it}^{*} - \tilde{\boldsymbol{\beta}}_i'\mathbf{f}_t + \hat{a}_i,\ t = 1, ..., T.$$

In order to derive the asymptotic distribution of the pooled ML estimators of φ, we assume that all the underlying variables are stationary, in which case under fairly standard conditions the consistency and the asymptotic normality of the pooled ML and mean group estimators (see below) of the parameters in (3.14) can be easily established. In particular, as both $T \to \infty$ and $N \to \infty$, the pooled ML estimator of $\boldsymbol{\delta}$ has the following asymptotic distribution:

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) \overset{a}{\sim} N\left\{\mathbf{0}, \left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\mathbf{Q}_{x_i x_i}\right]^{-1}\right\}, \tag{3.18}$$

where $\mathbf{Q}_{x_i x_i}$ are the probability limits of $T^{-1}\mathbf{x}_i'\mathbf{H}_i\mathbf{x}_i$.[4] The proof can be easily established using the results in Pesaran and Smith (1995) and Pesaran, Shin and Smith (1999).

Using these results, the joint null hypothesis $\boldsymbol{\delta} = 0$ can be tested simply by a Wald

[4]For this result to hold it is necessary that the limit of $N^{-1}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\mathbf{Q}_{x_i x_i}$ as $N \to \infty$ is a positive definite matrix.

statistic given by

$$Wald = \hat{\boldsymbol{\delta}}' \left[Var\left(\hat{\delta}\right)\right]^{-1} \hat{\boldsymbol{\delta}} = \left(\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \mathbf{x}_i' \mathbf{H}_i \mathbf{y}_i\right)' \left(\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \mathbf{x}_i' \mathbf{H}_i \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \mathbf{x}_i' \mathbf{H}_i \mathbf{y}_i\right), \tag{3.19}$$

where $\hat{\sigma}_i^2$ is the final consistent estimate of σ_i^2 . Then, under the null, we have $Wald \overset{a}{\sim} \chi_q^2$, where q is the dimension of $\mathbf{x}_{it}$. As a special case the single null of $\delta_i = 0$, $i = 1, 2, ..., q$, can be tested using the t-test given by

$$t = \frac{\hat{\delta}_i}{\sqrt{Var(\hat{\delta}_i)}}, \tag{3.20}$$

where $\hat{\delta}_i$ is an ith element of $\hat{\boldsymbol{\delta}}$, which converges to the standard normal distribution under the null.

3.4 Empirical application to UK Data

Several other authors have already discovered "anomalous" size and book-to-market effects in UK data. See for example Levis (1985, 1989), Strong and Xu (1997) and Hussain, Diacon and Toms (2000). In this section we address the issue of testing for factor price mis-specification and apply the traditional two pass regression method, the Fama and French (1993) time series procedures and the panel data approach above mentioned to a sample of UK stock returns. In particular we focus on the significance and importance of size and book-to-market effects in explaining returns within the context of a single factor or CAPM model and of a three factor model.

3.4.1 Data Description

The data consist of 408 monthly observations from July 1968 to June 2002 on 5603 UK firms quoted in the London Stock Exchange (LSE). The sample is comprehensive and also includes all dead firms therefore there are no problems related to survival bias. Stock market returns for the Financial Times All Share Index and for the individual companies are obtained by transforming the associated (monthly) return indices from

Datastream[5] into the monthly percentage returns (for example, the return index, RI, is the growth in value of a share holding over a month, assuming that dividends are re-invested to purchase additional units of an equity at the closing price applicable on the ex-dividend date). The unweighted excess return on individual firms, r, is obtained as the difference between the monthly return on individual firms above described and the monthly return on the risk-free asset which we take to be a 3-Month UK Treasury Bill. The excess return on the market portfolio, denoted r^m, is obtained as the difference between the monthly stock market return above mentioned and the monthly return on the risk-free asset. Book-to-Market value (BTM) and Market value (MV) are as given by Datastream, and proxy the firm characteristics of financial distress and size, respectively. BTM is defined as the ratio between book value and market value, where book values, measured in millions of pounds, are defined as net tangible assets, excluding intangible assets, less total liabilities, minority interest and preference stock. MV, measured also in millions of pounds, is the share price multiplied by the number of ordinary shares in issue.

Following Fama and French (1996), we construct the two additional factors meant to mimic the underlying risk factors in returns related to size and Book-to-Market. In order to do so we build six portfolios. At the end of each month from July 1968 to June 2002 all the LSE stocks are ranked on MV and BTM, independently. The median of MV value is used to split the sample into two groups, small and big. LSE stocks are also split into three BTM equity groups based on the break points for the bottom 30%, middle 40%, and top 30% of the ranked values of BTM for LSE stocks. The use of three groups for BTM but only two for MV is consistent with Fama and French (1993). Our portfolio SMB (small minus big), meant to mimic the risk factor in returns related to size, is the each month difference between the simple average

[5] *Datastream* turns out to be a rather imprecise source of information. For instance, *Datastream* posts zero returns for firms leading up to their official death dates. In facts those firms might have died before and suspended trading (zero returns) before they died, including the zeros will cause distressed firms (those that go bankrupt) to have lower average returns than would be true. Also, when firms die there is no final payoff recorded. Furthermore, monthly returns occasionally exceed 10000%, they may be stock consolidations missed by *Datastream* but this has not been confirmed. Notice that those extraordinarily big returns are driven by recorded price jumps and therefore also size jumps to very high levels. Those outliers have not been deleted from our sample. We acknowledge those imperfections and also report that to our knowledge there is no feasible alternative as other sources do not contain all the information needed.

of returns on the small-stock portfolios (S) and the simple average of returns on the big-stock portfolios (B). This difference should be largely free of the influence of BTM equity, focusing instead on the different return behavior of small and big stocks. The portfolio HML (high minus big) meant to mimic the risk factor in returns related to Book-to-Market equity is similarly constructed: HML is the monthly difference between the simple average of returns on the high BTM portfolios (H) and the average of returns on the low BTM portfolios (L). The difference between the two returns should be largely free of the size factor return, focusing instead on the different return behavior of high and low Book-to-Market equity. The evidence of the success of this procedure is that the correlation from 1968 to 2002 monthly returns for the size and Book-to-Market factors is only 0.06. True mimicking portfolios for the common risk factors in returns minimize the variance of firm-specific factors. The six size-Book-to-Market portfolios in SMB and HML are value-weighted. Using value-weighted components is in the spirit of minimizing the variance, since return variances are negatively related to size. Notice also that on average only 59 firms (out of 5603) per year have negative book equity. There are no firms with negative book equity till 1972 and they are very rare before 1985. The negative book equity firms are mostly concentrated in the last 12 years of the sample 1990-2002, and we do not include them in the test. Table 3.7 presents annual returns for r^m, r, SMB and HML and firm characteristics (MV and BTM) for the thirty-five years but also provides the mean, the standard error and the number of negative values over the full sample. Also Figures 3.1 and 3.2 represent the patterns of r^m, r, SMB and HML. Excess return and excess market returns are clearly correlated and their cyclical pattern follows the main events that hit the UK economy during the last thirty-five years. The mean of r^m is 4.9% and might represent a proxy for the average equity risk premium over a relatively long time period. It is higher than the 3.90% figure derived from a shorter period using UK data (Hussain, Diacon and Toms, 2000). Excess return and excess market returns reflect the recession between 1973 and 1974. Generally the '80s witnessed a rapid expansion of equity markets worldwide, which was accompanied by a particularly strong increase in international equity trading. As reported in the Bank of England Quarterly Bulletin (1988), the exceptionally rapid growth in international equity trading over the period can be attributed to an attempt to reduce risk through international portfolio diversification and the pursuit of international arbitrage opportunities emerging in the context of changing macroeconomic, tax and regulatory environments. The expansion of international equities has been associated with a greater interdependence of national stock markets and greater international interdependence of stock market ought to result in greater market efficiency. To the extent that funds can flow freely into markets in which assets are undervalued and out of markets where they are overvalued, prices tend to be based on more uniform risk-return criteria, enabling funds to be channelled into their most productive uses. On the other hand, the events of October 1987 interrupted the trend towards expansion. The negative values of r^m and r show the effect of the recession that affected the UK economy in the second half of 1990 and in 1991. Following the recession in the early '90s three other major events negatively influence $r^m$and r: namely, the Mexican crisis and political uncertainty in Europe between 1994 to 1995, the Asian crisis between 1997 and 1998 and the global slowdown starting from early 2001.

We now turn to SMB and HML. The SMB appears to be positive mainly for the first half of the study and negative for the second half of the study. In total, there are sixteen negative years out of thirty-five for SMB implying that small firms only outperformed large firms in nineteen out of thirty-five years. Clearly, there is evidence of a cyclical pattern in this factor. On the other hand, the HML variable appears to be mainly positive with only six negative years out of thirty-five and with these six years falling in the last two decades. This implies that high BTM firms outperformed low Book-to-Market firms for 83% of the sample's time span.

In order to analyze in depth the relation between average returns and βs we conduct the same steps of the Fama and French (1992) analysis. First we estimate "post-ranking" βs. In each month t, we rank firms by market value of equity and build 100 size pre-ranking betas. We classify firms into ten portfolios based on market value. For each firm in each market value portfolio for month t we estimate ("pre-ranking") betas using returns for the firm over the five-year period ending at time t. Firms in each size portfolio are then sorted by pre-ranking betas and classified in ten portfolios, resulting in 100 market value-β portfolios. We then calculate post-ranking betas using the complete set of (post-ranking) return observations.

Table 3.7. Annual returns and firm's characteristics

Table 3.7(a). Annual returns

Date	r^m	r	Date	r^m	r
1968	15.4	36.1	1987	1.1	28.1
1969	-16.3	-32.2	1988	8.2	6.3
1970	-4.5	16.3	1989	13.8	11.2
1971	26.2	71.9	1990	-16.3	-49.1
1972	14.9	39.2	1991	5.1	6.2
1973	-34.3	-71.1	1992	8.6	7.8
1974	-68.2	-109.7	1993	15.6	56.9
1975	84.4	144.7	1994	-4.5	-8.7
1976	-16.2	9.2	1995	12.9	19.1
1977	39.4	88.4	1996	7.7	17.1
1978	4.8	24.4	1997	14.1	21.3
1979	-3.1	-1.4	1998	3.5	-9.7
1980	14.6	7.2	1999	16.3	53.8
1981	-0.7	12.2	2000	-6.1	8.3
1982	13.8	19.7	2001	-14.1	-24.6
1983	10.8	37.8	2002	-9.9	-15.2
1984	14.3	28.8	Negative	12	10
1985	8.7	15.4	Mean	4.9	13.2
1986	10.8	34.2	Std Dev	6.2	10.6

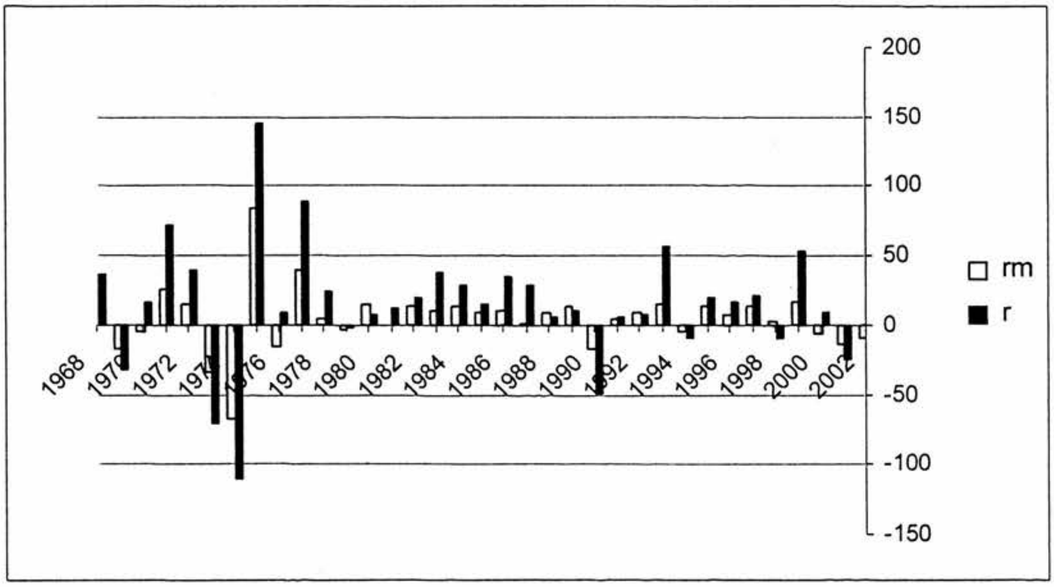

Figure 3.1 Annual returns

Table 3.7(b). Annual factor returns and firm's characteristics

Date	HML	SMB	BTM	MV	Date	HML	SMB	BTM	MV
1968	13.8	-3.7	0.74	39.8	1987	12.9	23.1	0.62	235.4
1969	2.9	3.1	0.84	34.7	1988	11.1	4.85	0.66	220.5
1970	5.1	3.2	0.97	30.3	1989	-4.2	-22.7	0.65	262.5
1971	7.8	21.4	1.04	29.2	1990	5.1	-21.5	0.82	272.5
1972	13.2	23.9	0.79	29.2	1991	-17.3	-2.62	0.95	315.1
1973	13.7	13.1	0.83	26.1	1992	7.6	-21.8	1.1	348.5
1974	31.7	1.4	1.58	15.9	1993	34.7	17.1	0.88	423.5
1975	27.9	-18.7	1.99	17.9	1994	5.4	9.75	0.67	463.7
1976	11.7	4.4	1.76	23.3	1995	-0.003	-6.42	0.71	461.1
1977	22.3	26.1	1.71	29.1	1996	11.1	-18.4	0.69	492.8
1978	7.1	19.8	1.42	38.1	1997	8.4	-14.1	0.65	545.7
1979	1.4	-0.8	1.38	44.7	1998	-11.2	-19.1	0.71	658.8
1980	-8.4	-21.2	1.65	49.4	1999	24.7	24.8	0.79	829.3
1981	18.1	10.1	1.72	57.6	2000	37.3	7.2	0.73	947.4
1982	-8.1	-7.6	1.67	63.5	2001	28.7	-4.6	0.77	856.3
1983	18.9	7.7	1.45	80.9	2002	23.2	-3.1	0.89	821.1
1984	8.8	-1.9	1.18	94.9	Negative	6	16	0	0
1985	11.7	2.6	1.03	128.1	Mean	11.1	2.1	1.1	256.5
1986	16.7	11.6	0.86	163.4	Std Dev	3.2	3.6	0.41	275.1

Notes: Annual average of returns and firm's characteristics. Negative is number of negative annual returns.

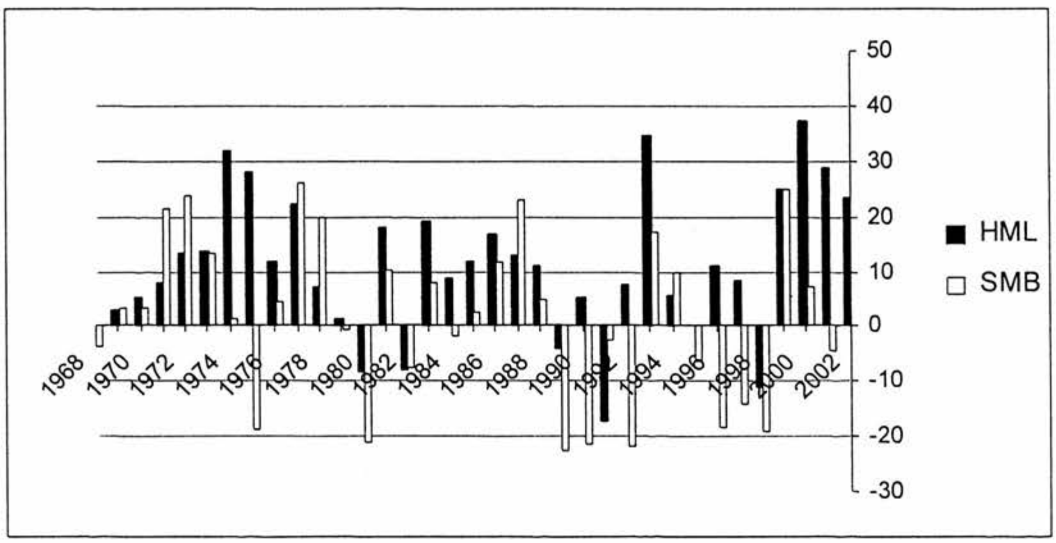

Figure 3.2. Annual factor returns

Tables 3.8 and 3.9 give various proprieties of portfolios formed on market value and pre-ranking beta, alone and in combination. Both tables report full-period post-ranking portfolio beta, estimated using monthly returns. Table 3.8 gives the portfolio values of average return, post-ranking β, market value and book-to-market for portfolios sorted according to (i) pre-ranking β and (ii) market value. In Table 3.8, when portfolios are formed on size alone, we observe the familiar negative relation between size and average return and a strong positive relation between average return and book-to-market. Contrarily to the findings presented in Fama and French (1992) here there is no evidence of a positive relation between average return and β. Average returns fall from 1.51% per month for the smallest size portfolio to 0.52% for the largest. Post-ranking β are flat or show a slight tendency to increase. Thus, a simple size sort seems to contradict the SBL prediction of a positive relation between average return and βs.

However, focusing on pre-ranking β, there is a clear positive relation between average return and post-ranking β. Average returns increase from 0.18% to 1.48% and post-ranking βs increase from 0.13% to 1.21%. The second row of Table 3.8 shows the correspondence of between the relative magnitudes of pre-ranking and post-ranking βs. There is a wider range of βs than the portfolios formed on size and the ranking is always preserved.

Table 3.9 gives proprieties of portfolios formed according to two-dimensional sorts. Panel A shows the average returns on 100 portfolios formed first on market value and then on pre-ranking beta. While less strong than the US evidence of Fama and French (1992), smaller market value portfolios generally produce higher average returns. Across β portfolios there is more consistency. Controlling for market value leaves a positive relation between β and average return: higher β portfolios produce higher average returns. The post-ranking βs and average return in Table 3.9 show a tendency to increase in each size decile. In contrast, within the columns of the average returns and βs matrices, average returns decrease while βs are rather flat and slightly increase with increasing size. The two-pass sort on size and β in Table 3.9 says that there is a negative relationship between size and average return and, when controlling for size, there is a positive relation between β and average returns.

Finally, we follow the same basic methodology of Fama and French (1993) and build another 25 portfolios formed on size and Book-to-Market equity. At the end of each month from July 1968 to June 2002 all the LSE stocks are ranked on market value and on Book-to-Market independently and then split into five size and five book-to-market groups. We construct 25 portfolios from the intersections of the size and Book-to-Market quintiles and calculate the value-weighted monthly excess returns on the portfolios. We also calculate Book-to-Market value and Market value for the 25 portfolios as the simple average of Book-to-Market value and Market value within each portfolio. The 25 portfolios formed on size and book-to-market equity produce a wide range of average excess returns from 0.51 % to 2.73 % (see Table 3.10). The portfolios confirm our previous results and those presented in Fama and French (1992) and Fama and French (1993): there is evidence of a negative relation between size and average return and a positive relation between average return and Book-to-Market equity. In all the BTM quintiles average returns tend to decrease with portfolio firm size. With the exception of a slight dip in the second BTM quintile, average returns increase with BTM for any given size quintile. Finally, taken together, the five portfolios in the largest size quintile account for on average about 56 % of the total value. The portfolio in both the largest size and the second BTM quintile alone accounts for more than 26 % of the combined value of the 25 portfolios. This finding slightly differs from Fama and French (1993) where the portfolio in both the largest size and the lowest BTM quintile contains the highest concentration of market values.

Table 3.8. Descriptive statistics for 100 stock portfolios formed on size or pre-ranking β

	1	2	3	4	5	6	7	8	9	10
Panel A: Portfolios Formed on Size										
Return[1]	1.57	0.84	0.64	0.55	0.5	0.56	0.53	0.49	0.52	0.52
β^2	0.55	0.6	0.6	0.67	0.69	0.69	0.72	0.72	0.66	0.54
$\ln(MV^3)$	0.03	1.02	1.6	2.1	2.59	3.1	3.6	4.3	5.1	6.8
$\ln(BTM)$	0.32	0.05	-0.07	-0.16	-0.22	-0.33	-0.43	-0.5	-0.58	-0.59
Panel B: Portfolios Formed on Pre-Ranking β										
Return	0.18	0.25	0.37	0.52	0.59	0.68	0.74	0.93	1.01	1.48
β	0.13	0.33	0.44	0.52	0.59	0.69	0.76	0.83	0.97	1.21
$\ln(MV)$	3.1	3.03	3.03	3.03	3.04	3.03	3.04	3.03	3.03	3.0
$\ln(BTM)$	-0.28	-0.22	-0.21	-0.22	-0.26	-0.24	-0.21	-0.23	-0.27	-0.37

Notes: [1]Return is the time series average of the monthly portfolios returns, in percentage.[2]β the time series average of the monthly portfolio βs. [3] $\ln(MV)$ and $\ln(BTM)$ the time series averages of the monthly average values of these variables in each portfolios.

Table 3.9. Descriptive statistics for 100 stock portfolios formed on size and pre-ranking β

		Pre-ranking βs									
		Low	2	3	4	5	6	7	8	9	High
	Panel A: Average Monthly Returns										
	Small	1.32	0.32	0.36	0.97	1.26	1.49	1.49	1.65	2.49	4.34
	2	0.25	0.1	0.29	0.4	0.78	0.93	0.84	1.09	1.23	2.44
	3	0.39	-0.04	0.18	0.54	0.53	0.39	0.91	0.98	1.14	1.46
	4	-0.11	0.23	0.52	0.73	0.49	0.45	0.47	0.7	0.85	1.37
Size	5	-0.11	0.24	0.3	0.35	0.34	0.57	0.56	0.89	0.65	1.32
	6	-0.08	0.5	0.3	0.53	0.37	0.81	0.71	0.96	0.84	0.81
	7	-0.1	0.02	0.54	0.42	0.68	0.62	0.64	0.71	0.94	0.82
	8	-0.12	0.33	0.42	0.35	0.4	0.65	0.7	0.93	0.48	0.72
	9	0.09	0.4	0.25	0.4	0.54	0.47	0.52	0.71	0.77	0.97
	Big	0.27	0.42	0.54	0.49	0.5	0.45	0.63	0.62	0.76	0.5
	Panel B: Post-Ranking βs										
	Small	-0.17	0.11	0.19	0.29	0.48	0.73	0.74	0.79	0.93	1.44
	2	0.017	0.22	0.3	0.45	0.56	0.68	0.68	0.83	1	1.33
	3	0.04	0.25	0.33	0.46	0.54	0.61	0.77	0.88	0.94	1.2
	4	0.09	0.35	0.44	0.52	0.61	0.75	0.78	0.86	0.98	1.31
Size	5	0.14	0.32	0.48	0.58	0.59	0.78	0.84	0.89	1.11	1.22
	6	0.18	0.42	0.59	0.57	0.63	0.74	0.8	0.84	0.96	1.21
	7	0.25	0.4	0.57	0.63	0.69	0.7	0.8	0.9	1.01	1.21
	8	0.25	0.46	0.56	0.64	0.67	0.75	0.85	0.81	1.05	1.18
	9	0.29	0.46	0.52	0.57	0.65	0.64	0.71	0.84	0.88	1.07
	Big	0.21	0.33	0.45	0.44	0.52	0.53	0.61	0.62	0.8	0.92

Notes: Time series average of the monthly portfolios returns formed on market value and pre-ranking beta.

Table 3.10. Descriptive statistics for 25 stock portfolios formed on size and Book-to-Market equity

		Book-to-Market				
		Low	2	3	4	High
	Small	Average MV^1				
	2	187.6	211.3	142.9	71.49	34.11
	2	190.4	214.1	145.7	74.26	36.88
Size	3	196.6	220.4	151.9	80.54	43.16
	4	218.2	241.9	173.5	102.1	64.74
	Big	780.1	803.8	735.4	663.9	626.6
		Average BTM				
	Small	0.96	1.12	1.25	1.43	2.01
	2	0.71	0.86	0.99	1.17	1.75
Size	3	0.61	0.76	0.89	1.07	1.65
	4	0.55	0.71	0.84	1.02	1.59
	Big	0.52	0.67	0.82	0.99	1.5
		Average r				
	Small	1.3	1.22	1.73	2.05	2.73
	2	0.81	0.72	1.23	1.55	2.23
Size	3	0.67	0.59	1.09	1.42	2.11
	4	0.66	0.58	1.09	1.42	2.11
	Big	0.58	0.51	1.01	1.34	2.03

Notes:[1] MV and BTM are measured in millions of pounds; r is measured in percentage per month.

3.4.2 Empirical results

Table 3.11 shows the time-series average of the slopes from the month-by-month Fama-MacBeth regressions of the cross-section of the stock returns on size, β and book-to-market. We run the regression at the individual security level, where βs are now calculated individually. As in Fama and French (1992) size, $\ln(MV)$, helps explain the cross-section of average stock returns. The average slope from monthly regressions of returns on size alone is -0.17%, with a t-statistic of -4.37. This negative relation persists in all the regressions, independently of the inclusion of other explanatory variables. The size effect is thus robust and confirms the results shown in Table 3.9. In contrast to the consistent explanatory power of size, the Fama and MacBeth (1973) regressions show that market β, in isolation, does not help explain average stock returns for 1968-2002. Similar findings are interpreted in Fama and French (1992) as "a shot straight to the heart of the Sharpe, Lintner, and Black model". On the other hand, in the regressions of returns on β in combination with other variables, β has a significant positive coefficient and this supports the thesis that beta commands a positive risk premium in equilibrium expected returns. Furthermore, the TP regression confirms the importance of book-to-market equity in explaining the cross-section of average stock returns. The average coefficient of $\ln(BTM)$ is 0.55% with a t-statistic of 12.5. As in Fama and French (1992), the book-to-market relation is stronger than the size effect. On the other hand, book-to-market equity does not replace size in explaining average returns: when both $\ln(MV)$ and $\ln(BTM)$ are included in the regressions, the average size slope is still significant. As a bottom line, these results indicate that unlike the simple relation between β and average return, the univariate relations between average return and size and book-to-market are strong. In the multivariate tests, the negative relation between size and average return is robust to the inclusion of the other variables. If assets are priced rationally, our results suggest that stock risks are multidimensional. Specifically, we find that size and book-to-market are characteristics associated with average returns in the 1968-2002 returns on LSE stocks. One dimension of risk is therefore proxied by size and another dimension is proxied by book-to-market.

Table 3.11. Average slopes from Month-by-Month Regressions of Stock Returns on β, MV, BTM.

α^1	β	$\ln(MV)$	$\ln(BTM)$
0.34 (1.99)	0.55 (1.75)		
1.28 (4.07)		−0.17 (−4.37)	
0.92 (3.55)			0.55 (12.52)
1.14 (4.21)		−0.11 (−2.62)	0.51 (10.3)
0.68 (3.98)	0.69 (2.21)	−0.17 (−4.58)	
0.46 (2.28)	0.69 (2.18)		0.57 (13.85)
0.62 (3.66)	0.79 (2.54)	−0.11 (−2.74)	0.53 (11.83)

Notes: The values inside (.) indicate the t-ratio; $^1\alpha$ is the intercept of the regression.

In light of these findings, we extend our analysis and apply Fama and French (1993) procedure to our sample of UK firms quoted in the LSE. Fama and French (1993) suggest that firm specific characteristics such as size and distress proxies are really picking up the effects of missing factors. They propose two additional factors SMB and HML that will destroy the significance of all of the usual characteristic variables. In this preliminary analysis, we follow Fama and French (1993) and check whether the average return spread is accounted for only by the spread in βs or also by other factors. We use excess returns on 25 portfolios, formed on size and book-to-market equity, as the dependent variable in the time series regression. The explanatory variables include the returns on market portfolio of stocks, SMB and HML. As in Fama and French (1993), the role of stock market factors in returns is developed in three steps. In the time series regressions stock returns are regressed first only on the excess market return, then on SMB and HML and finally on market return, SMB and HML. Table 3.12 shows the results of the regressions. In the first case the market leaves much variation in stock returns that might be explained by other factors. The R^2 values range from 0.45 to 0.59. When only SMB and HML are considered as explanatory variables R^2 range from 0.04 to 0.2. Finally in the case where the three stock-market factors are included in the regression R^2 improve but, contrarily to Fama and French (1993), never reach high values, implying that the three factor model is still mis-specified. In the three factor model, the factor SMB is always significant except for the big size quintile. The factor HML is always significant except for the third book-to-market quintile. Both the HML and SMB might capture shared variation in stock return that is missed by the market factor. As in Fama and French (1993), in every book-to-market quintile, the slopes on SMB decrease monotonically from smaller- to bigger-size quintiles. Similarly in every size quintile of stocks, the HML slopes increase monotonically from strong negative values for the lowest-book-to-market quintile to strong positive values for the highest-book-to-market quintile. Generally speaking, the significance of the SML and HML factors is certainly lower than in Fama and French (1993) despite adding SMB and HML to the regressions pushes the βs for stocks towards 1.0: low βs move up towards 1.0 and high βs move down.

Table 3.12. Time series regressions of excess returns on factors

Table 3.12(a). Regression of excess returns on the excess market return r^m:$r_t = a + br_t^m + e_t$

		Book to Market									
		Low	2	3	4	High	Low	2	3	4	High
				b					$t(b)$		
	Small	1.17	1.18	1.18	1.17	1.24	21.7	22.4	22.6	22.3	22.4
	2	1.26	1.27	1.27	1.26	1.33	22.4	22.9	23.1	22.7	22.9
	3	1.31	1.32	1.32	1.32	1.38	22.6	23.2	23.4	23.1	23.4
	4	1.37	1.38	1.39	1.38	1.44	22.7	23.2	23.3	23.1	23.6
	Big	1.23	1.32	1.23	1.23	1.3	18.4	23.2	18.7	18.9	20.1
Size				R^2					$s(e)$		
	Small	0.54	0.56	0.57	0.57	0.58	6.78	6.61	6.56	6.64	6.98
	2	0.55	0.56	0.57	0.57	0.58	7.11	6.9	6.93	7.03	7.31
	3	0.56	0.57	0.58	0.57	0.58	7.33	7.18	7.14	7.21	7.47
	4	0.56	0.57	0.58	0.57	0.59	7.65	7.51	7.48	7.54	7.73
	Big	0.45	0.57	0.46	0.48	0.51	8.43	7.19	8.33	8.21	8.14

Table 3.12(b). Regression of excess returns on mimicking returns for size (SMB) and book-to-market (HML) factors :$r_t = a + sSMB_t + hHML_t + e_t$

		Book to Market									
		Low	2	3	4	High	Low	2	3	4	High
				s					$t(s)$		
	Small	0.36	0.33	0.34	0.33	0.36	6.73	6.31	6.46	6.29	6.73
	2	0.37	0.35	0.36	0.34	0.37	6.49	6.11	6.24	6.08	6.53
	3	0.36	0.34	0.35	0.34	0.37	6.11	5.76	5.85	5.71	6.13
	4	0.36	0.34	0.35	0.34	0.37	5.76	5.45	5.57	5,42	5.87
	Big	0.27	0.34	0.25	0.24	0.27	4.27	5.72	3.98	3.85	4.36
				h					$t(h)$		
	Small	-0.33	-0.11	0.13	0.49	0.73	-2.26	-0.71	0.91	3.44	4.89
	2	-0.28	-0.05	0.17	0.54	0.77	-1.81	-0.36	1.14	3.5	4.88
Size	3	-0.32	-0.09	0.15	0.51	0.75	-1.94	-0.57	0.92	3.17	4.51
	4	-0.31	-0.08	0.15	0.51	0.74	-1.86	-0.51	0.86	3.01	4.29
	Big	-0.54	-0.08	-0.07	0.28	0.52		-0.52	-0.43	1.67	3.02
				R^2					$s(e)$		
	Small	0.12	0.11	0.12	0.16	0.21	9.44	9.45	9.39	9.31	9.61
	2	0.11	0.08	0.1	0.13	0.18	10.1	10.1	1005	9.97	10.2
	3	0.09	0.08	0.09	0.11	0.16	10.5	10.5	10.5	10.4	10.6
	4	0.08	0.07	0.08	0.11	0.15	11.06	11.1	11.03	10.9	11.2
	Big	0.06	0.07	0.04	0.06	0.09	11.07	10.6	11.2	11.03	11.1

Table 3.12(c). Regression of excess returns on the excess market return, r^m and the mimicking returns for size (SMB) and Book-to-Market (HML) factors: $r_t = a + br_t^m + sSMB_t + hHML_t + e_t$

		Book to Market									
		Low	2	3	4	High	Low	2	3	4	High
				b					$t(b)$		
	Small	1.12	1.13	1.13	1.12	1.17	20.8	21.1	21.1	20.8	20.9
	2	1.21	1.23	1.22	1.21	1.27	21.5	21.6	21.6	21.2	21.5
	3	1.28	1.29	1.29	1.28	1.33	22.02	22.04	22.03	21.7	21.9
	4	1.34	1.35	1.35	1.35	1.4	22.1	22.1	22.06	21.7	22.2
	Big	1.22	1.29	1.24	1.23	1.28	18.1	22.03	18.02	18.2	19.1
				s					$t(s)$		
	Small	0.13	0.12	0.13	0.13	0.17	3.33	2.99	3.41	3.44	4.14
	2	0.12	0.11	0.12	0.12	0.16	3.04	2.71	3.09	3.11	3.8
	3	0.1	0.09	0.1	0.1	0.13	2.43	2.15	2.47	2.5	3.19
	4	0.08	0.07	0.09	0.09	0.12	1.99	1.71	2.06	2.09	2.78
	Big	0.007	0.08	0.01	0.01	0.04	0.15	2.11	0.21	0.25	0.95
Size				h					$t(h)$		
	Small	-0.46	-0.23	0.001	0.36	0.59	-4.56	-2.33	0.01	3.64	5.76
	2	-0.42	-0.2	0.03	0.4	0.63	-3.97	-1.87	0.33	3.76	5.84
	3	-0.46	-0.24	-0.001	0.36	0.59	-4.22	-2.21	-0.001	3.31	5.33
	4	-0.47	-0.24	-0.001	0.35	0.58	-4.11	-2.12	-0.08	3.07	5.04
	Big	-0.68	-0.23	-0.22	0.14	0.37	-5.4	-2.14	-1.7	1.32	2.96
				R^2					$s(e)$		
	Small	0.58	0.57	0.58	0.58	0.6	6.53	6.51	6.48	6.46	6.6
	2	0.58	0.57	0.58	0.57	0.59	6.9	6.89	6.86	6.85	6.92
	3	0.59	0.58	0.58	0.58	0.6	7.14	7.11	7.1	1.07	7.15
	4	0.59	0.58	0.58	0.57	0.6	7.47	7.46	7.46	7.43	7.44
	Big	0.48	0.58	0.47	0.48	0.51	8.16	7.12	8.32	8.2	8.07

Notes: R^2 and residual standard error, $s(e)$, are adjusted for degrees of freedom.

Table 3.13 shows our suggested panel-based ML estimation and test results for anomaly effects for the single factor model and for the three factor model. Here, to accommodate the possible time varying nature of the underlying factor pricing models (including structural breaks), we also examine the three different decades, separately. The first subsample goes from 1968 to 1980, the second from 1981 to 1990 and the third from 1991 to 2002. Under the null hypothesis the Wald test examines the joint significance of the homogeneous coefficients on MV and BTM in explaining excess returns. The rejection of the null suggests that the underlying factor pricing model is mis-specified. We also re-estimate excluding each of the variables in turn and perform a t-test for the significance of the included variable's homogenous coefficient. We present both the results for the sample as a whole and the analogous results for the three selected subsamples.

Looking at the full sample estimates in the single factor (r^m) model, we see that when the model includes both MV and BTM, the value of the Wald test indicates massive significance of these terms. When only one characteristic is included in the regression, the coefficient on BTM is positive and significant whereas the coefficient on MV is significantly negative. These findings are consistent with Fama and French's (1996) argument that MV and BTM proxy for a macro"distress" factor with low BTM/MV firms being more exposed to bankruptcy risk and therefore, paying a higher return. But, when BTM and MV are jointly included in the regression, the accounting variable Book-to-Market equity has consistent explanatory power for average returns whereas the coefficient on MV becomes insignificantly positive. Strong and Xu (1997) have also obtained similar results that when BTM is included in the regression, the coefficient on MV turns out to be insignificant.

Turning to the analysis of the subperiods, we also find that the value of the Wald test for the joint significance of MV and BTM still indicates high significance of these terms for all subperiods. Next, when only one characteristic is included in the regression, the coefficients on BTM are always positive and significant whereas the coefficients on MV are always negative and significant. When both BTM and MV are included in the regression, the coefficient on MV remains negative but becomes significant only for the second subsample of '80s. On the other hand, the accounting variable Book-to-Market equity has consistent explanatory power for average returns with t-statistics in the range 2.7 to 7.5.

Table 3.13 also summarizes the results for the tests of anomaly effects for the three-factor model. In the full sample, the coefficient on MV is insignificant and small, but it becomes significant and negative when the model excludes BTM although the value is very close to zero. In the first two subperiods the coefficients on MV turn out to be significant but for the third, its value is close to zero and insignificant. The coefficients on BTM are significant except for the first period.

More importantly, comparing the three and one factor models, on average, the magnitude of the BTM coefficients are slightly lower in the former but the signs of the coefficients are the same in both. Although the values of the Wald test are generally lower than in the single factor model, they still indicate that the variables that proxy for characteristics are on the whole highly significant. Strictly speaking, this implies that the three factor model is still mis-specified, although, in terms of fit, it is an improvement over a single factor model.

In the panel data estimation, our general findings that the inclusion of BTM tends to destroy the significance of MV [see also Strong and Xu (1997)] and that the latter remains significant on the whole, only in the 1980's subsample are interesting and deserve further comment. The high significance of MV during the 80's, might be due to the cross sectional behavior of cash flows during this decade as we now argue. Berk (1995) shows that even when firm size is irrelevant to asset pricing, *measured* size (*i.e.* MV) *will* be (spuriously) statistically negatively related to expected returns. The intuition is simple. *Ceteris paribus*, firms whose prices are high, have by definition high MV. But high price firms (again *ceteris paribus*, particularly with respect to cash flow) are those with low expected return. Hence there will be a statistically negative relationship between MV and expected returns even when intrinsic firm size is irrelevant to asset pricing and does not affect expected returns. Importantly, Berk goes on to show that the significance of this negative spurious affect decreases with the cross sectional variance of cash flow. We would argue that compared with the 70's and 90's, the 80's (with perhaps the exception of 1987) was a decade that was relatively "calm" for stock markets. Although it gives no information from the cross sectional dimension, Table 3.7 clearly shows that the value of stocks experienced large cyclical fluctuations in the 70's and 90's, but grew more or less steadily in the 80's. During periods of relative calm it is indeed quite likely that the cross sectional variation in cash flow across firms will not be as great as in volatile periods. Hence, following Berk' s theoretical results, we might expect the MV bias to have been relatively high in the 80's but relatively low in the 70's and 90's. Our empirical findings could therefore be interpreted as supportive of Berk' s view that MV is a poor proxy for priced risk and its significance is mainly attributable to spurious coefficient bias.

Table 3.13. Pooled estimation and test results for anomaly effects

	Single Factor Model[1]				Three Factor Model[2]			
	68-02	68-80	81-90	91-02	68-02	68-80	81-90	91-02
Joint[3]								
δ_{BTM}	1.86 (10.70)	2.12 (6.87)	.92 (2.77)	5.27 (7.48)	1.50 (8.97)	.11 (0.37)	.65 (1.99)	4.78 (7.25)
δ_{MV}	.0004 (1.88)	−.034 (−0.32)	−.005 (−3.99)	−.002 (−0.79)	.0001 (0.77)	−.02 (−2.89)	−.007 (−5.42)	.0004 (1.48)
$Wald$	118.2	67.4	49.8	62.2	87.5	9.12	63.8	52.4
Single[4]								
δ_{BTM}	1.74 (10.71)	2.29 (7.55)	1.63 (5.81)	5.39 (7.85)	1.45 (9.33)	0.25 (0.88)	1.61 (5.82)	4.57 (7.09)
δ_{MV}	−.0003 (−1.97)	−.048 (−4.46)	−.008 (−6.67)	−.0008 (−2.55)	−.0004 (−2.64)	−.03 (−2.99)	−.008 (7.73)	−.0001 (−0.13)

Notes: Values in (.) show the t-ratio. Results are derived from the following regressions $^1r = \alpha+\beta_{r_m}r+\delta_{BTM}BTM+\delta_{MV}MV+u$; $^2r = \alpha+\beta_{r_m}r+\beta_{smb}HML+\beta_{smb}SML+\delta_{BTM}BTM+\delta_{MV}MV+u$. In ^3BTM and MV are used jointly, in ^{4} separately.

Finally, we carry out a mean group test advanced by Pesaran and Smith (1995) and assess the "average" significance of factor betas and intercepts in the panel as a whole[6]. In particular, we test the joint null, $H_0 : \beta_i = 0,\ i = 1,...,N$ against the (one-sided) alternative hypotheses $H_1 : \beta_i > 0$ for $i = 1,...,N$, and thus construct the mean group t statistic as

$$\bar{t}_{NT}(\beta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} t_T(\beta_i), \tag{3.21}$$

where $t_T(\beta_i)$ is an individual t-test for $\beta_i = 0$. Under the null as $N, T \to \infty$ and $\frac{N}{T} \to 0$, it would be possible to show under certain additional assumptions that [see Shin and Snell (2001)], $\bar{t}_{NT}(\beta) \to_d N(0,1)$.

The test results summarized in Table 3.14 indicate that market betas remain significant on average overall, in the three sub periods and despite the introduction of our two asset specific effects (i.e. size and book-to-market distress). Also, the values of the market betas change only slightly when the characteristics proxies are added. Table 3.14 also shows the results for intercepts and factor loadings for HML and SMB. The factor loadings on HML are mostly negative and their significance decreases only slightly when characteristics are added to the regression, meaning that

[6] In some ways, the current application is an ideal environment for mean group testing because the error terms are cross sectionally independent (*i.e.* idiosyncratic) under the null, an assumption which is required by this analysis but which for many other applications may be considered too strong.

the value of the coefficients and their average significance are mostly influenced by the negative values attached to the lowest-book-to-market quintile. On the other hand, the significance of factor loadings on SMB increases when characteristics are added to the regression, and the value of the coefficient becomes negative when the whole period and the third subperiod are considered. The intercepts are always found to be significant and their significance generally increases when characteristics are included in the regressions. Interestingly, the value of the intercepts decreases in a three factor model without characteristics although it never turns out to be significantly close to zero. Generally speaking the increasing significance of the intercept and of the factor loadings for SMB in the three factor model with characteristics might compensate for the slight decrease of the significance of the characteristics showed in Table 3.13 although those effects never offset each other.

In summary our findings from the TP regression unlike the Sharpe Lintner and Black model show the importance of characteristics like size and Book-to-Market in explaining average returns either when they are considered separately or when they are both included in the regression. Those findings are only partially confirmed from the panel data analysis. Here, while BTM is mostly significant the size characteristic loses most of its significance when considered together with BTM, and it remains significant only in the second decade of analysis. Another important result that contradicts Fama and French (1992) and Strong and Xu (1997) is that in the TP estimations the market beta helps in explaining average returns even when characteristics like size and Book-to-Market are added to the regression, implying that the single factor model is not sufficient to explain average returns. In fact, our empirical findings from the time series analysis enhance this belief. Here, there is evidence that the three factor model, in terms of fit, is an improvement over the single factor but is still mis-specified. The panel data estimations also support this argument. Indeed, the key aspects of the results obtained with the panel approach are that BTM remains significant even when size and book to market *factors* are present and that the market betas remain significant on average despite the introduction of our two asset specific size and book-to-market distress effects. Moreover, the values of the factors betas change only slightly when the characteristics proxies are added. This is in line with the results obtained with the multifactor approach but contradicts most of the empirical findings to date [particularly, Fama and French (1995)] which tend to show that characteristics such as firm size and book to market become insignificant when the standard CAPM specification is augmented by Fama and French's size and book to market factors.

Table 3.14. Mean Group estimation and test results for alphas and betas

		68-02	68-80	81-90	90-02
(i)	$\bar{\beta}_{r_m}$	1.29	1.29	1.34	1.23
	$\bar{t}_{NT}\,(\beta_{r_m}=0)$	111.7	74.2	56.5	52.0
	$\bar{\alpha}$	0.68	0.80	0.58	0.63
	$\bar{t}_{NT}\,(\bar{\alpha}=0)$	9.42	5.86	4.60	6.05
(ii)	$\bar{\beta}_{r_m}$	1.28	1.27	1.33	1.22
	$\bar{t}_{NT}\,(\beta_{r_m}=0)$	110.7	73.5	56.4	51.7
	$\bar{\alpha}$	-1.38	-0.88	0.53	-3.35
	$\bar{t}_{NT}\,(\bar{\alpha}=0)$	-19.2	-6.57	3.91	-33.4
(iii)	$\bar{\beta}_{r_m}$	1.29	1.23	1.38	1.24
	$\bar{t}_{NT}\,(\beta_{r_m}=0)$	117.5	76.4	58.8	54.9
	$\bar{\beta}_{hml}$	-.49	-1.1	-.4	-.07
	$\bar{t}_{NT}\,(\beta_{hml}=0)$	-24.7	-27.7	-9.7	-2.8
	$\bar{\beta}_{smb}$	0.59	0.96	0.67	0.57
	$\bar{t}_{NT}\,(\beta_{smb}=0)$	8.2	6.94	5.04	5.74
	$\bar{\alpha}$	0.59	0.64	0.67	0.57
	$\bar{t}_{NT}\,(\bar{\alpha}=0)$	8.20	6.94	5.04	5.75
(iv)	$\bar{\beta}_{r_m}$	1.29	1.23	1.36	1.23
	$\bar{t}_{NT}\,(\beta_{r_m})$	116.9	76.6	58.8	54.7
	$\bar{\beta}_{hml}$	-.47	-1.06	-.41	-.06
	$\bar{t}_{NT}\,(\beta_{hml}=0)$	-24.1	-27.5	-9.9	-2.6
	$\bar{\beta}_{smb}$	-1.02	1.74	1.13	-3.4
	$\bar{t}_{NT}\,(\beta_{smb}=0)$	-14.1	12.5	8.7	-34.5
	$\bar{\alpha}$	-1.02	1.75	1.14	-3.42
	$\bar{t}_{NT}\,(\bar{\alpha}=0)$	-14.1	12.5	8.65	-34.6

Notes: The Models correspond to the estimation of the following regressions (i): $r=\alpha+\beta_{r_m}r+u$; (ii): $r=\alpha+\beta_{r_m}r+\delta_{BTM}BTM+\delta_{MV}MV+u$; (iii): $r=\alpha+\beta_{r_m}r+\beta_{smb}HML+\beta_{smb}SML+u$;.and (iv): $r=\alpha+\beta_{r_m}r+\beta_{smb}HML+\beta_{smb}SML+\delta_{BTM}BTM+\delta_{MV}MV+u$. In (ii) and (iv), all estimates are computed conditional on the pooled ML estimates of δ_{HML} and δ_{BTM} shown in Table 3.13.

3.5 Conclusions

In this analysis we address the issue of testing for factor price mis-specification via TP regression methods, the Fama and French (1993) time series procedure and via a panel data approach. The first two approaches have been broadly used in the literature and are presented here as preliminary to the panel data analysis. While the benefits of using panel data techniques have been completely ignored in the literature here we propose this approach as an alternative of the conventional two path regression. We have presented a logically natural and theoretically coherent panel data framework within which to analyse asset return anomalies and derived the appropriate estimation and inference techniques within this framework together with their relevant asymptotic properties. In an empirical application, the TP regression results unlike the Sharpe Lintner and Black model show the importance of characteristics like size and Book-to-Market in explaining average returns either when they are considered separately or when they are both included in the regression. Those findings are only partially confirmed when we apply our panel data proposed approach. Here, while BTM is mostly significant the size characteristic loses most of its significance when considered together with BTM, and it remains significant only in the second decade of analysis. However, it is worth noticing that in the TP method at the first pass the whole time period has been used to calculate beta and at the second pass the cross section regression has been computed at the last period of the sample using all the cross section observations available. Hence, the difference between the results obtained with the TP method and the panel data analysis might be due to the different time periods considered in the analysis. A more accurate comparison could be done applying the TP method to the three subsample periods and then testing for the significance of the difference in the parameters as well as for parameters shifts. On the other hand, the results from the panel data analysis fully support the time series evidence that the three factor model still appears mis-specified. Indeed, the key aspects of the results obtained with the panel approach are that BTM remains significant even when size and book to market *factors* are present and that the market betas remain significant on average despite the introduction of our two asset specific size and book-to-market distress effects. This conforms with the results obtained with the multifactor approach but contradicts most of the empirical findings to date [particularly, Fama and French (1995)] which tend to show that characteristics such as firm size and book to market become insignificant when the standard CAPM specification is augmented by Fama and French's size and book to market factors.

Chapter 4

Gravity Models of the Intra-EU Trade: Application of the Hausman-Taylor Estimation in Panels with Heterogeneous Time-specific Common Factors

4.1 Introduction

The gravity model of international trade flows states that the size of trade flows between two countries is determined by supply conditions at the origin, demand conditions at the destination and stimulating or restraining forces related to the specific flows between the two countries. In particular, the gravity model seems to be well suited for the trade policy analysis and has been widely used a baseline model for estimating the impact of a variety of policy issues regarding regional trading groups, currency unions and various trade distortions, *e.g.* Bougheas, Demetriades and Morgenroth (1999), De Grauwe and Skudelny (2000), Glink and Rose (2001), Martinez-Zaroso and Nowak-Lehmann (2001) and De Sousa and Disdier (2002). Since the seminal paper by Anderson (1979), there have also been some attempts to explicitly derive the prediction of the gravity model from different structural models such as Ricardian models, Heckscher-Olin models and Increasing Returns to Scale models, *e.g.* Bergstrand (1990), Markusen and Wigle (1990) and Leamer (1992). As argued by Davis (2000), it is nowadays remarkable to observe that in the space of a little more than a decade the gravity model has gone from theoretical orphan to having several competing claims to maternity.

Recently, it is criticised that the use of conventional cross-section estimation is mis-specified since it is not able to deal with bilateral (exporter and/or importer) heterogeneity, which is extremely likely to be present in bilateral trade flows. In this regard a panel-based approach will be desired because heterogeneity issues can be modelled by including country-pair "individual" effects. Therefore, most of recent empirical studies adopt the panel data approach to the gravity model of international trade flows. In particular, Matyas (1997) argues that the correct econometric specification should be the so-called "triple index model", where time, exporter and importer effects are specified as fixed and unobservable. But, Egger and Pfaffermayr (2002) clearly demonstrate that when the Matays' triple index model is extended to include bilateral trade interaction effects, then this generalized three way specification is in fact identical to a conventional double index model with time and bilateral effects only. A number of panel estimation techniques such as the pooled OLS, the Fixed Effects Model, the Random Effects Model have been applied in various contexts. However, the assumption that unobserved individual effects are uncorrelated with all the regressors is convincingly rejected in almost all studies. Therefore, the Fixed Effects estimation is the most preferred estimation method to avoid the inconsistent estimation, *e.g.* Matyas (1997,1998), Egger (2000), Martinez-Zaroso Nowak-Lehmann (2001), Cheng and Wall (2002), Brun, Carrere, Guillaumont, and de Melo (2002), Egger and Pfaffermayr (2002), and De Sousa and Disdier (2002).

However, it is worth noting that the fixed effects approach does not allow for estimating coefficients on time invariant variables such as distance or common language dummies, though the consistent estimation of such effects are equally important in many situations. Cheng and Wall (2002) simply suggest to estimate the regression of the (estimated) individual effects on individual-specific variables by the OLS. But, this approach clearly ignores the potential correlation between individual specific variables and (unobserved) individual effects and therefore, the resulting estimates are likely to be severely biased. In order to properly address this issue we need to employ the Hausman and Taylor (1981, hereafter HT) instrumental variable estimation technique, which allows us to consistently estimate the coefficients on time invariant variables as well. In this context, recently, Brun, Carrere, Guillaumont and de Melo (2002) and De Sousa and Disdier (2002) attempt to apply the HT estimation to gravity models of international trade.

Most recent empirical studies also emphasise the importance of explicitly allowing for the time specific effects in order to capture business cycle effects or deal with the generic globalization issues. The conventional approach extends the benchmark model simply by incorporating the fixed $T-1$ time dummies in the regressions. Using this extended model the empirical investigation of the pattern of bilateral trade flows is mostly conducted by the Fixed Effects estimation along with the HT estimation, *e.g.* Matyas (1997), Matyas and Harris (1998), De Sousa and Disdier (2002) and Egger (2002).

In this analysis we follow recent developments of panel studies surrounding the common time effects, *e.g.* Ahn, Lee and Schmidt (2001), Ng and Bai (2001), Pesaran (2002) and Phillips and Sul (2002), and advance an alternative estimation framework in which we explicitly allow for the existence of observed and/or unobserved common time-specific factors and individual responses to those common factors are heterogeneous across country pairs. This approach also has an additional advantage to allow for certain degrees of cross section dependence via heterogeneous common time factors, and thus we may avoid the potential bias of the conventional uncorrected estimates. In particular, we aim to generalize the HT estimation in this extended panel data model and develop the underlying econometric theory. More importantly, we propose to employ an alternative source of instruments in addition to the conventional (internal) instruments suggested by HT; namely, some of heterogeneous common factors under the assumption that they are correlated with individual specific variables but not with unobserved individual effects.

We apply our proposed (extended) HT estimation technique along with the conventional approaches to a comprehensive analysis of the sources of bilateral trade amongst the 15 European countries over 1960-2001 using both the triple and the double indexed versions of the gravity equation, where we consider as the dependent variable the logarithm of real export in the former and the logarithm of total trade (sum of real export and real import) in the latter. In selecting the basic empirical specification we first consider the impacts of core explanatory variables: measures of economic size of trading partners such as GDP and population, and the distance. We then augment the basic specification by adding various variables such as common language, common border, free trade area and currency union membership dummies. Finally, we follow recent theoretical developments [*e.g.* Helpman (1987), Hummels and Levinsohn (1995) and Egger (2002)] and include variables measuring both similarity in relative size of trading countries and differences in relative factor endowments.

We now summarise our main empirical findings below. First, the impact of the GDP variables is always significantly positive, whereas the impact of population variables is found to be mostly insignificant. Second, the impacts of free custom union membership are all positively significant, whilst the results are mixed for the impacts of the EMU. Third, the impact of similarity in relative size of trading countries are mostly significant and positive, while the impact of differences in relative factor endowments (RLF) are somewhat ambiguous. Turning to the estimation results for the impacts of individual specific variables, the impacts of distance, common language dummy and common border dummy are mostly significantly negative, positive and positive, respectively, as expected. A notable finding is that once the correlation between the common language dummy and unobserved individual effect is accommodated by the HT estimation, there is evidence that the effects of the variables that may proxy for geographical distance, *i.e.* distance and common border dummy, might compensate each other whereas the role of cultural affinities proxied by common language dummy becomes more important in explaining the pattern of bilateral trade flows. The exception is the estimation results obtained using the conventional $T-1$ fixed dummies: The HT estimates of the impact of distance are surprisingly positive but insignificant, the impacts of common language dummy are significant but seem to be too large, and common border dummy loses its statistical significance. This observation clearly suggests the potential advantage of our proposed approach over the conventional one based on the fixed time dummies.

Furthermore, comparing the estimation results for the benchmark case without allowing for time effects and our proposed model with unobserved common time factors, we find that there are two notable differences between them. First, the impacts of RLF are found to be significant and positive in the benchmark case, but become insignificant in an extended model. Next, the impact of EMU is found to be mostly insignificant, but becomes significantly positive in the double index version of the benchmark model. First, we notice that the impact of RLF on total trade flows might not be unambiguous since the total trade flows are the sum of inter- and intra-industry trades, and RLF is positively correlated only with the intra-industry trade. Second, empirical evidence on the impact of EMU on trade flows is mixed in the literature; Rose (2000), Glink and Rose (2002) find a rather large positive effect of currency union on trade, while a number of recent studies find negative or insignificant effects on trade of a monetary union [Persson (2001), de Souza (2002) and Pakko and Wall (2002) and De Nardis and Vicarelli (2003)]. In particular, de Souza (2002) proposes an explanation that either the periods are too short after an introduction of the Euro to use the EMU dummy as an adequate proxy for monetary union membership or forward looking agents anticipate and thus discount the increase of trade associated with the EMU membership. In this regard we also expect that the impacts of EMU are yet to be significant. Therefore, we may conclude that the estimation results obtained using our proposed extended model with unobserved common time factors seem to be more sensible. This may reflect that it is also important to allow for a certain degree of cross section dependence via unobserved common time factors, otherwise the resulting estimates would be severely biased.

The plan of the Chapter is as follows: Section 2 presents an overview on gravity models of international trade flows. Section 3 develops the extended HT estimation methodology for heterogeneous panels with observed and unobserved common factors. Section 4 presents a comprehensive empirical application to the gravity model of an intra-EU trade. Section 5 concludes with further discussions.

4.2 Overview on Gravity Models of International Trade Flows

Since early 1940s, the gravity model has been applied to a wide variety of goods and factors of production moving across regional and national boundaries under differing circumstances. For example, the model has been successfully applied to explain the determinants of varying types of flows, such as migration, flows of buyers to shopping centers, recreational traffic or commuting flows and patient flows to hospitals. In particular, in the context of international trade flows, the gravity model states that the size of trade flows between two countries is determined by supply conditions at the origin, demand conditions at the destination and stimulating or restraining forces related to the trade flows between the two countries, *e.g.* Oguledo and MacPhee (1994). Empirically, the gravity model has been well suited for trade policy analysis and thus it has been widely used as a baseline for estimating the impact of a variety of policy issues regarding regional trading groups, currency unions and various trade distortions, *e.g.* Bougheas Demetriades and Morgenroth (1999), De Grauwe and Skudelny (2000), Glink and Rose (2001) and De Sousa and Disdier (2002). Core explanatory variables used to explain the volume of trade across a pair of countries in the gravity model are measures of economic size of trading partners and of the distance between them. Moreover, empirical works to date are often augmented by various variables such as common language, common border, free trade area and currency union membership dummies.

Despite its widespread empirical use, the gravity model was earlier criticized because it lacked theoretical foundations. Nowadays, it is certainly no longer true that the gravity model is without a theoretical basis. Since the seminal paper by Anderson (1979) it has been increasingly recognized that the prediction of the gravity model can be derived from different structural models such as Ricardian models, Heckscher-Olin (H-O) models and increasing returns to scale (IRS) models of the New Trade Theory. These three types of models differ by the way product specialization is obtained in equilibrium: technology differences across countries in Ricardian model, factor proportion differences in the H-O model, and increasing returns at the firm level in the IRS model, see Helpman (1987), Bergstrand (1990), Markusen and Wigle (1990), Leamer (1992) and Eaton and Kortum (2002).

Although the gravity model per se cannot be used to test the validity any of these trade theories against each other, its empirical success is mainly due to its ability to incorporate most of the empirical phenomena observed in international trade. In order to reconcile theory and empirical evidence, Evenett and Keller (1998) address the so called 'model identification' issue and try to determine which models generate gravity-like trade predictions. The H-O model predicts that the trade will be exclusively inter-industry (defined as trade in goods with different factor intensities), whereas the IRS model anticipates that trade is intra-industry. Using a cross-sectional study on a sample of almost all industrialized countries Evenett and Keller (1998) find a robust evidence that an IRS-based trade theory provides an important reason why the gravity equation fits trade flows well. This implies that volume of international trade among industrialized countries is likely to be determined mainly by the extent of product specialization and factor proportions differences, though it is also acknowledged that these findings do not rule out the possibility that Ricardian technology differences might be what is really behind intra-industry trade. See also Deardoff (1998). As highlighted by Davis (2000), it is quite remarkable to observe that in the space of a little more than a decade the gravity model has gone from a theoretical orphan to having several competing claims to maternity.

We now turn to the issue of econometric specifications in details. Most of earlier empirical studies relied upon the use of cross-section estimation techniques. We begin with the following typical gravity equation of the international trade:

$$y_{hft} = \alpha_0 + \theta_t + \boldsymbol{\beta}_{1t}'\mathbf{x}_{hft} + \boldsymbol{\beta}_{2t}'\mathbf{x}_{ht} + \boldsymbol{\beta}_{3t}'\mathbf{x}_{ft} + \boldsymbol{\beta}_{4t}'\mathbf{z}_{hf} + u_{hft}, \tag{4.1}$$

for $h = 1, ..., N$, $f = 1, ..., N$, $h \neq f$, $t = 1, ..., T$, where y_{hft} is the dependent variable (say, the volume of trade from home country h to target country f at time t), $\mathbf{x}_{hft}$ are explanatory variables with variation in all the three dimensions (say, exchange rates between local currencies), $\mathbf{x}_{ht}$, $\mathbf{x}_{ft}$ are explanatory variables with variation in h or f and t (say, GDP or population), $\mathbf{z}_{hf}$ are explanatory variables that do not vary over time but vary in h and f (say, distance), and the disturbance terms u_{hft} are assumed to be *iid* with zero mean and constant variance across all h, f, t. Then, (4.1) is estimated by the cross-section OLS for each year, where α_0 and θ_t cannot be separately identified. However, it is well-known that this cross-section OLS estimation will ignore any of heterogeneous characteristics related to bilateral trade relationship. Even though they are difficult to measure in general, heterogeneity is extremely likely to be present in bilateral trade flows across different pair of countries. For instance, a country would export different amounts of the same product to the two different countries, even if their GDPs are identical and they are equidistant from the exporter. Since the cross-section OLS estimates clearly fail to account for these heterogeneous factors, they are likely to suffer from substantial heterogeneity bias.

A panel-based approach will be more desirable in order to deal with heterogeneity issues as the effects of such determinants can be modelled by including country-pair "individual" effects and the source of heterogeneity-induced inconsistency could be avoided. Imposing $\beta_{jt} = \beta_j$ for all t and $j = 1, ...4$, and $\theta_t = 0$ in (4.1), we obtain the following pooled panel data model:

$$y_{hft} = \alpha_0 + \boldsymbol{\beta}_1' \mathbf{x}_{hft} + \boldsymbol{\beta}_2' \mathbf{x}_{ht} + \boldsymbol{\beta}_3' \mathbf{x}_{ft} + \boldsymbol{\beta}_4' \mathbf{z}_{hf} + u_{hft}. \tag{4.2}$$

The pooled OLS estimator obtained from (4.2) does not still deal with the issue of heterogeneity bias. In this regard, almost all of recent studies have adopted the fixed effects estimation approach to the gravity model of international trade flows.

Matyas (1997, 1998) claims that the gravity model based on the pooled specification (4.2) is mis-specified, which might lead to an incorrect inference as estimates of parameters may artificially be inflated or deflated. In particular, Matyas (1997) proposes that the proper econometric specification of the gravity model should be a three-way model:

$$y_{hft} = \alpha_0 + \alpha_h + \gamma_f + \theta_t + \boldsymbol{\beta}_1' \mathbf{x}_{hft} + \boldsymbol{\beta}_2' \mathbf{x}_{ht} + \boldsymbol{\beta}_3' \mathbf{x}_{ft} + \boldsymbol{\beta}_4' \mathbf{z}_{hf} + u_{hft}, \tag{4.3}$$

where one dimension is time-specific effect (θ_t), and the other two are time invariant export and import country-specific effects (α_h and γ_f), and it is assumed that these effects are unobservable and thus specified as fixed effects. Clearly, the unduly strict restrictions $\alpha_h = \gamma_f = \theta_t = 0$ for all h, f, and t are imposed in (4.2). Estimating both models (4.2) and (4.3), he finds a statistically significant evidence against restrictions, $\alpha_h = \gamma_f = \theta_t = 0$. This approach is subsequently extended by Harris and Matyas (1998), and Matyas, Konya and Harris (2000) by augmenting various regressors such as foreign currency reserves and real exchange rates.

Egger and Pfaffermayr (2002) demonstrates that when the Matays' model (4.3) is extended to include bilateral trade interaction effects such as

$$y_{hft} = \alpha_0 + \alpha_h + \gamma_f + \theta_t + \alpha_{hf} + \boldsymbol{\beta}_1'\mathbf{x}_{hft} + \boldsymbol{\beta}_2'\mathbf{x}_{ht} + \boldsymbol{\beta}_3'\mathbf{x}_{ft} + \boldsymbol{\beta}_4'\mathbf{z}_{hf} + u_{hft}, \quad (4.4)$$

then this generalized three way specification is in fact identical to a two way model with time and bilateral effects only. This implies that the Matyas' model (4.3) is also likely to be mis-specified, since it does not span the whole vector space of possible treatments of explaining variations in bilateral trade and ignoring such bilateral trade interactions may lead to biased estimation. In general, the bilateral effect accounts for any time invariant historical, geographical, political, cultural and other bilateral influences which will lead to deviations from country pair's 'normal' propensity to trade. Since most of these influence usually remain unobserved, including bilateral interaction effects is the natural way of controlling them.

Cheng and Wall (2002) also focus on the issue of heterogeneity bias and propose the following fixed effects model (FEM):

$$y_{hft} = \alpha_0 + \alpha_{hf} + \theta_t + \boldsymbol{\beta}_1'\mathbf{x}_{hft} + \boldsymbol{\beta}_2'\mathbf{x}_{ht} + \boldsymbol{\beta}_3'\mathbf{x}_{ft} + \boldsymbol{\beta}_4'\mathbf{z}_{hf} + u_{hft}. \quad (4.5)$$

It is argued that the fixed effects are a result of ignorance because we do not know which variables are responsible for heterogeneity bias in practice. Indeed, those cultural, historical and political factors are difficult to observe and measure. Thus, they suggest to allow each pair of countries to have its own dummy variable that may be correlated with both the bilateral trade and explanatory variables. The main feature that distinguishes it from Matyas' model is the inclusion of country-pair effects which are allowed to differ accordingly with the direction of trade, *i.e.* $\alpha_{hf} \neq \alpha_{fh}$. In this regard, (4.3) can be seen as a special case of (4.5), where arbitrary cross-country restrictions on the country-pair effect are imposed, *i.e.* $\alpha_{hf} = \alpha_h + \gamma_f$. Cheng and Wall (2002) also consider the two other models: the symmetric fixed effect (SFE) and the difference fixed effect model (DFE). The former specification imposes the restriction that country-pair effects are symmetric, *i.e.* $\alpha_{hf} = \alpha_{fh}$, whilst the latter model applies first differencing to (4.5) so as to eliminate the fixed effects. Based on the statistical finding that the restrictions imposed in (4.2), the symmetry restriction on the country-pair effects and those needed to obtain the DFE specification are all significantly rejected, they conclude that the FEM (4.5) will be the most robust econometric specification of the gravity model of international trade.

Further applications of the augmented gravity model (4.5) have also been considered. De Grauwe and Skudelny (2000) analyse the effect of exchange rates variability on trade flows by including the variance of monthly nominal exchange rate returns, and find that its FEM estimate is significantly negative. Glink and Rose (2002) analyse the effect of a country joining (or leaving) a currency union by adding a binary time varying dummy variable for the same currency. Based on the Hausman test results that both POLS and REM estimators are inconsistent, they suggest to use the FEM and find that the currency union has a strong positive effect on trade in the context of a large number of countries for the fifty post-war years. See also Persson (2001) and Pakko and Wall (2002) who find contradictory evidence on the impact of the currency union.

However, it is worth noting that the fixed effects approach does not allow for estimating coefficients on time invariant variables such as distance, common border or common language dummies. Although it is sometimes difficult to find an appropriate measure of economic distance and of controlling for contiguity (for example, considering Canada and the US, China and Russia, and Argentina and Chile are all equivalently contiguous pairs), it is still important to find relatively precise effects on trade flows of those variables. To address this issue, Cheng and Wall (2002) simply suggest to estimate the additional regression of the (estimated) individual effects on individual-specific variables by the OLS. See also Martinez-Zaroso and Nowak-Lehmann (2001) for a similar two-step approach to an analysis of determinants of bilateral trade flows between European Union and Mercosur countries. However, this approach clearly ignores the potential correlation between individual specific variables and (unobserved) individual effects and therefore, the resulting estimated impacts of individual specific variables are likely to be biased. In order to properly address the issue of correlation between regressors (including both time-varying and time-invariant) and unobserved individual effects we need to employ the Hausman and Taylor (1981, hereafter HT) instrumental variable estimation technique, which allows us to obtain consistent estimation of the coefficients on time invariant variables as well. In this context, Brun, Carrere, Guillaumont and de Melo (2002) attempt to apply the HT estimation by using infrastructure and population as instruments for standard trade-barrier function such as distance, common language and common border dummies, assuming that they are not correlated with individual effects.

The triple index model as given in (4.5) is not the only way of representing the panel data-based gravity model of international trade. A more conventional double index-based panel data specification have also been applied in which case explanatory variables are expressed as a combination of characteristics of trading partners, *e.g.* Egger (2001) and Glink and Rose (2002). Thus we now also consider the following double index panel data model:

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{z}_i + \alpha_i + \theta_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., N, \tag{4.6}$$

where an index i represents each country-pair hf such that $\alpha_i = \alpha_{hf} = \alpha_h + \gamma_f$ as in Cheng and Wall (2002). Notice that variables in $\mathbf{x}_{it}$ are defined as a combination of features of the countries in each pair, but importantly embrace variables, $\mathbf{x}_{hft}$ that vary in all the three dimensions, and variables, $\mathbf{x}_{ht}$ and $\mathbf{x}_{ft}$ that vary only with one partner of trade and time, respectively. Time invariant regressors such as distance, common language and common borders dummies are now included in $\mathbf{z}_i$ that coincide with $\mathbf{z}_{hf}$. Beginning with a theoretical model of monopolistic competition advanced by Dixit and Stiglitz (1977) and Krugman (1980), De Sousa and Disdier (2002) attempt to investigate the role of consumer's preferences as well as tariff and non-tariff barriers in explaining border effects on trade flows among Hungary, Romania and Slovenia, European Union (EU) and Central European Free Trade Agreement (CEFTA) countries. Using (4.6) and including the fixed time effects θ_t simply given by the $T-1$ fixed time dummies, they apply the HT estimation and aim to consistently estimate the impacts of individual country's characteristics like distance, common border or language on intra-European trade flows, where $\mathbf{x}_{it}$ include relative production, relative price and association or free trade agreement, and $\mathbf{z}_i$ comprehend distance and common border variables. In particular, they find that once these correlations are properly eliminated, the significance of the distance is strongly reduced and the coefficient of on common border becomes insignificant.

Motivated by the New Trade Theory initiated by Krugman (1979), which attempts to explain trade patterns under monopolistic competition and increasing returns, Helpman (1987) suggests that the share of intra-industry trade in bilateral trade flows should be larger for countries with similar incomes per capita or similar characteristics in general. Helpman estimates (4.1) by the cross-section OLS estimation for fourteen countries for every year from 1970 to 1981, where the share of intra-industry trade is used as the dependent variable and some combined measures of trading partners' incomes and relative country size are suggested as the regressors that are meant to proxy for size, similarity in size and difference in relative factor endowments of trading partners, and finds that there is a positive correlation between the share of intra-industry trade and similarity in income per capita. Hummels and Levinsohn (1995) extend Helpman' s analysis into a panel data framework. In similar veins, Egger (2000) attempts to explain the total volume of export (the sum of inter- and intra-industry volumes) in terms of the geographical distance between two trading partners, the relative factor endowments, the relative size of two countries (GDP) and their overall economic space. His empirical findings generally confirm the importance of allowing for both heterogeneity and correlation between explanatory variables and individual effects.

In summary we may conclude that the FEM along with the HT is the most preferred estimation technique in the analysis of gravity model of international trade, because we need to deal with heterogeneous individual effects and its correlation with both time-varying and time invariant regressors to avoid any potential biases. In next section we will generalize the HT estimation in presence of observed common time-specific factors and develop the underlying econometric theory.

4.3 The Hausman-Taylor Estimation in Heterogeneous Panels with Time-specific Common Factors

Noticing that both triple and double index versions of the gravity model of trade, (4.5) and (4.6), can be expressed as a conventional double index panel-data model, we begin with

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{z}_i + \varepsilon_{it}, \; i = 1, ..., N, \; t = 1, ..., T, \tag{4.7}$$

$$\varepsilon_{it} = \alpha_i + \theta_t + u_{it}, \tag{4.8}$$

where the error term ε_{it} is composed of three parts; namely, α_i is an individual effect that accounts for the effect of all possible time invariant determinants of trade and might be correlated with some of the explanatory variables $\mathbf{x}_{it}$ and $\mathbf{z}_i$, θ_t is the time-specific effects common to all trade pairs that is meant to correct for the impact of all the possible country-pair invariant trade determinants such as potential trend or business cycle, and u_{it} is a zero mean idiosyncratic random disturbance uncorrelated across country-pairs and over time. The conventional assumptions are that these three components are independent of each other.

We can further generalize (4.8) such that the individual responses to variations of the common time-specific effects are heterogeneous across the country pairs. This suggests that (4.8) can be modelled as

$$\varepsilon_{it} = \alpha_i + \lambda_i f_t + u_{it}, \tag{4.9}$$

where λ_i capture heterogeneous responses that trade flows between trading countries might have with respect to the time-specific common factors, f_t. It is clearly seen that the pooled or fixed effects estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in (4.7) will be less efficient without properly accommodating the error component structure given by (4.9). More importantly, in the case where some or all of the regressors in $\mathbf{x}_{it}$ and $\mathbf{z}_i$ are likely to be correlated with f_t, the uncorrected estimator will be severely biased. There is now a growing number of panel studies using (4.9) explicitly. See for example Ahn, Lee and Schmidt (2001), Ng and Bai (2001), Pesaran (2002), Phillips and Sul (2002) and Bai (2002). Additional advantage of this approach is to allow for certain degrees of cross section dependence via heterogeneous time-specific effects. To accommodate this potentially important issue, we now combine (4.7) and (4.9). Here we consider the two cases. First, we simply assume that all of the time-specific common effects are observable in which case we have

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{z}_i + \boldsymbol{\lambda}_i'\mathbf{f}_t^* + \varepsilon_{it}, \ i = 1, ..., N, \ t = 1, ..., T, \qquad (4.10) \\
\varepsilon_{it} &= \alpha_i + u_{it}, \qquad (4.11)
\end{aligned}
$$

where $\mathbf{f}_t^*$ are observed multiple time-specific factors. The distinguishing features of the above model are: first, it considers explicitly the impacts of time-specific factors $\mathbf{f}_t^*$ instead of the conventional fixed time effects for example to investigate the business cycle or the globalization issues, and secondly, it does not impose the homogeneous restriction on the coefficients on $\mathbf{f}_t^*$. Given that $\mathbf{f}_t^*$ usually measure the common macro shocks or policies, it is quite natural to expect that individual's responses will be different from each other. In both cases of observed and unobserved common time-specific effects, we follow the Pooled Correlated Common Effect (PCCE) estimation approach advanced by Pesaran (2002), and extend the model (4.10) to

$$
y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{z}_i + \boldsymbol{\lambda}_i'\mathbf{f}_t + \varepsilon_{it}, \ i = 1, ..., N, \ t = 1, ..., T, \qquad (4.12)
$$

where we assume that there is a single unobserved time-specific common effect and then $\mathbf{f}_t$ is the augmented set including $\mathbf{f}_t^*$ and the cross sectional averages of y_{it} and $\mathbf{x}_{it}$, namely $\overline{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$ and $\overline{\mathbf{x}}_t = N^{-1}\sum_{i=1}^{N} \mathbf{x}_{it}$. We note in passing that the specification (4.12) generalizes the standard gravity model such that bilateral trade flows are now explained by variables that take into account difference in factor endowments, difference in relative size and relative price effect (together with dummies that reflect trade association or resistance), time-specific common factors and country-specific variables that measure distance, border and cultural similarities.

In what follows we will work on (4.12) and (4.11) without loss of generality. Here notations are: $\mathbf{x}_{it} = (x_{1,it}, x_{2,it}, ..., x_{k,it})'$ is a $k \times 1$ vector of variables that vary over individuals and time periods, $\mathbf{z}_i = (z_{1,i}, z_{2,i}, ..., z_{g,i})'$ is a $g \times 1$ vector of individual-specific variables, $\mathbf{f}_t = (f_{1,t}, f_{2,t}, ..., f_{l,t})'$ is an $l \times 1$-vector of time-specific variables, and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)'$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_g)'$, $\boldsymbol{\lambda}_i = (\lambda_{1,i}, \lambda_{2,i}..., \lambda_{l,i})'$ are conformably defined column vectors of parameters, respectively. Finally, we follow Hausman and Taylor (1981), decompose $\mathbf{x}_{it} = (\mathbf{x}'_{1it}, \mathbf{x}'_{2it})'$ and $\mathbf{z}_i = (\mathbf{z}'_{1i}, \mathbf{z}'_{2i})'$, and rewrite (4.10) by

$$y_{it} = \boldsymbol{\beta}'_1 \mathbf{x}_{1it} + \boldsymbol{\beta}'_2 \mathbf{x}_{2it} + \boldsymbol{\gamma}'_1 \mathbf{z}_{1i} + \boldsymbol{\gamma}'_2 \mathbf{z}_{2i} + \boldsymbol{\lambda}'_i \mathbf{f}_t + \varepsilon_{it}, \tag{4.13}$$

where $\mathbf{x}_{1it}$, $\mathbf{x}_{2it}$ are k_1- and k_2-vectors, $\mathbf{z}_{1i}$, $\mathbf{z}_{2i}$ are g_1- and g_2-vectors, and $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ are conformably defined column vectors.

We now make the following assumptions:

Assumption 1. (i) $u_{it} \sim iid\left(0, \sigma_u^2\right)$. (ii) $\alpha_i \sim iid\left(\alpha, \sigma_\alpha^2\right)$. (iii) $E\left(\alpha_i u_{jt}\right) = 0$ for all i, j, t. (iv) $E\left(\mathbf{x}_{it} u_{js}\right) = \mathbf{0}$, $E\left(\mathbf{f}_t u_{is}\right) = \mathbf{0}$ and $E\left(\mathbf{z}_i u_{jt}\right) = \mathbf{0}$ for all i, j, s, t, so all the regressors are exogenous with respect to the idiosyncratic errors, u_{it}. (v) $\mathbf{x}_{1it}$, $\mathbf{z}_{1i}$ and f_t are uncorrelated with α_i for all i, t, whereas $\mathbf{x}_{2it}$ and $\mathbf{z}_{2i}$ are correlated with α_i. (vi) Both N and T are sufficiently large.

Assumption 1 is standard in the panel data literature. In particular, we need to use prior information to distinguish columns of $\mathbf{x}$ and $\mathbf{z}$ which are correlated with the individual unobservable effect, α_i and those which are not. Assumption (vi) is necessary to consistently estimate (nuisance) heterogenous parameters, $\boldsymbol{\lambda}_i$.

We now develop the estimation theory for all the parameters in (4.13), which involves the two steps. First, we rewrite (4.10) as

$$y_{it} = \alpha_i + \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\gamma}' \mathbf{z}_i + \boldsymbol{\lambda}'_i \mathbf{f}_t + u_{it}, \ i = 1, ..., N, \ t = 1, ..., T, \tag{4.14}$$

and obtain the consistent estimator of $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \mathbf{x}'_i \mathbf{M}_T \mathbf{x}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{x}'_i \mathbf{M}_T \mathbf{y}_i \right), \tag{4.15}$$

where

$$\underset{(T\times 1)}{\mathbf{y}_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}; \ \underset{(T\times 1)}{\mathbf{1}_T} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \ \underset{(T\times l)}{\mathbf{f}} = \begin{pmatrix} \mathbf{f}_1' \\ \mathbf{f}_2' \\ \vdots \\ \mathbf{f}_T' \end{pmatrix}, \ \underset{(T\times k)}{\mathbf{x}_i} = \begin{pmatrix} \mathbf{x}_{i1}' \\ \mathbf{x}_{i2}' \\ \vdots \\ \mathbf{x}_{iT}' \end{pmatrix},$$

$\mathbf{H}_T = (\mathbf{1}_T, \mathbf{f})$ is a $T \times (l+1)$ matrix and $\mathbf{M}_T = \mathbf{I}_T - \mathbf{H}_T \left(\mathbf{H}_T'\mathbf{H}_T\right)^{-1} \mathbf{H}_T'$. Next, the consistent estimators of $\boldsymbol{\lambda}_i$ can be obtained from the following regression:

$$\tilde{y}_{it} = b_i + \boldsymbol{\lambda}_i'\mathbf{f}_t + \tilde{u}_{it}, \ i = 1, ..., N, \ t = 1, ..., T, \tag{4.16}$$

where $\tilde{y}_{it} = y_{it} - \hat{\boldsymbol{\beta}}_{FE}'\mathbf{x}_{it}$ and $b_i = \alpha_i + \boldsymbol{\gamma}'\mathbf{z}_i$.

Assuming that all the underlying variables are stationary[1], in which case under fairly standard conditions, the consistency and the asymptotic normality of the FE estimator of $\boldsymbol{\beta}$ can be easily established. In the current context, as $(N,T) \to \infty$ jointly, we have

$$\sqrt{NT}\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\beta_{FE}}\right), \tag{4.17}$$

where the consistent estimator of $\boldsymbol{\Sigma}_{\beta_{FE}}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{\beta_{FE}} = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{x}_i'\mathbf{M}_T\mathbf{x}_i}{T}\right)^{-1}\hat{\sigma}_u^2, \tag{4.18}$$

where $\hat{\sigma}_u^2$ is a consistent estimator of σ_u^2 provided by

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N}\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i}{N(T-1-l)-k}, \tag{4.19}$$

and $\hat{\mathbf{u}}_i = (\hat{u}_{i1}, ..., \hat{u}_{iT})'$ with

$$\hat{u}_{it} = \tilde{y}_{it} - \hat{b}_i - \hat{\boldsymbol{\lambda}}_i'\mathbf{f}_t = y_{it} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it} - \hat{b}_i - \hat{\boldsymbol{\lambda}}_i'\mathbf{f}_t, \ i = 1, ..., N, \ t = 1, ..., T. \tag{4.20}$$

[1]Here, the assumption of stationarity might be too restrictive and has been considered mainly for convenience. Less restrictive assumptions on the stochastic proprieties of the series will be taken in consideration in further studies.

However, the above (extended) FE estimation will also wipe out any individual specific variables in $\mathbf{Z}_i$ from (4.10) or (4.13). In order to consistently estimate γ_1 and γ_2 on individual specific variables, we notice that (4.14) can be written as

$$d_{it} = \alpha_i + \gamma_1' \mathbf{z}_{1i} + \gamma_2' \mathbf{z}_{2i} + u_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{4.21}$$

where

$$d_{it} = y_{it} - \boldsymbol{\beta}' \mathbf{x}_{it} - \boldsymbol{\lambda}_i' \mathbf{f}_t,\ i = 1, ..., N,\ t = 1, ..., T. \tag{4.22}$$

Assuming that $\alpha_i \sim (\alpha, \sigma_\alpha^2)$, we rewrite (4.21) as

$$d_{it} = \alpha + \gamma_1' \mathbf{z}_{1i} + \gamma_2' \mathbf{z}_{2i} + \alpha_i^* + u_{it} = \alpha + \gamma' \mathbf{z}_i + \varepsilon_{it}^*,\ i = 1, ..., N,\ t = 1, ..., T, \tag{4.23}$$

where $\alpha_i^* \sim (0, \sigma_\alpha^2)$ and $\varepsilon_{it}^* = \alpha_i^* + u_{it}$ is a zero mean process by construction. Rewriting (4.23) in matrix notation we have

$$\mathbf{d}_i = \alpha \mathbf{1}_T + \mathbf{z}_{1i} \mathbf{1}_T \gamma_1 + \mathbf{z}_{2i} \mathbf{1}_T \gamma_2 + \boldsymbol{\varepsilon}_i^*,\ i = 1, ..., N, \tag{4.24}$$

$$\mathbf{d} = \alpha \mathbf{1}_{NT} + \mathbf{Z}_1 \gamma_1 + \mathbf{Z}_2 \gamma_2 + \boldsymbol{\varepsilon}^*, \tag{4.25}$$

where

$$\underset{(NT\times 1)}{\mathbf{d}} = \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_N \end{pmatrix};\ \underset{(NT\times 1)}{\mathbf{1}_{NT}} = \begin{pmatrix} \mathbf{1}_T \\ \mathbf{1}_T \\ \vdots \\ \mathbf{1}_T \end{pmatrix},\ \underset{(T\times g)}{\mathbf{Z}_j} = \begin{pmatrix} \mathbf{z}_{j1} \mathbf{1}_T \\ \mathbf{z}_{j2} \mathbf{1}_T \\ \vdots \\ \mathbf{z}_{jN} \mathbf{1}_T \end{pmatrix},\ j = 1, 2,\ \underset{(NT\times 1)}{\boldsymbol{\varepsilon}^*} = \begin{pmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_N^* \end{pmatrix}.$$

Replacing $\mathbf{d}$ by its consistent estimate, $\hat{\mathbf{d}} = \left\{ \hat{d}_{it}, i = 1, ..., N,\ t = 1, ..., T, \right\}$, where

$$\hat{d}_{it} = y_{it} - \hat{\boldsymbol{\beta}}' \mathbf{x}_{it} - \hat{\boldsymbol{\lambda}}_i' \mathbf{f}_t,\ i = 1, ..., N,\ t = 1, ..., T, \tag{4.26}$$

we now have

$$\hat{\mathbf{d}} = \alpha \mathbf{1}_{NT} + \mathbf{Z}_1 \gamma_1 + \mathbf{Z}_2 \gamma_2 + \boldsymbol{\varepsilon} = \mathbf{C} \boldsymbol{\delta} + \boldsymbol{\varepsilon}^*, \tag{4.27}$$

where $\mathbf{C} = (\mathbf{1}_{NT}, \mathbf{Z}_1, \mathbf{Z}_2)$ and $\boldsymbol{\delta} = (\alpha, \gamma_1', \gamma_2')'$. Here we notice that approximation errors stemming from the use of $\hat{\mathbf{d}}$ in (4.27) are (asymptotically) negligible. To deal with the nonzero correlation between $\mathbf{Z}_2$ and $\boldsymbol{\alpha}$ or $\boldsymbol{\alpha}^*$, we need to find the following $NT \times (1+g_1+h)$ matrix of instrument variables: $\mathbf{W} = (\mathbf{1}_{NT}, \mathbf{Z}_1, \mathbf{W}_2)$, where $\mathbf{W}_2$ is an $NT \times h$ matrix of instrument variables for $\mathbf{Z}_2$ with $h \geq g_2$ for identification. The advantage of the HT estimation is that the instrument variables for $\mathbf{Z}_2$ can be obtained internally, and they suggest to use $\mathbf{QX}_1$ as the instruments for $\mathbf{Z}_2$. See also Amemiya and MaCurdy (1986) and Breush, Mizon and Schimidt (1989) for additional source of instruments.

But in this analysis we suggest to use an alternative source of instruments as follow: for this we rewrite (4.14) as

$$y_{it} = b_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \cdots + \lambda_{li}f_{lt} + u_{it}, \tag{4.28}$$

where $b_i = \alpha_i + \gamma'\mathbf{z}_i$. Define

$$\theta_{jit} = \hat{\lambda}_{ji}f_{jt},\ j = 1, ..., l,\ i = 1, ..., N,\ t = 1, ..., T,$$

where $\hat{\lambda}_{ji}$ are consistent estimates of heterogenous factor loadings λ_{ji}, and similarly define the $NT \times 1$ matrix,

$$\boldsymbol{\Theta}_j = \begin{pmatrix} \mathbf{f}_j\hat{\lambda}_{j1} \\ \mathbf{f}_j\hat{\lambda}_{j2} \\ \vdots \\ \mathbf{f}_j\hat{\lambda}_{jN} \end{pmatrix},\ \underset{(T\times 1)}{\mathbf{f}_j} = \begin{pmatrix} f_{j,1} \\ f_{j,2} \\ \vdots \\ f_{j,T} \end{pmatrix},\ j = 1, ..., l.$$

We now make the following assumption:

Assumption 2. λ_{ji}, $j = 1, ..., l_1$, are correlated with $\mathbf{z}_{2i}$, but not correlated with α_i, whilst λ_{ji}, $j = l_1 + 1, ..., l$, are correlated with both $\mathbf{z}_{2i}$ and α_i.

Assumption 2 implies that some of individuals' heterogeneous responses with respect to common factors $\mathbf{f}_t$ are correlated with $\mathbf{Z}_2$, but not with individual effects. In fact, the nature and implication of this assumption is basically the same as those of Assumption 1(v). Under Assumption 1(v) and Assumption 2, we now obtain the following instrument matrix for $\mathbf{Z}_2$, $\mathbf{W}_2 = (\mathbf{QX}_1, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, ..., \boldsymbol{\Theta}_{l_1})$, where the dimension of $\mathbf{W}_2$ is $NT \times h$ with $h = k_1 + l_1$. Note that the HT method requires the means of the variables in X_1 to be uncorrelated with the effect α_i. Together with this assumption here we also require that the heterogenous factor loadings are uncorrelated with the effect. The question of whether W_2 is a legitimate set of instruments depends on what we assume about the correlation between Z_2 and the effects. If these additional instruments are valid the extended HT estimator is at least as efficient as the HT. As Breush, Mizon and Schimidt (1989) observe, for a given sample it is observable whether the use of extra instruments increases the explanatory power of the model. This naturally depends on the data set and the context, and it is therefore a suitable subject for empirical investigation.

The consistent estimator of $\boldsymbol{\delta}$ is obtained by the GLS-IV estimation. Premultiplying $\mathbf{W}'$ by (4.27), we have

$$\mathbf{W}'\hat{\mathbf{d}} = \mathbf{W}'\mathbf{C}\delta + \mathbf{W}'\varepsilon^* \tag{4.29}$$

and therefore we obtain the GLS estimator of $\boldsymbol{\delta}$ by

$$\hat{\delta}_{GLS} = \left[\mathbf{C}'\mathbf{W}\mathbf{V}^{-1}\mathbf{W}'\mathbf{C}\right]^{-1}\mathbf{C}'\mathbf{W}\mathbf{V}^{-1}\mathbf{W}'\hat{\mathbf{d}}, \tag{4.30}$$

where $\mathbf{V} = Var\left(\mathbf{W}'\varepsilon\right)$. The feasible GLS estimator is obtained by replacing $\mathbf{V}$ by its consistent estimator. For this, we first obtain an initial consistent estimation of $\hat{\boldsymbol{\delta}}$ by the OLS estimator from (4.27) and construct a consistent estimate of $\boldsymbol{\varepsilon}^*$ by

$$\hat{\varepsilon}^*_{OLS} = \hat{\mathbf{d}} - \mathbf{C}\hat{\delta}_{OLS}, \tag{4.31}$$

where $\hat{\varepsilon}^*_{OLS} = \left(\hat{\varepsilon}^{*\prime}_{OLS,1}, ..., \hat{\varepsilon}^{*\prime}_{OLS,N}\right)'$. Using this we estimate the initial consistent estimate of $\mathbf{V}$ by

$$\hat{\mathbf{V}}_{(1)} = \sum_{i=1}^{N} \mathbf{w}_i'\hat{\varepsilon}^*_{OLS,i}\hat{\varepsilon}^{*\prime}_{OLS,i}\mathbf{w}_i, \tag{4.32}$$

where $\mathbf{w}_i$ is the $T \times (1 + g_1 + h)$ instrument matrix for individual i, defined in $\mathbf{W} = (\mathbf{w}_1', ..., \mathbf{w}_N')'$, and estimate the feasible GLS (FGLS) estimator of $\boldsymbol{\delta}$ by

$$\hat{\delta}^{(1)}_{FGLS} = \left[\mathbf{C}'\mathbf{W}\hat{\mathbf{V}}^{-1}_{(1)}\mathbf{W}'\mathbf{C}\right]^{-1}\mathbf{C}'\mathbf{W}\hat{\mathbf{V}}^{-1}_{(1)}\mathbf{W}'\hat{\mathbf{d}}. \tag{4.33}$$

Next, we construct the GLS residuals by

$$\hat{\varepsilon}_{GLS} = \hat{\mathbf{d}} - \mathbf{C}\hat{\boldsymbol{\delta}}^{(1)}_{FGLS},$$

estimate $\mathbf{V}$ and $\boldsymbol{\delta}$ further by

$$\hat{\mathbf{V}}_{(2)} = \sum_{i=1}^{N} \mathbf{w}_i' \hat{\varepsilon}^{*}_{GLS,i} \hat{\varepsilon}^{*\prime}_{GLS,i} \mathbf{w}_i$$

$$\hat{\boldsymbol{\delta}}^{(2)}_{FGLS} = \left[\mathbf{C}'\mathbf{W}\hat{\mathbf{V}}^{-1}_{(2)}\mathbf{W}'\mathbf{C}\right]^{-1} \mathbf{C}'\mathbf{W}\hat{\mathbf{V}}^{-1}_{(2)}\mathbf{W}'\hat{\mathbf{d}}. \tag{4.34}$$

This iteration will be repeated until the convergence occurs, *e.g.* $\left|\hat{\boldsymbol{\delta}}^{(j)}_{FGLS} - \hat{\boldsymbol{\delta}}^{(j-1)}_{FGLS}\right| < 0.0001$, $j = 1, 2, \ldots$ Once we have obtained the final converged FGLS estimator, its covariance matrix will be computed by

$$Var\left(\hat{\boldsymbol{\delta}}_{FGLS}\right) = \left\{\left[\mathbf{C}'\mathbf{W}\hat{\mathbf{V}}^{-1}_{FGLS}\mathbf{W}'\mathbf{C}\right]^{-1}\right\}. \tag{4.35}$$

Under fairly standard conditions the consistency and the asymptotic normality of the FGLS estimator of $\boldsymbol{\delta}$ can also be easily established. When both N and T tend to infinite[2], we have

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{FGLS} - \boldsymbol{\delta}\right) \stackrel{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\delta_{FGLS}}\right), \tag{4.36}$$

where the consistent estimator of $\boldsymbol{\Sigma}_{\delta_{FGLS}}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{\delta_{FGLS}} = \left[\frac{\mathbf{C}'\mathbf{W}}{NT}\left(\frac{\hat{\mathbf{V}}_{FGLS}}{NT}\right)^{-1}\frac{\mathbf{W}'\mathbf{C}}{NT}\right]^{-1}. \tag{4.37}$$

In next section we investigate the potential usefulness of the extended HT estimation methodology developed above by applying it to the gravity model of international trade among the European countries along with the conventional approaches based on the fixed time dummies.

[2]This assumption turns out to be rather imprecise. Further studies will follow in order to specify how N and T go to infinity jointly.

4.4 Empirical Application to the Intra-EU Trade

In this section we will provide a comprehensive analysis of the determinants of bilateral trade flows amongst 15 European countries using both triple and double indexed versions of the gravity equation, (4.5) and (4.6), where we consider as the dependent variable the logarithm of real export in (4.5) and the logarithm of total trade (sum of real export and real import) in (4.6). (For detailed definition of all the variables see the Data Appendix.) In each case we consider the three different specifications.

First, the basic model specifies that bilateral export or trade only depends on the mass of the countries (measured by GDP and population) and barrier to trade (measured by distance). A high level of income in the exporting country indicates a high level of production, which increase availability of goods for exports, whereas a high level of income in importing country suggest higher imports. Therefore, we expect the positive impacts of those variables on trade flows. The effect of population is not unambiguous as disputed in the literature. Here we follow Bergstrand (1989) and interpret that a positive (negative) impact of exporter population indicates that the exports tend to be labor (capital) intensive goods, whilst a positive (negative) impact of importer population indicates that the exports tend to be necessity (luxury) goods. As noted by Baldwin (1994), however, both impacts might be negative as larger countries are sometimes regarded as self-efficient. On the other hand, the effect of transportation costs proxied by geographical distance between capital cities is certainly expected to be negative on trade flows. Notice that in the double indexed version both GDP and population are expressed as a combined measure of trading partners.

Second, we consider the augmented specification, where trade flows are also allowed to depend on variables that take into account free trade agreements and common currency union as well as time invariant dummies for common language and common border. The variable CEE is a dummy that is equal to one when both countries belong to the European Community and is expected to exert a positive impact. See also De Grauwe and Skudelny (2000), Glink and Rose (2001), Martinez-Zaroso and Lehmann (2001), Cheng and Wall (2002), De Sousa and Disdier (2002) for an analysis of the effects of regional trading blocks. We also add the time-varying dummy variable EMU which is equal to one when both trading partners adopt the same currency. The issue on the benefits of joining a common currency union has recently been getting more attention since the introduction of the Euro in 1999. Since an official motivation behind the EMU project (European Commission, 1990) is that a single currency will reduce the transaction costs of trade within member countries, the impact of EMU on trade flows is expected to be positive. But, the empirical evidence is mixed. Glink and Rose (2002) have analysed the trade data for almost all countries in the world and found evidence of a rather large positive effect of currency union on trade. Interestingly, this finding is not consistent with the earlier studies that fail to find a significant link between exchange rate stability and trade, *e.g.* Branda and Mendez (1988), Belanger, Gutierrez and Raynauld (1992) and Frankel and Wei (1993). See also a number of recent studies that find negative or insignificant effects on trade of a monetary union, *e.g.* Persson (2001) and Pakko and Wall (2002). In particular, de Souza (2002), and De Nardis and Vicarelli (2003) investigate the effect of the EMU in the euro area over the last two decades and find no significant evidence of a robust relationship between EMU and trade. The common language dummy Lan has a value equal to one when both countries speak the same official language and is meant to capture similarity in cultural and historical backgrounds between trading countries. The shared border dummy (Bor) is equal to one when the trading partners share a border, which is a proxy for geographical proximity. Obviously, both effects on bilateral trade flows are expected to be positive.

Finally in the full specification version of the gravity equation, we also aim to follow recent developments of the New Trade Theory advanced by Helpman (1987), Hummels and Levinsohn (1995) and Egger (2001, 2002) and thus add variables such as RLF and SIM. The variable RLF measures the difference in terms of relative factor endowments (proxied by per capita GDPs) between two countries and takes a minimum value of zero when there is equality in relative factor endowments. The larger is this difference, the higher is the volume of inter-industry (and the total) trade will be, and the lower the share of the intra-industry trade. The variable SIM captures the relative size of two countries in terms of GDP. This index is bounded between zero (absolute divergence in size) and 0.5 (equal country size). The larger this measure is (meaning that the more similar two countries are), the higher the share of the intra-industry trade will be. We note in passing that these variables have been considered to mainly explain trends of the intra-industry trade share. For example, Helpman (1987) finds a negative correlation between the intra-industry trade share and RLF, and a positive correlation between the intra-industry trade share and SIM, which is interpreted as supporting evidence of the theory of IRS and imperfect competition in international trades. Since our analysis aims to explain the patterns of both intra-industry trades and the total trade flows (sum of inter- and intra-industry trades), the impact of RLF might not be unambiguous on total trade flows. We also consider the impact of (logarithm of) real exchange rates (RER) between two countries, which is defined as the price of the foreign currency per the home currency unit and which is meant to capture the relative price effects. A depreciation of the home currency relative to the foreign currency (an increase in RER) should lead to more export and less import for the home country. The effect of real exchange rates on total trade flow will be positive (negative) if the export component of the total trade is significantly larger than the import component. For similar lines of studies see De Grauwe and Skudelny (2000), Matyas, Konya and Harris (2000) and Egger and Pfaffermayr (2002). Here we drop the population variables from the full specification in order to avoid collinearity as RLF is a linear combination of GDP and population.

4.4.1 Explanatory Data Analysis

The data used cover a period of 42 years (1960-2001) whereas the country sample contains all of the 15 EU member countries, namely Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain, Sweden, United Kingdom where Belgium and Luxemburg are treated as a single country, counting 182 country-pairs in the triple index version of the gravity model (4.5) and 91 country-pairs in the double index version (4.6).

Table 4.1 reports some of summary figures presented in the *Statistical Yearbook* (Eurostat, 1997) and shows that the intra-EU trade has always been a considerable part of EU' s total trade (currently it is almost two-thirds). Since 1960, the intra-EU trade share declined as a percentage of the total EU trade only three times. During the periods 1973-1975 and 1979-1981, the relative importance of the intra-EU trade fell sharply due to price increases in primary goods. As a result, the total value of the extra-EU imports went up, raising total value of extra-EU trade. Even when the internal market was introduced in 1993, the relative importance of intra-EU trade has declined. But this may be a purely statistical phenomena due to the fact that the collection of the intra-EU data has been reorganized since 1993.

In general, the intra-EU trade volumes were positively affected by the enlargement of the European Community, *e.g.* with the accession of new member states (Greece, Portugal and Spain) in the 1980s and with the German unification at the beginning of the 1990s, see *Single Market Review* (European Commission, 1997). Also, the enlargement of the EU in 1995 with Austria, Sweden and Finland has significantly increased the intra-EU trade volume: for example, the intra-EU share of total EU trade before the three new member states joined the EU was 58% in 1994, whereas it reached around 64% in 1995, see *External and intra-European Union trade: Statistical Yearbook* (Eurostat, 1996). This clearly suggests that one of main factors behind the increasing importance of intra-EU trade within the total EU trade is clearly the stronger link among member states over the last few decades.

Table 4.1 also shows that an intra-EU trade trends along with the total EU GDP. But, the fact that the trade volume between EU countries grows faster than GDP is further evidence of the increasing integration of EU market. The *Single Market Review* (European Commission, 1997) summarizes that the growth of the intra-EU trade, initiated by the programme to complete the single market implemented in the mid-1980s, leads to major changes for the European economies. The measures taken consist mainly of a liberalization of trade in products and services through the abolition of non-tariff barriers, border formalities, a liberalization of public procurement practices and the mutual recognition of technical standards.

Table 4.1 Descriptive and Summary statistics

Panel A	1960	1970	1980	1990	2000
Share of US on Extra-EU trade	16.5*	26.3*	33.8**	19**	21.9***
Share of Intra-EU on EU trade	37.2*	49.8*	50.5**	59.7**	61.7***
Share of Export on Intra-EU trade	52.4*	51.6*	51.1**	49.7**	51.2***
Panel B	60/70	70/80	80/90	90/00	
Average Growth of GDP	8.9	16.4	7.8	3.5	
Average Growth of Intra-EU trade	11.5	17.3	9.3	5.8	
Average Growth of Total EU trade	10.3	20.1	7.2	3.9	
Average Growth of Bilateral Exchange Rate	0.12	7.9	-1.4	-3.7	

Notes: Source: Trade Policy Review of the European Union: a Report by the Secretariat of the WTO, WTO (2002) and Statistical Yearbook, Eurostat (1997). '*' denotes values for EU9 (Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxemburg, Netherlands), '**' for EU12 (EU9 plus Greece, Portugal and Spain) and '***' for EU15 (EU12 plus Denmark, Sweden and United Kingdom) countries, respectively.

Table 4.1 also shows that an intra-EU trade trends along with the total EU GDP. But, the fact that the trade volume between EU countries grows faster than GDP is further evidence of the increasing integration of EU market. The *Single Market Review* (European Commission, 1997) summarizes that the growth of the intra-EU trade, initiated by the programme to complete the single market implemented in the mid-1980s, leads to major changes for the European economies. The measures taken consist mainly of a liberalization of trade in products and services through the abolition of non-tariff barriers, border formalities, a liberalization of public procurement practices and the mutual recognition of technical standards. Also included are the liberalization of factors movements and deregulation of sectors formerly subject to tight national regulation. The anticipation by economic agents of the completion of the single market caused a drive towards strong industrial restructuring at the microeconomic level, notably through merges and acquisitions by both European and non-European companies. Liberalization would also tend to lower prices through increased competition and foster a concentration of resources in more efficient use. These effect would translate into sizable welfare gains, increases in GDP, and increase competitiveness *vis-a-vis* non-member states. On the other hand, as Jacquemin and Sapir (1990) notice, the concentration of European industries might also create or foster dominant position which lead to higher domestic prices. This lowers trade barriers against imports from the rest of the world, meaning more extra-EU and less inter-EU import. Table 4.1 actually shows that in our sample the share of exports is generally higher than the share of imports within EU trades. In light of these figures we therefore expect that positive effects of an increase in real exchange rates on exports will dominate negative impacts on imports. As a result its influence on total trades is expected to be positive.

The *Single Market Review* (European Commission, 1997) further reports that the removal of barriers to the mobility of goods leads to an increase in trade flows within the Community, and most of this increase is of the intra-industry type. Intra-industry, boosted by similarity of the trading nations, may lead to cost-free adjustments, increased efficiency and welfare gains associated with variety. In contrast, inter-industry trade, traditionally associated with comparative advantages of nations, may lead to more costly adjustments, as trade and specialization move factors from contested, export-oriented industries. Figure 4.1 shows the evolution of trade in intra-EU trade between 1980 and 1994.[3] At the beginning of the 1980s the most important trade was the inter-industry type (share of 45%), but it started to decline from the mid-1980s onwards. The resulting increase in the share of intra-industry is essentially due to a trade boost in vertically differentiated products that are predominant in the largest European countries, *e.g.* Germany and France since about 1986 and the UK since 1989. This is consistent with evidence that intra-industry trade accounts for a substantial fraction of total trade among industrialized countries, see Deardoff (1984) and Evenett and Keller (1998). Molle (1997) states that contrary to what some had expected, both EU and EFTA has not produced specialization among countries along lines of traditional trade theory predicting that one country will be specialized in one good and the other in the other good on the basis of comparative advantages. In fact, at

[3]The share of intra-industry trade is measured by the traditional Grubel-Lloyd indicator, whereas inter-indusrty trade is represented by the so called 'one way trade'. The Grubel-Lloyd (1975) index is defined as $GL = 1 - \frac{X_j - M_j}{X_j + M_j}$ and measures the amount of intra-industry trade in a particular product group j. The value ranges from zero to unity representing a situation of zero and 100 percent intra-industry trade respectively. When X_j or M_j equal to zero there is no overlap of export or import so no intra-industry trade will take place. On the other hand if $X_j = M_j$, matching will be completed and $GL = 1$. Total trade is decomposed in three trade types according to their similarity in price (proxy for quality) and to overlap in trade: two-way trade in similar products (significant overlap and low price differences); two-way trade in vertically differenced products (significant overlap and high price differences); one-way trade (no or no significant overlap).

the beginning of the 1960s, it became clear that specialization occurred within sectors and consumers have benefited from the resulting increased range of products available. The more similar the demand structures of two countries are, the more intensive are the potential trade between these two countries, see also Linder (1961).

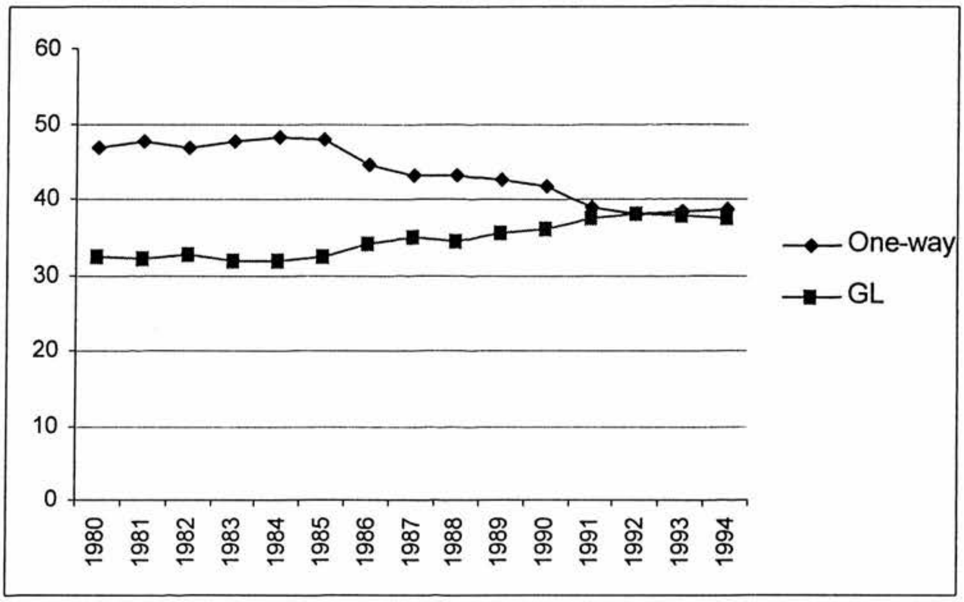

Figure 4.1. Evolution of trade in intra-EU trade 1980-1994

Notes: The Grubel-Lloyd indicator measures the share of intra-industry trade; One- way trade represents inter-industry trade. Source: *Single Market Review*, European Commission *(1997):* Trade patterns inside the single market, page 50.

4.4.2 Estimation results

We now briefly discuss alternative estimation procedures used to estimate (4.5) and (4.6): namely, the pooled OLS (POLS), the between estimation (BTW), the fixed effect model (FEM), the random effect model (REM) and Hausman and Taylor (HT) instrumental variable estimation. The POLS estimation is likely to gain in efficiency due to the increased number of observations but estimation results would be biased due to neglected (individual) heterogeneity. The between estimator runs an OLS regression on the time averages of cross section pairs, but is also likely to subject to the potential heterogeneity bias. The FEM explicitly takes into account the bilateral trade heterogeneity by specifying that all explanatory variables are assumed to be correlated with unobserved fixed individual effects, though it also wipes out any of time invariant variables. On the other hand, under the stronger assumption that unobserved individual effects are randomly distributed but uncorrelated with all regressors, the REM allows us to estimate the parameters on both time-varying and time-invariant variables, simultaneously. The validity of this assumption should be tested by using the Hausman test (1978), and when this assumption is rejected, we will use the HT estimator to consistently estimate the impacts of time-invariant variables.

We consider the two different scenarios: first, we estimate both (4.5) and (4.6) without including any time-specific effects, which we call the benchmark case. Secondly, we follow most recent empirical studies that also emphasise the importance of explicitly allowing for the time specific effects in either (4.5) or (4.6), *e.g.* Matyas (1997), Matyas and Harris (1998), De Sousa and Disdier (2002) and Egger (2002). Since we analyse the trade data over the longer time span, this issue should be addressed properly for instance in order to capture business cycle effects or deal with the generic globalization issues. We consider the three extensions: first, we extend the benchmark model by incorporating the conventional fixed time-specific dummies in the regressions. Secondly, we will use our proposed approach described in Section 3, namely by incorporating observed and unobserved single common time factor, respectively.

Tables 2 present alternative panel-data estimation results for the triple and double index gravity models of bilateral trades amongst the 15 EU countries. Since the validity of the REM assumption that there is no correlation between explanatory variables and unobserved individual effects is convincingly rejected in all cases considered, we will discuss estimation results mainly with the FEM results. For overwhelmingly similar empirical evidence see Egger (2001), Brun, Carrere, Guillaumont and de Melo (2002), Cheng and Wall (2002) and De Sousa and Disdier (2002) and Glink and Rose (2002)

Starting from the full specification of the triple index version (see Table 4.2(a)), almost all the FEM estimation results are statistically significant and consistent with our *a priori* expatiations. Both GDPs of home and foreign country have a positive effect on real exports and a depreciation of the home currency leads to an increase in export flows. Similarity in size and relative difference in factor endowments between trading partners help to boost real exports although the impact of RLF is much smaller than the impact of SIM. This finding clearly reflects the fact that the intra-industry trade is the main part of the total EU trade as described in subsection 4.1. A custom union membership also boosts real exports significantly, though the effect of EMU appears negative but insignificant. Although both REM and the POLS estimation results are likely to be biased because of correlation between regressors and unobserved individual effects, both estimation results are relatively consistent with the corresponding FEM results. Only the coefficient on EMU is positive but insignificant. Next, the BTW estimates appear to be mostly insignificant (SIM, RLF, CEE, EMU). This may be a clear indication of severe bias problem expected over the relatively long time span considered in our estimation, though we might expect to obtain different results over different time periods since the between estimator is based on a regression on time averages of cross section pairs. Turning briefly to the basic and augmented specifications, we find that only the impact of importer population is significant and negative, which leads us to conclude that the exports within EU countries are most likely to be luxury goods.

Table 4.2(b) reports the estimation results for the double index version, (4.6). Though they are mostly consistent with those of the triple index model, there are two notable differences: first, the impact of EMU on the total trade is now positive and significant. Hence, the EMU seems to have a more positive impact on imports than on exports contrary to the evidence observed after the completion of the single market. Secondly, the impact of SIM on the total trade (mostly via the impact on the intra-industry trade) is much higher (1.17 versus 0.35). Once again the effect of income variable is highly significant, whereas the impact of population is insignificant. This reinforces the previous finding in the triple index version, but may also imply that the size or the mass effect is likely to be captured mostly by income variable rather than population. (Considering that both GDP and population are proxies for the economic size of trading partners and they are highly correlated, this might indicate a certain degree of collinearity.) We also note that the magnitude of the FEM coefficient on the total GDP is somewhat larger than its OLS counterpart, a consistent finding with the previous empirical study by Matyas, Konya and Harris (2000) who argue that (correctly) allowing for heterogeneous bilateral effects is likely to increase the magnitude of the impact of GDP.[4]

One of our main purposes of the current study is an investigation of consistent estimation (and thus precise evaluation) of the impacts of individual specific variables. We consider both (inconsistent) OLS and (consistent) HT estimations and summarise such estimation results in Table 4.2(c). Here we assume *a priori* that Lan is the only time invariant variable correlated with unobserved individual effects (as common language is a proxy for cultural, historical, linguistic proximity, it is highly likely to be correlated with unobserved individual effects). We then employ two different sets of instrument variables. The first instrument set (HT1) contains only real exchange rates (RER), the second set (HT2) adds size related variables such as $GDPs$, SIM and RLF. Following de Sousa and Disdier (2002) we do not consider time-varying dummy variables that represent free trade agreement and currency union as valid instruments. As expected *a priori*, all estimation results show that distance has a negative effect on exports and trades, while common language and common border have positive effects on them. Here a notable finding is that once the correlation between Lan and unobserved individual effect is accommodated by the HT estimation, then the impacts of distance decrease (in absolute value) as compared to the OLS counterpart, whilst the impacts of both common language and common border dummies increase, especially the former. Furthermore, when we use the broad set of instruments (HT2), the distance variable loses significance. This result might be plausible given the fact that both distance and common border proxy geographical distance, the effects of which might compensate each other (the correlation coefficient between them is about 0.6). Overall, this result suggests that the role of cultural affinities will become more important in explaining the pattern of bilateral trade flows once the correlation between Lan and unobserved individual effect is appropriately handled.

[4] Most empirical studies find that estimates of the income coefficient are well over unity, see for example Matyas (1997), Egger (2000), Egger and Pfaffermayr (2000), Martinez-Zaroso Nowak-Lehmann (2001), Cheng and Wall (2002), and De Sousa and Disdier (2002).

Table 4.2(a) Alternative Panel Data Estimation Results for Triple Index Models

	Basic Model				Augmented Model				Full Model			
	POLS	BTW	FEM	REM	POLS	BTW	FEM	REM	POLS	BTW	FEM	REM
Con	-11.8 $(.241)^2$	-6.05 (1.89)		-16.5 (.59)	-12.7 (.266)	-8.4 (1.94)		-15.9 (.745)	-9.96 (.211)	-10.1 (1.24)		-15.3 (.73)
GDP_h	0.73 (.015)	0.49 (.106)	0.64 (.035)	0.78 (.03)	0.68 (.015)	0.51 (.101)	0.55 (.034)	0.67 (.029)	0.76 (.007)	0.73 (.041)	0.49 (.03)	0.73 (.021)
GDP_f	1.25 (.015)	0.99 (.106)	1.48 (.035)	1.55 (.03)	1.21 (.015)	1.03 (.101)	1.4 (.034)	1.44 (.029)	0.87 (.007)	0.85 (.041)	1.43 (.03)	1.18 (.021)
POP_h	0.01 (.019)	0.27 (.124)	0.06 (.127)	-.01 (.058)	0.04 (.019)	0.25 (.121)	-.02 (.124)	0.07 (.057)				
POP_f	-.52 (.019)	-.25 (.124)	0.72 (.127)	-.61 (.058)	-.48 (.019)	-.26 (.121)	0.64 (.124)	-.54 (.057)				
SIM									0.11 (.013)	0.04 (.071)	0.35 (.051)	0.30 (.041)
RLF									0.03 (.007)	0.02 (.05)	0.03 (.007)	0.03 (.007)
RER									0.1 (.003)	0.09 (.019)	0.08 (.007)	0.09 (.007)
CEE					0.28 (.019)	-.12 (.202)	0.28 (.013)	0.29 (.014)	0.33 (.019)	-.08 (.19)	0.31 (.014)	0.32 (.013)
EMU					0.07 (.044)	-2.21 (1.44)	-.008 (.023)	-.007 (.023)	0.17 (.043)	-1.1 (1.49)	-.004 (.023)	0.006 (.024)
Dist	-1.05 (.015)	-1.16 (.086)		-.92 (.079)	-.79 (.018)	-.92 (.109)		-.71 (.097)	-.68 (.019)	-.77 (.11)		-.57 (.096)
Lan					0.51 (.028)	0.48 (.16)		0.59 (.155)	0.25 (.029)	0.27 (.163)		0.41 (.152)
Bor					0.45 (.029)	0.52 (.167)		0.41 (.161)	0.51 (.028)	0.60 (.163)		0.44 (.156)

Notes: Here the dependent variable is logarithm of real export. POLS stands for the pooled OLS estimator, BTW the between estimator; FEM fixed effects estimator and REM random effects estimator, respectively. Figures in (.) indicate the standard error. Hausman statistic rejects the null hypothesis of no correlation between explanatory variables and unobserved individual effects in all cases considered.

Next, we consider an extended model in order to capture business cycle effects or deal with the generic globalization issues. We first follow the conventional approach and include the $T-1$ fixed dummy variables (not T dummies to avoid multicollinearity) in the corresponding regressions, (4.5) and (4.6), that are common to all country pairs. Notice here that the impacts of fixed time dummies are assumed to be homogeneous. Table 4.3 reports the related estimation results. Although most estimation results for both triple and double index specifications follow similar patterns as obtained in Table 4.2, there are a few notable discrepancies between them (mainly in the context of the FEM estimation results). First, the impact of EMU is now mostly significantly positive. Second, the impact of the GDPs seem to be somewhat too large. Third, the impact of SIM on exports is no longer significant (see Table 4.3(a)), whereas the impact of SIM on total trades is still significant and larger (see Table 4.3(b)). Finally, turning to (consistent) estimation of the impacts of individual specific variables, we find: the HT estimates of the impact of distance are surprisingly positive but insignificant, the impact of common language dummy (in the HT estimation) are much larger than in Table 4.2(c), but common border dummy loses its statistical significance.

Table 4.2(b). Alternative Panel Data Estimation Results for Double Index Models

	Basic Model				Augmented Model				Full Model			
	OLS	BTW	FEM	REM	OLS	BTW	FEM	REM	OLS	BTW	FEM	REM
Con	-7.49 (.361)	0.02 (2.64)		-16.5 (.94)	-7.9 (.39)	-2.3 (2.78)		-15.4 (1.18)	-10.9 (.247)	-9.71 (1.55)		-13.9 (.88)
GDP	1.68 (.03)	0.98 (.234)	2.2 (.031)	2.3 (.023)	1.5 (.032)	1.08 (.231)	1.94 (.031)	2.1 (.025)	1.57 (.012)	1.54 (.08)	1.81 (.019)	1.79 (.018)
POP	-.52 (.039)	0.24 (.281)	0.03 (.164)	-.83 (.099)	-.38 (.039)	0.12 (.279)	0.01 (.154)	-.74 (.095)				
SIM									0.88 (.017)	0.81 (.104)	1.17 (.055)	1.14 (.045)
RLF									0.03 (.008)	0.01 (.06)	0.03 (.008)	0.03 (.008)
RER									0.09 (.004)	0.09 (.023)	0.06 (.009)	0.07 (.008)
CEE					0.47 (.03)	0.14 (.321)	0.14 (.028)	0.36 (.016)	0.32 (.022)	-.03 (.244)	0.31 (.016)	0.32 (.016)
EMU					0.22 (.069)	0.89 (.778)	0.31 (.016)	0.14 (.028)	0.2 (.051)	-1.4 (1.84)	0.08 (.027)	0.09 (.027)
Dist	-1.2 (.022)	-1.29 (.129)		-.97 (.124)	-.93 (.028)	-1.03 (.169)		-.77 (.156)	-.64 (.022)	-.71 (.13)		-.6 (.116)
Lan					0.36 (.043)	0.24 (.255)		0.51 (.247)	0.23 (.034)	0.24 (.201)		0.41 (.185)
Bor					0.42 (.045)	0.58 (.269)		0.39 (.258)	0.52 (.034)	0.61 (.201)		0.44 (.19)

Notes: Here the dependent variable is logarithm of real export; Hausman statistic rejects the null hypothesis of no correlation between explanatory variables and unobserved individual effects in all cases considered. See also notes to Table 4.2(a).

Table 4.2(c). Hausman and Taylor Estimation Results

	Triple index model			Double index model		
	OLS	HT1	HT2	OLS	HT1	HT2
Dist	−.57 (.026)[2]	−.43 (.208)	−.34 (.208)	−.6 (.021)	−.38 (.192)	−.34 (.199)
Lan	0.45 (.041)	1.05 (.755)	1.57 (.72)	0.45 (.034)	1.56 (.707)	1.8 (.695)
Bor	0.43 (.042)	0.53 (.282)	0.61 (.289)	0.43 (.035)	0.6 (.258)	0.64 (.275)

Notes: Here we consider the full specifications and the slope coefficients are already reported as FEM estimates in Tables 4.2(a)-(b). The sets of instruments used in the HT estimation are as follows: $\{RER\}$ for HT1 and $\{RER, GDP, SIM, RLF\}$ for HT2. See also notes to Tables 4.2(a)-(b).

Table 4.3(a). Alternative Panel Data Estimation Results for Triple Index Models with Time Dummies

	Basic Model			Augmented Model			Full Model		
	OLS	FEM	REM	OLS	FEM	REM	OLS	FEM	REM
Con	−7.15 (.314)		−19.3 (.859)	−9.64 (.323)		−20.2 (.996)	−10.6 (.215)		−18.7 (.997)
GDP_h	0.51 (.018)	0.97 (.044)	0.91 (.041)	0.54 (.017)	0.91 (.043)	0.86 (.041)	0.73 (.007)	1.04 (.045)	0.84 (.029)
GDP_f	1.03 (.018)	1.82 (.044)	1.68 (.041)	1.06 (.017)	1.76 (.043)	1.62 (.041)	0.84 (.007)	1.97 (.045)	1.32 (.029)
POP_h	0.24 (.021)	0.03 (.123)	−.14 (.063)	0.21 (.021)	−.007 (.12)	−.11 (.062)			
POP_f	−.28 (.021)	0.69 (.123)	−.73 (.063)	−.32 (.021)	0.65 (.12)	−.7 (.062)			
SIM							0.07 (.012)	0.02 (.052)	0.25 (.041)
RLF							0.02 (.007)	0.02 (.007)	0.02 (.006)
RER							0.09 (.003)	0.08 (.009)	0.06 (.008)
CEE				0.11 (.022)	0.29 (.014)	0.29 (.015)	0.15 (.022)	0.33 (.015)	0.31 (.015)
EMU				0.13 (.061)	0.19 (.031)	0.23 (.032)	0.18 (.059)	0.17 (.032)	0.23 (.032)
Dist	−1.15 (.015)		−.86 (.081)	−.89 (.018)		−.61 (.099)	−.75 (.019)		−.52 (.1)
Lan				0.51 (.027)		0.64 (.156)	0.28 (.028)		0.62 (.154)
Bor				0.44 (.028)		0.41 (.162)	0.53 (.028)		0.31 (.157)

Notes: Here we augment the models in 4.2(a) by adding the time-specific fixed effects; Hausman statistic rejects the null hypothesis of no correlation between explanatory variables and unobserved individual effects in all cases considered. See also notes to Table 4.2(a).

Table 4.3(b). Alternative Panel Data Estimation Results for Double Index Models with Time Dummies

	Basic Model			Augmented Model			Full Model		
	OLS	FEM	REM	OLS	FEM	REM	OLS	FEM	REM
Con	−1.75 (.12)		−15.5 (.75)	−3.37 (.44)		−16.5 (1.45)	−10.2 (.257)		−20.2 (1.2)
GDP	1.03 (.038)	2.5 (.084)	2.3 (.077)	1.07 (.037)	2.48 (.081)	2.23 (.074)	1.53 (.013)	3.05 (.078)	2.22 (.053)
POP	0.18 (.045)	−.49 (.154)	−.96 (.111)	0.11 (.045)	−.44 (.147)	−.9 (.108)			
SIM							0.84 (.017)	1.42 (.055)	1.27 (.049)
RLF							0.02 (.008)	0.02 (.007)	0.02 (.007)
RER							0.09 (.003)	0.09 (.01)	0.06 (.009)
CEE				0.17 (.034)	0.35 (.036)	0.31 (.017)	0.17 (.026)	0.32 (.017)	0.31 (.017)
EMU				0.11 (.091)	0.32 (.017)	0.37 (.036)	0.21 (.07)	0.22 (.034)	0.28 (.035)
Dist	−1.3 (.021)		−1.01 (.125)	−1.05 (.027)		−.76 (.158)	−.69 (.022)		−.44 (.123)
Lan				0.35 (.041)		0.52 (.248)	0.26 (.034)		0.65 (.189)
Bor				0.43 (.043)		0.41 (.258)	0.54 (.033)		0.28 (.195)

Notes: Here we augment the models in 4.2(b) by adding time-specific fixed effects; Hausman statistic rejects the null hypothesis of no correlation between explanatory variables and unobserved individual effects in all cases considered. See also notes to Tables 4.2(a) and 4.2(b).

Table 4.3(c). Hausman and Taylor Estimation Results with Time Dummies

	Triple index model			Double index model		
	OLS	HT1	HT2	OLS	HT1	HT2
Dist	−.14 (.036)	0.27 (.331)	0.52 (.388)	−.07 (.045)	0.38 (.43)	0.67 (.529)
Lan	0.92 (.057)	3.1 (1.17)	4.71 (1.31)	0.94 (.072)	3.27 (1.52)	5.06 (1.78)
Bor	.06 (.059)	0.4 (0.44)	0.7 (.54)	0.03 (.074)	0.39 (.583)	0.71 (.734)

Notes: See notes to Tables 4.2(c), 4.3(a) and 4.3(b).

We move to address an alternative approach of allowing for common time factors; namely we consider our proposed approach as developed in Section 3. We find from Table 4.1 that the share of EU trade with the US has always been a consistent part of the extra-EU trade. For example, it is reported in *Trade policy review of the European Union: A Report by the Secretariat of the WTO* (2002) that the percentage of export (import) from Europe to the US increases from around the 10% (10%) of the total volume of EU export in 1960 to around the 25% (20%) in 2000. Hence, we expect that certain characteristics of the US will also help in further explaining the pattern of the intra-EU exports and/or total trades. In this regard, we consider the EU and the US as two main trade blocks and augment the model with the US reference variables, which we regard as observed common time factors. Here we simply choose the variable of $RERT_t$ that will capture any of the relative price effects between the European currencies and the US dollar.[5] We expect that a depreciation of the European currency with respect to the US dollar (an increase in $RERT_t$) should result in more extra-EU exports to and less extra-imports from the US, though its impact on the intra-EU trade will be ambiguous. We thus consider the model (4.13) for both triple and double index versions, where $f_t = RERT_t$ with heterogeneous parameters, and focus only on the FEM combined with the HT estimation results. Under our maintained assumption that common language dummy is only correlated with unobserved individual effects, we consider the four different instrument sets, denoted HT1, HT2, HT3 and HT4, respectively, where HT1 and HT2 are exactly the same before, namely $HT1 = \{RER\}$ and $HT2 = \{RER, GDPs, SIM, RLF\}$, whilst HT3 and HT4 are the sets combining HT1 and HT2 respectively with $\hat{\lambda}_i RERT_t$. Remind that we follow our theoretical discussion in Section 3 and use $\hat{\lambda}_i RERT_t$ as an additional source of instrument in HT3 and HT4.

Table 4.4 summarizes there results. First, looking at the results for the triple index model (Table 4.4(a)), we find that signs and significances of coefficients are preserved, though the magnitudes of the coefficients are somewhat different from the previous estimates reported in Table 4.4.2(a). But, the coefficient on EMU is surprisingly negative and significant. The HT estimates of coefficients on individual specific variables all show the expected signs, but the language dummy loses its statistical significance. Next turning to the double index model (Table 4.4(b)), most FEM estimates are similar to those shown in Table 4.2(b) with the following main difference: the coefficients on EMU and RLF are both insignificant. The HT estimates of the impacts of individual specific variables show more or less the similar patterns to Table 4.2(c), namely, the distance variable becomes insignificant whilst the language variable becomes more

[5] We construct $RERT_t$ like RER_{it}. Here the home currency is the European currency, *i.e.* ECU till 1998 and Euro from 1999 to 2001, and the foreign currency is the US Dollar. See also Data Appendix. We have also tried a different US reference variable such as the US GDP, and found the qualitatively similar results as described in the text.

important in explaining the pattern of trade flows.

Table 4.4. FEM and HT Estimation Results with an Observed Time Factor

Table 4.4(a). Triple Index Model

	FEM[1]	OLS	HT1	HT2	HT3	HT4
GDP_h	1.09 (.035)[2]					
GDP_f	0.88 (.035)					
SIM	0.21 (.055)					
RLF	0.01 (.005)					
RER	0.16 (0.01)					
CEE	0.33 (.011)					
EMU	−.06 (.018)					
Dist		−.43 (.02)	−.23 (.344)	−.43 (.161)	−.13 (.367)	−.43 (.161)
Lan		0.25 (.032)	1.26 (1.68)	0.25 (.472)	1.85 (1.81)	0.24 (.473)
Bor		0.49 (.033)	0.65 (.312)	0.49 (.222)	0.74 (.319)	0.49 (.222)

Notes: Here we augment the models in 4.2(a) by the observed time-specific factor, $RERT$. See also notes to 4.2(a). The sets of instruments used in the HT are: $\{RER\}$ for HT1, $\{RER, GDP, SIM, RLF\}$ for HT2, and HT3 and HT4 are HT1 and HT2 respectively combined with $\{\gamma_i RERT_t\}$.

Table 4.4(b). Double Index Model

	FEM[1]	OLS	HT1	HT2	HT3	HT4
GDP	2.02 (.03)[2]					
SIM	1.4 (.062)					
RLF	0.009 (.006)					
RER	0.11 (.011)					
CEE	0.31 (.012)					
EMU	−.008 (.019)					
Dist		−.4 (.024)	−.38 (.208)	−.19 (.227)	−.35 (.211)	−.17 (.227)
Lan		0.39 (.038)	0.44 (.641)	1.45 (.671)	0.64 (.627)	1.6 (.081)
Bor		0.43 (.041)	0.45 (.317)	0.61 (.309)	0.48 (.312)	0.63 (.311)

Notes: Here we augment the models in 4.2(b) by the observed time-specific factor, $RERT$. See also notes to 4.4(a).

We notice in passing that the choice of observed common factors might be somewhat arbitrary in general and that there is always a possibility of missing factors. In this regard, there is still a room for further improving previous estimation results, and we now take an alternative approach based on the assumption that the common time factors are unobserved and their impacts are heterogeneous. This approach has two advantages: first, we may avoid inevitable arbitrariness and difficulty in selecting observed common factors. Secondly and more importantly, this approach is also able to allow for certain degrees of cross section dependence via heterogeneous time-specific effects, and thus to avoid the potential bias of uncorrected estimates as described earlier. Here we follow the PCCE estimation methodology advanced by Pesaran (2002) to deal with this issue and thus consider the model (4.13), where we now have $f_t = \{\overline{y}_t, \overline{TGDP}_t, \overline{SIM}_t, \overline{RLF_t}, \overline{RER_t}\}$ and the bar over the variable indicates the cross sectional average of the variable of interest, namely $\overline{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$ and so on.[6] As before, we focus only on the FEM combined with the HT estimation results and maintain the assumption that common language dummy is only correlated with unobserved individual effects. We now consider the following four different instrument sets: HT1=$\{RER\}$ and HT2=$\{RER, GDPs, SIM, RLF\}$, whilst HT3 and HT4 are the sets combining $\left\{\hat{\lambda}_{1i}\overline{y}_t, \hat{\lambda}_{2i}\overline{TGDP}_t, \hat{\lambda}_{3i}\overline{SIM}_t, \hat{\lambda}_{4i}\overline{RLF_t}, \hat{\lambda}_{5i}\overline{RER_t}\right\}$ with HT1 and HT2 respectively.[7]

We provide these estimation results in Table 4.5. First, from Table 4.5(a) for the triple index model, we find that the impacts of foreign GDP, RLF and EMU are all insignificant, while the impact of CEE is smaller than reported in Table 4.2(a). The HT estimation results show that the distance is significantly more negative while both common language and border dummies become insignificant. Next turning to Table 4.5(b) for the double index model, most FEM estimates are quite similar to those reported in Table 4.2(b). Main differences are: the coefficients on EMU and RLF are both insignificant while the impact of CEE is now much smaller. The HT estimates of the impacts of individual specific variables confirms similar findings to

[6] We do not include cross sectional average of the CEE and EMU dummies to avoid the potential multicollinearity problem. We also notice that $\overline{TGDP}_t = \overline{GDP}_{ht} = \overline{GDP}_{ft}$.

[7] In practice, the subset of $\left\{\hat{\lambda}_{1i}\overline{y}_t, \hat{\lambda}_{2i}\overline{TGDP}_t, \hat{\lambda}_{3i}\overline{SIM}_t, \hat{\lambda}_{4i}\overline{RLF_t}, \hat{\lambda}_{5i}\overline{RER_t}\right\}$ can be parsimoniously used as instruments, though here we use all for convenience.

those reported in Table 4.2(c). Interestingly, once the instrument set is augmented with $\left\{\hat{\lambda}_{1i}\overline{y}_t, \hat{\lambda}_{2i}\overline{TGDP}_t, \hat{\lambda}_{3i}\overline{SIM}_t, \hat{\lambda}_{4i}\overline{RLF}_t, \hat{\lambda}_{5i}\overline{RER}_t\right\}$, we find that all individual specific variables (distance, common language and border) become strongly significant with expected signs. This may indicate the potential importance of using additional source of instruments.

Table 4.5. FEM and HT Estimation Results with an Unobserved Time Factor

Table 4.5(a). Triple Index Model

	FEM[1]	OLS	HT1	HT2	HT3	HT4
GDP_h	1.26 (.075)[2]					
GDP_f	−.03 (.075)					
SIM	0.16 (.081)					
RLF	−.0001 (.005)					
RER	0.03 (.011)					
CEE	0.12 (.013)					
EMU	−.001 (.018)					
Dist		−.97 (.039)	−.72 (.299)	−.92 (.268)	−.6 (.302)	−.94 (.254)
Lan		0.31 (.062)	1.57 (1.26)	0.5 (1.12)	2.25 (1.21)	0.46 (1.01)
Bor		0.56 (.064)	0.76 (.385)	0.59 (.371)	0.87 (.396)	0.59 (.364)

Notes: Following Pesaran (2002) we augment the model estimated in 4.2(a) by multiple factors $\{\overline{y}_t, \overline{TGDP}_t, \overline{SIM}_t, \overline{RLF}_t, \overline{RER}_t\}$. The sets of instruments used in the HT are: $\{RER\}$ for HT1, $\{RER, GDP, SIM, RLF\}$ for HT2, and HT3 and HT4 are HT1 and HT2 respectively combined with $\{\lambda_{1i}\overline{y}_t, \lambda_{2i}\overline{TGDP}_t, \lambda_{3i}\overline{SIM}_t, \lambda_{4i}\overline{RLF}_t, \lambda_{5i}\overline{RER}_t\}$. See also notes to 4.2(a).

Table 4.5(b). Double Index Model

	FEM	OLS	HT1	HT2	HT3	HT4
GDP	1.63 (.115)					
SIM	1.11 (.093)					
RLF	−.001 (.005)					
RER	0.03 (.014)					
CEE	0.14 (.013)					
EMU	−.01 (.017)					
Dist		−.73 (.023)	−.38 (.23)	−.41 (.233)	−.46 (.218)	−.48 (.209)
Lan		0.51 (.037)	2.33 (0.78)	2.17 (.76)	1.89 (.638)	1.82 (.628)
Bor		0.44 (.038)	0.72 (0.28)	0.7 (.266)	0.65 (.255)	0.65 (.245)

Notes: Following Pesaran (2002) we augment the model estimated in 4.2(b) by multiple factors, $\{\overline{y}_t, \overline{TGDP}_t, \overline{SIM}_t, \overline{RLF_t}, \overline{RER_t}\}$. See also notes to 4.5(a).

Comparing the above three extended estimation results and evaluating them in light of our *a priori* expectations, we may reach to the following conclusion: first, the results obtained using the conventional $T-1$ fixed dummies (with their homogeneous impacts) are least satisfactory, which might indicate that the conventional approach may be too limited to accommodate the time effects. Second, the estimation results with an observed time factor are somewhat mixed in the sense that most estimation results are relatively sensible for the double index model, but not quite for the triple index model. Finally, the estimation results with unobserved time factor (in conjunction with the PCCE estimation) are similar to but somewhat better than those obtained using the observed common time factor. In particular, the results of Table 4.5(b) for the double index model for explaining the patterns of bilateral total trade flows are mostly sensible. Therefore, this observation may suggest the potential advantage of our proposed approach using over the conventional one based on the fixed time dummies.

We now summarise our main findings in a broad context combining all of the above estimation results together but mainly with estimation results in Tables 2 and 5. We begin with the triple index model in explaining the pattern of bilateral real exports. The impact of the GDP variables is mostly significant and positive with the total impact being just under 2. Only the impacts of foreign population are found to be significant but negative. The impact of similarity in relative size of trading countries are mostly significant and positive, ranging between 0.16 and 0.35. The impact of differences in relative factor endowments are mostly significant and positive, ranging between 0.01 and 0.03. The impacts of CEE are all positive and significant, mostly around 0.3. The results are mixed for the impacts of EMU, but mostly insignificant in both Tables 2(a) and 5(a). The impacts of distance are mixed in Table 4.2(c), but become significantly negative in Table 4.5(a). The impacts of common language are mixed in Table 4.2(c) but become insignificant in Table 4.5(a). The impacts of common border are mostly significant and range between 0.49 and 0.76. Next, we move on to the double index model in explaining the pattern of bilateral real total trades. The impacts of GDP are all significant and positive, ranging between 1.63 and 2.02. The impacts of population are insignificant. The impacts of SIM are all significant and positive, ranging between 1.11 and 1.4, which are significantly larger than its impacts on exports only. The impacts of RLF are significantly positive in Table 4.2(b), but insignificant in Table 4.5(b). The impacts of CEE are all significantly positive. The impact of EMU is significantly positive in Table 4.2(b), but becomes insignificant in Table 4.5(b). The impacts of distance are mostly significantly negative, the impacts of common language are mostly significantly positive, and the impacts of common border are mostly significantly positive.

Though the above estimation results and their interpretations are more or less consistent with our *a priori* expectations, we notice that there are two conflicting findings between the benchmark estimation results in Table 4.2 and the results of our preferred extended model in Table 4.5; namely, the role of the RFL and EMU variables. The impacts of RLF are found to be significant and positive in Table 4.2, but become insignificant in Table 4.5, whilst the impacts of EMU are found to be mostly insignificant, but only become significantly positive in Table 4.2(b). As mentioned earlier, the impact of RLF on total trade flows might not be unambiguous since the total trade flows are the sum of inter- and intra-industry trades. Next, we earlier discussed that empirical evidence on the impact of EMU on trade flows is mixed. In particular, de Souza (2002) argues that either the periods are too short after an introduction of the Euro to use the EMU dummy as an adequate proxy for monetary union membership, or forward looking agents anticipate and thus discount the increase of trade associated with the EMU membership. In this regard we also expect that the impacts of EMU are yet to be significant. Along this line of logic we may conclude that the estimation results obtained using the extended model with unobserved common time effects seem to be much more sensible. This observation may suggest that it is also important to allow for a certain degree of cross section dependence unobserved common time effects, otherwise the resulting estimates would be severely biased. Lastly, although it is difficult to judge the relative fit of the various models proposed,[8] few general implications on European trade can be drawn from the results presented. The Intra-EU trade flows appear to be more influenced by GDP, intra industry trade, common trade agreements, transportation costs and cultural proximities rather than by population, relative factor endowments and monetary unions. In sum, all variables that proxy for similarity give a positive and consistent impact except for the common currency dummy. These results are somehow expected given the characteristics of the data (see the Explanatory Data Analysis) and provide evidence in favor of the IRS-based trade theory.

4.5 Conclusions

In this analysis we follow recent developments of panel studies surrounding the common time effects, *e.g.* Ahn, Lee and Schmidt (2001), Ng and Bai (2001), Pesaran (2002) and Phillips and Sul (2002), and advance an alternative estimation framework in which we explicitly allow for the existence of observed and unobserved common time-specific factors and individual responses to those common factors are heterogeneous across country pairs. In the context of this extended panel model we generalize the HT estimation methodology and develop the underlying econometric theory. More importantly, we propose to employ an alternative source of instruments in addition to the conventional (internal) instruments suggested by HT; namely, some of heterogeneous common factors under the assumption that they are correlated with individual specific variables but not with unobserved individual effects. We apply our proposed

[8]Neither relative fit measures nor tests for overidentifying restrictions have been constructed for the HT methods so far. This will be object of future research as those are essential tools for choosing amongst the various models and instruments sets here proposed.

(extended) HT estimation technique along with the conventional approaches to a comprehensive analysis of the gravity equation of bilateral trade amongst the 15 European countries over 1960-2001. Empirical results clearly demonstrate that our proposed approach fits the data reasonably well and its estimations results are sensible in a number of different dimensions. In particular, we first find that our proposed (extended) HT estimation provides much more sensible results than the conventional approach based on the fixed time dummies. Further, we also notice that our proposed HT estimation results produce more sensible predictions on the impacts of differences in factor endowments and of the common currency dummy than the conventional approach with and without fixed time dummies. This observation may indicate the importance of allowing for a certain degree of cross section dependence unobserved common time effects, otherwise the resulting estimates would be severely biased.

A number of extensions will be desirable. First, it would be interesting to analyse the gravity models of international trade over different periods of time. For instance, as discussed in the subsection on explanatory data analysis, the impacts of intra- and inter-industry trades will be different over different periods of time. If so, we might expect that the role of certain explanatory variables such as RLF and EMU also change accordingly over different periods of time. Second, it would be worth investigating the effect of globalization on transport costs more explicitly. For instance, transport and communication revolutions should lead to a dispersion of economic activity. Although this dispersion did not occur with the reduction in transportation costs during the first wave of the globalization in the 20th century, the second wave of globalization associated with recent information and communication technologies revolution should lead to an integrated equilibrium view of the 'death of the distance'. Hence, it would be interesting to study the effect of an 'augmented' trade-barrier function which make transport costs both dependent on and independent of distance in addition to the standard trade-barrier function that only comprehend variables like distance, common language and common border dummies as employed in the current analysis, see Brun, Carrere, Guillaumont and de Melo (2002). Finally, once the data over the longer time periods will be available, we will reexamine the issue concerning the impacts of the Euro on the bilateral intra-EU trade as the insignificantly estimated impact of the EMU dummy might be due to the shortage of observations.

4.6 Data Appendix

We now describe how the variables are constructed. All variables are converted in constant dollar prices with 1995 as the base year. Bilateral exports and imports are defined as logarithms of real export $\left(X_{hft}^{R}\right)$ and real imports $\left(M_{hft}^{R}\right)$, X_{hft}^{R} and M_{hft}^{R} are obtained by $X_{hft}^{R} = X_{hft}^{N} \times \frac{100}{XPI_{US}}$, $M_{hft}^{R} = M_{hft}^{N} \times \frac{100}{MPI_{US}}$, where X_{hft}^{N} and M_{hft}^{N} are bilateral export and import measured in millions of current US dollars, and XPI_{US} and MPI_{US} are the US export and import price indices. Then, the total volume of trade is given by $Trade = \ln\left(X_{hft}^{R} + M_{hft}^{R}\right)$. GDP of home and foreign country are defined as logarithms of GDP_{ht}^{R} and GDP_{ft}^{R}, where GDP_{ht}^{R} and GDP_{ft}^{R} are gross domestic products at constant dollar of country h and f, respectively, and the total GDP is defined as $TGDP_{it} = \ln\left(GDP_{ht}^{R} + GDP_{ft}^{R}\right)$. GDPs are originally expressed in million Euro for the twelve countries that joined the European Monetary Union (Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain) and in millions of current national currency for Denmark, Sweden and UK (GDP^{N}). In the last three cases the original nominal values of GDP have been deflated by the GDP deflator ($GDPD$, 1995 = 100) of the respective countries whereas for the remaining countries the European GDP deflator has been used. We also convert GDPs in US dollar at the exchange rate of 1995 (mean over period) in order to exclude the effect of a dollar depreciation or appreciation as follow:

$$GDP_{hft}^{R} = GPD_{hft}^{N} \times \frac{100}{GDPD_{ht}} \times \left(\frac{US\$}{NC_{h}}\right)_{1995},$$

where NC_h stands for national currency of the home country. Population of home and foreign countries are defined as logarithms of POP_{ht} and POP_{ft}, where POP_{ht} and POP_{ft} are the population of country h and f measured in million of inhabitants and the total population is defined as $TPOP_{it} = \ln\left(POP_{ht} + POP_{ft}\right)$. Next, we construct SIM_{it} and RLF_{it} respectively by

$$SIM_{it} = \ln\left[1 - \left(\frac{GDP_{ht}^{R}}{GDP_{ht}^{R} + GDP_{ft}^{R}}\right)^{2} - \left(\frac{GDP_{ft}^{R}}{GDP_{ft}^{R} + GDP_{ht}^{R}}\right)^{2}\right],$$

$$RLF_{it} = \ln\left|PGDP_{ft}^{R} - PGDP_{ht}^{R}\right|,$$

where $PGDP$ is per capita GDP. Real exchange rates in constant dollars at 1995 are defined as $RER_{it} = NER_{it} \times XPI_{US}$, where NER_{it} is nominal exchange rate between currencies h and f in year t in terms of dollars. Lastly, the distance between countries is measured as the great circle distance between national capitals in kilometers.

The data sources are as follows: Export and import price indices are collected from OECD *Economic Outlook*, GDP deflators from World Bank *World Development Indicators*, and bilateral nominal export and import data (X^N and M^N) from OECD, *Statistical Compendium*, Main Economic Indicator, Yearly Statistic of Foreign Trade in current dollars, GDP from IMF *International Financial Statistics*, Economic Concept View, National Accounts, per capita GDP (already converted in constant dollars) from the World Bank *World Development Indicators*, population from the World Bank *World Development Indicators*, and NER from OECD, *National Accounts*, Volume I.

Chapter 5

Concluding Remarks

This research has concentrated on applying some of the most recent developments on the issue of panel data analysis to different topics in Economics. Empirical applications of models and methods presented have strongly confirmed the adequacy of this analysis to investigate economic phenomena.

In the First Chapter we apply the GMM methodology for testing for the validity of the PIH using micro panel data. Regardless of the evident theoretical importance of the PIH in explaining intertemporal choice of consumption [see Cochrane (1989)] some of the empirical tests, mainly conducted with aggregate data, do not provide support for the theory [see Hall (1978), Flavin (1981), Mankiw and Shapiro (1985) West (1988), Deaton and Campbell (1989), Campbell and Mankiw (1990) and Galì (1991)]. On the other hand, tests performed on micro data generally provide evidence in favor of the Hypothesis [see Zeldes (1989), Runkle (1991), Attanasio and Weber (1993) and DeJuan and Seater (1999)]. In our analysis we aim to shed more light on the issue of empirical tests of the PIH. We test the PIH via three alternative approaches, *i.e.* Flavin (1981), Euler equation, and Friedman (1957) characteristic tests, using data from the British Household Panel Survey (BHPS). In conformity with the recent studies conducted using micro data the PIH receives general support from our data when the appropriate panel data techniques are applied and consistent estimates are provided. The most relevant result is that testing a model suitable for aggregate consumption, *i.e.* Flavin (1981), with panel data provides evidence in favor of the PIH. Our findings can be considered as a piece of evidence in favor of the thesis that empirical tests of the

PIH, based on aggregate data, might suffer from mis-specifications or overlook some fundamental characteristics of micro data and therefore vitiate the results that lead to rejection of the PIH [see Attanasio and Weber (1993), Attanasio and Browing (1995) and DeJuan and Seater (1999)].

In the Second Chapter we to test for anomalies on factor pricing models via a panel data approach. Our investigation provides a theoretically coherent example to which panel data techniques dealing with both homogeneous and heterogeneous parameters can be applied. The suggested panel-based anomaly tests have one clear advantage over the conventional TP-based tests: they are based on full information maximum likelihood estimates so that they do not suffer from the errors in variable problem and have all the usual asymptotic properties associated with likelihood tests. In addition the panel technique adopted yields parameter estimates of firm specific effects that (under the alternative) are fully efficient. We apply both the conventional and the panel data approach to a large data set of UK stock returns between 1968 and 2002. The results from the panel data approach interestingly show, contrary to the results of Fama and French (1996), that adding macro factors does not drive out the significance of a standard single market factor and firm specific characteristics remain significant even in multifactor model. We tentatively argue that those findings could be a result of the greater efficiency of our estimates and power of our testing procedure.

Finally in the Third Chapter we analyse gravity models of international trade flows via a partially heterogenous panel data model. We emphasise the importance of explicitly allowing for the time specific effects and allow for the existence of common observable and unobservable time-specific effects to affect bilateral trade. In this framework we extend the conventional panel data approach and generalize the HT estimation. We develop the underlying econometric techniques and suggest an alternative source of instruments. We apply our proposed (extended) HT estimation technique along with the conventional approaches to a comprehensive analysis of the sources of bilateral trade amongst the 15 European countries over 1960-2001. In the empirical results the gravity model seems to fit the data well. Our empirical findings clearly suggest the potential advantage of our proposed approach over the conventional one based on the fixed time dummies. Furthermore, comparing the estimation results for the benchmark case without allowing for time effects and our proposed model with unobserved common time factors, we may conclude that the estimation results obtained using our proposed extended model with unobserved common time factors seem to be more sensible. This may reflect that it is also important to allow for a certain degree of cross section dependence via unobserved common time factors, otherwise the resulting estimates would be severely biased.

A number of extensions and modifications to the models proposed are possible. First, we might extend the analysis in the First Chapter to the analysis of the impacts of a number of different variables on the choice of intertemporal consumption. The BHPS is a rather comprehensive source of information on individual and household. It might be particularly interesting to evaluate the role of the expectations and the role of intergenerational transfers (the BHPS contains information on expectation of future income as well as on pensions or benefits) in the intertemporal choice of consumption. Along this line of logic the differences between the PIH and the LCH might be better analysed and possibly tested [see Jappelli and Modigliani (1998)]. Also in the Third Chapter, the incorporation of the effect of globalization on transport costs would be worth investigating. Transport and communication revolutions should lead to a dispersion of economic activity. Although this dispersion did not occur with the reduction in transportation costs during the first wave of the globalization in the XX century, the second wave of globalization associated with recent information and communication technologies revolution should lead to an integrated equilibrium view of the 'death of the distance'. Hence, it would be interesting to study the effect of transport costs that are both dependent and independent of geographical distance [see Brun, Carrere, Guillaumont and de Melo (2002)]. Additionally it would be interesting to analyse the impacts that the explanatory variables have on trade flows in different period of time and especially to analyse the role of the Euro on bilateral intra-EU trade once that more observations will be available to the analysis.

Bibliography

[1] Ahn, S.C. and C. Gadarowski (2001) "Two-Pass Cross-Section Regression of Factor Pricing Models" Unpublished Manuscript, Arizona State University.

[2] Ahn, S.G., Lee, Y.H. and P. Schmidt (2001) "GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects" *Journal of Econometrics*, 101, 219-255.

[3] Ahn, S. and P. Smidth (1995) "Efficient Estimation of Models for Dynamic Panel Data" *Journal of Econometrics*, 68, 5-27.

[4] Alonso-Borrego, C. and M. Arellano (1999) "Symmetrically Normalized Instrumental-Variable Estimation using Panel Data" *Journal of Business and Economic Statistics*, 17, 39-49.

[5] Alvarez, J. and M. Arellano (1998) "The Time Series and Cross-Section Asymptotic of Dynamic Panel Data Estimator", CEMFI Working Paper No.9808.

[6] Amemiya, T. and T. MaCurdy (1986) "Instrumental Variables Estimation of an Error Components Model" *Econometrica*, 54, 869-880.

[7] Anderson, P.S. (1979) "A Theoretical Foundation for the Gravity Equation" *American Economic Review*, 69, 106-116.

[8] Ando, A. and F. Modigliani. (1960) "The Permanent Income Hypothesis and the Life Cycle Hypothesis of Saving Behavior: Comparison and Tests" In Friend, I. and Jones, R. eds. Consumption and saving, vol.2.University of Pennsylvania Press.

[9] Ando, A. and F. Modigliani (1963) "The Life Cycle Hypothesis of Saving: Aggregate Implications and Tests" *American Economic Review*, 53, 55-84.

[10] Arellano, M. and O. Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Components Models" *Journal of Econometrics*, 68, 29-51.

[11] Attanasio, O. (1998) "Consumption Demand" NBER Working Paper No.6466.

[12] Attanasio, O. and M. Browning (1995) "Consumption over the Life Cycle and over the Business Cycle" *American Economic Review*,.85, 1118-37.

[13] Attanasio, O. and G. Weber (1995) "Is the Consumption Growth Consistent with Intertemporal Optimisation? Evidence from the Consumption Expenditure Survey" *Journal of Political Economy*, 103, 1121-55.

[14] Bai, J. (2003) "Inferential Theory for Factor Models of Large Dimensions" *Econometrica*, 71,135-171.

[15] Bai, J. and S. Ng (2002) "Determining the Number of Factors in Approximate Factor Models" *Econometrica, 70, 191-221.*

[16] Baldwin, R. E. (1994) "Towards an Integrated Europe" London, CEPR.

[17] Balestra, P. and M. Nelrove (1966) "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas" *Econometrica,* 34, 585-612.

[18] Bank of England (1988), "Recent Developments," *Bank of England Quarterly Bulletin.*

[19] Banz, R.W. (1981) "The Relationship Between Return and Market Value of Common Stocks," *Journal of Financial Economics, 9, 3-18.*

[20] Belanger, D., S. Gutierrez and R. Raynauld (1992) "The Impact of Exchange Rate Variability on Trade Flows: Further Results on Sectorial U.S. Import from Canada" *North American Journal of Economics and Financė*, Spring, 888-92.

[21] Bergstrand, J.H. (1985) "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence" *The Review of Economics and Statistics*, 67, 474-481.

[22] Bergstrand, J.H. (1989) "The Generalized Gravity Equation, Monopolistic Competition, and Factor Proportions Theory of International Trade" *Review of Economics and Statistics*, 71, 143-153.

[23] Bergstrand, J.H. (1990) "The Heckscher-Olin-Samuelson Model, the Linder hypothesis and the Determinants of Bilateral Intra-industry Trade" *Economic Journal*, 100, 1216-29.

[24] Berk, J.B. (1995) "A Critique of Size-related Anomalies" *Review of Financial Studies*, 8, 275-286.

[25] Black, F. (1972) "Capital Market Equilibrium with Restricted Borrowing" *Journal of Business*, 45, 444-454.

[26] Black, F., M.C. Jensen and M. Scholes (1972) "The Capital Asset Pricing Model: Some Empirical Tests" in *Studies in the Theory of Capital Markets*, ed. by M.C. Jensen. Praeger, New York.

[27] Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" *Journal of Econometrics*, 87, 115-143.

[28] Blundell, R., Bond, S. and F. Windmeijer (2000) "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator" in B. Baltagi ed., Nonstationary Panels, Panel Cointegration and Dynamic Panels, Elsevier Science.

[29] Bond, S., Hoeffler, A. and Temple, J. (2001) "GMM Estimation of Empirical Growth Models" Institute for Fiscal Studies Discussion Paper No.01/525.

[30] Bougheas, S., P.O. Demetriades and E.L.W. Morgenroth (1999) "Infrastructure, Transport Costs and Trade" *Journal of International Economics*, 47, 169-189.

[31] Brada, J.C. and J.A. Mendez (1988) "Exchange Rate Risk, Exchange Rate Regime and the Volume of International Trade" *Kylos*, 41, 263-80.

[32] Breusch T., G. Mizon and P. Schmidt (1989) "Efficient Estimation using Panel Data" *Econometrica*, 57, 695-700.

[33] Brown, T.M. (1952) "Habit Persistence and Lags in Consumer Behaviour" *Econometrica*, 20, 355-371.

[34] Brun, J., Carrere, C., Guillaumont, P. and J. de Melo (2002) "Has distance died? Evidence from a panel gravity model" Unpublished manuscript, University of d'Auvergne.

[35] Bun, M.J.G. and F.J.G.M. Klaassen (2002) "The Importance of Dynamic in Panel Gravity Models of Trade" Unpublished manuscript, University of Amsterdam

[36] Cagan, P. (1955) "The Monetary Dynamics of Hyperinflation" in Friedman, M. (ed) Studies in the quantity theory of money. Chicago University Press.

[37] Campbell, J.Y. and A. Deaton (1989) "Why Is Consumption So Smooth?" *Review of Economic Studies*, vol.56, 357-373.

[38] Campbell, J.Y. and N.G. Mankiw (1990) "Permanent Income, Current Income and Consumption" *Journal of Business and Economic Statistics*, 8, 265-279.

[39] Chan, K.C. and N. Chen (1991) "Structural and Return Characteristics and Small and Large Firms" *Journal of Finance, 46*, 1467-1484.

[40] Cheng, I. and H.J. Wall (2002) "Controlling Heterogeneity in Gravity Models of Trade" Working Paper 1999-010C, Federal Reserve Bank of St. Louis.

[41] Christ, C.F (1963) "Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Y. Grunfeld. Stanford" University Press.

[42] Cochrane, J.H. (1989) "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-rational Alternatives" *American Economic Review*, 79, 319-337.

[43] Cochrane, J.H. (2001), *Asset Pricing*. Princeton University Press, Oxford.

[44] Cornwell, C., Schmidt, P. and R. Sickles (1990) "Production frontiers with cross-sectional and time-series variation in efficiency levels" *Journal of Econometrics*, 46, 185-200.

[45] Daniel, K., Sheridan, T. and K.C.J. Wei (2001) "Explaining the Cross-Section of Stock Returns in Japan: Factors or Characteristics?" *Journal of Finance* 2, 743-766.

[46] Daniel, K.T. and S. Titman (1997), "Evidence of the Characteristics of Cross Sectional Variance in Stock Returns" *Journal of Finance,* 52, 1-33.

[47] Darby, M.R. (1974) "The Permanent Income Hypothesis: A Restatement" *Quarterly Journal of Economics*, vol.88, 228-250.

[48] Davis, D.R. (2000) "Understanding International Trade Patters: Advantages of the 1990s" Unpublished Manuscript, Columbia University.

[49] De Grauwe, P. and F. Skudelny (2000) "The Impact of EMU on Trade Flows" *Weltwirtschaftliches Archiv*, 136, 381-397.

[50] De Nardis, S. and C. Vicarelli (2003) "The Impact of the Euro on Trade" ENEPRI, Working Paper No.17.

[51] De Sousa, J and A. Disdier (2002) "Trade, Border Effects, and Individual Characteristics: A Proper Specification" Unpublished Manuscript, University of Paris 1.

[52] de Souza L.V. (2002) "Trade Effects of Monetary Integration in Large, Mature Economies: a Primer on the European Monetary Union" Keil Working Paper No.1137.

[53] Deardorff, A.V. (1984) "Testing Trade Theories and Predicting Trade Flows" in *Handbook of International Economics*, 1.

[54] Deardorff, A.V. (1998) "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" NBER Working Paper No.5377.

[55] Deaton, A.(1991) "Saving and Liquidity Constraints" *Econometrica*, 59, 1221-48.

[56] Deaton, A. (1992) "Understanding Consumption" Oxford University Press.

[57] DeJuan, J.P. and J.J. Seater (1998) "Testing Permanent Income Hypothesis with Classification Methods" Unpublished Working Paper.

[58] DeJuan, J.P. and J.J. Seater (1999) "The PIH: Evidence from the Consumer Expenditure Survey" *Journal of Monetary Economics*, 43, 351-376.

[59] Dimson, E. (1979) "Risk Measurement When Shares Are Subjected to Infrequent Trading," *Journal of Financial Economics,* 7, 197-226.

[60] Dixit, A. and J. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity" *American Economic Review*, 67, 297-308.

[61] Duesenberry, J.S (1949) "Income, Saving and Theory of Consumption Behaviour" Harvard University Press.

[62] Eaton, J. and S. Kortum (2002) "Technology, Geography, and Trade" *Econometrica*, 70, 1741-79.

[63] Egger, P. (2001) "On the Problem of Endogenous Unobserved Effects of the Gravity Models" Working Paper, Austrian Institute of Economic Research, Vienna.

[64] Egger, P. (2002) "A Note on the Proper Econometric Specification of the Gravity Equation" *Economics Letters*, 66, 25-31.

[65] Egger, P. and M. Pfaffermayr (2003) "The Proper Econometric Specification of the Gravity Equation: A Three-way Model with Bilateral Interaction" Working Paper, Austrian Institute of Economic Research, Vienna.

[66] Evenett, S.J. and W. Keller (2002) "On the Theories Explaining the Success of the Gravity Equation" *Journal of Political Economy*, 110, 281-316.

[67] Fama, E.F. and K.R. French (1992) "The Cross-Section of Expected Stock Returns" *Journal of Financial Economics*, 47, 427-465.

[68] Fama, E.F. and K.R. French (1993) "Common Risk Factors in the Returns on Stocks and Bonds" *Journal of Finance*, 33, 3-56.

[69] Fama, E.F. and K.R. French (1996) "Multifactor explanations of Asset-Pricing Anomalies" *Journal of Finance*, 51, 555-584.

[70] Fama, E. and J.D. Macbeth (1973) "Risk Return and Equilibrium: Empirical Tests" *Journal of Political Economy*, 71, 607-636.

[71] Ferber, R. (1973) "Consumer Economics, A Survey" *Journal of Economic Literature,* 11, 1303-42.

[72] Flavin, M.A. (1981) "The Adjustment of Consumption to Changing Expectation about Income" *Journal of Political Economy,* 89, 974-1009.

[73] Frenkel, J.A. and S. Wei. "Trade Blocs and Currency Blocs" NBER Working Paper No.4335.

[74] Friedman, M. (1957) "A Theory of Consumption Function" Princeton University Press.

[75] Friedman, M. and S. Kuzten (1945) "Income from Independent Professional Practice" NBER.

[76] Galì, J. (1991) "Budget Constraints and Time Series Evidence on Consumption" *American Economic Review,*.81, 1238-53.

[77] Gibbons, M.R. (1982) "Multivariate Tests of Financial Models: A New Approach" *Journal of Financial Economics*, 10, 3-27.

[78] Gibbons, M.R., S.A. Ross and J. Shanken (1989) "A Test of the Efficiency of a Given Portfolio" *Econometrica*, 57, 1121-52.

[79] Glink, R. and A.K. Rose (2002) "Does a Currency Union Affect Trade? The Time Series Evidence" *European Economic Review*, 46, 1125-51.

[80] Grauer, R.R. (1999) "On the Cross-Sectional Relation between Expected Return, Betas and Size" *Journal of Finance,* 54, 774-789.

[81] Greene, W.H. (1997) "Econometric Analysis" Prentice-Hall Inc.

[82] Grubel, H.G. and P.J. Lloyd (1975) "Intra-industry Trade, the Theory and Measurement of International Trade in Differentiated Products" Macmillan Eds, London.

[83] Guilliotin, Y. and P. Sevestre (1994) "Estimations de Fonctions de Gains sur Donnees de Panel: Endogeneite du Capital Humain et Effects de la Selection" Economies et Prevision, 116, 119-135.

[84] Hall, R.E. (1978) "Stochastic Implication of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence" *Journal of Political Economy*, 86, 67-86.

[85] Harris, M.N. and L. Matyas (1998) "The Econometrics of Gravity Models" Melbourne Institute Working Paper No.5/98.

[86] Hausman, J.A. and W.E. Taylor (1981) "Panel Data and Unobservable Individual Effect" *Econometrica*, 49,1377-98.

[87] Helpman, E. (1987) "Imperfect Competition and International Trade: Evidence from Fourteen Industrialized Countries" *Journal of the Japanese and International Economies*, 1, 62-81.

[88] Hummels, D. and J. Levinsohn (1995) "Monopolistic Competition and International Trade: Reconsidering the Evidence" *Quarterly Journal of Economics*, 110, 798-836.

[89] Hussain, S.I., Diacon, S.R. and J.S. Toms (2000) "Asset Pricing Models and Market Anomalies: The UK Evidence" Unpublished Manuscript, University of Nottingham.

[90] Im, S.K., M.H. Pesaran and Y. Shin (2003) "Testing for Unit Roots in Heterogeneous Panels" *Journal of Econometrics*, 115, 53-74.

[91] Jacquemin, A and A. Sapir (1990) "Competition and Imports in the European Market" in *European Integration: Trade and Industry*. Cambridge University Press.

[92] Jagannathan, R. and Z. Wang (1998) "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-sectional Regression" *Journal of Finance*, 53, 1285-1309.

[93] Jegadeesh, N. and S. Titman (1993) "Returns to Buying Winners and Selling Losers: Implications for Stock market Efficiency" *Journal of Finance*, 45, 881-898.

[94] Johnson, M.B. (1971) "Household behaviour" Penguin eds.

[95] Keim, D.B. (1983) "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence" *Journal of Financial Economics,* 12, 13-32.

[96] Kim, J., De Marchi, N. and Morgan, M. (1995) "Empirical Model Particularities and Belief in the Natural Rate Hypothesis" *Journal of Econometrics*, 67, 81-102.

[97] Kiviet, J.F. (1995) "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models" *Journal of Econometrics*, 68, 53-78.

[98] Krugman, P.R. (1979) "Increasing Returns, Monopolistic Competition and International Trade" *Journal of International Economics*, 9, 469-479.

[99] Krugman, P.R. (1980) "Scale Economies, Product Differentiation, and Pattern of Trade" *American Economic Review,* 70, 950-959.

[100] Kuznet, S. (1946) "National Income: A Summary of Findings" NBER, NY Arno Press.

[101] Lakonishok, J., Shleifer, A. and R.W. Vishny (1994) "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance, 49, 1541-78.*

[102] Lawrence, E.C. (1991) "Poverty and Rate of Time Preference: Evidence from Panel Data" *Journal of Political Economy*, 99, 54-77.

[103] Leamer, E. (1992) "The Interplay of the Theory and Data in the Study of International Trade" in *Issues in Contemporary Economics*, 2 New York University Press.

[104] Levis, M. (1985) "Are Small Firms Big performers" *The Investment Analyst*, 76, 21-27.

[105] Levis, M. (1989) "Stock Market Anomalies: A Reassessment based on the UK Evidence" *Journal of Banking and Finance*, 13, 675-696.

[106] Linder, S.B. (1961) "An Essay on Trade and Transformation" John Wiley and Son Eds.

[107] Lintner, J. (1965) "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets" *Review of Economics and Statistics*, 47, 13-37.

[108] Lucas, R.E. (1976) "Econometric Policy Evaluation: A Critique" in The Phillips Curve and Labor Markets. K. Brunner and A.H, Meltzer eds.

[109] Ludvigson, S. and C.H. Paxon (1999) "Approximation Bias in Linearized Euler Equation" NBER Technical Working Paper No.236.

[110] Mankiw, N.G. and M.D. Shapiro (1985) "Trends, Random Walks and Tests of the Permanent Income Hypothesis" *Journal of Monetary Economics*, 16, 165-174.

[111] Mariger, R.P. and K. Shaw (1993) "Unanticipated Aggregate Disturbance and Tests of Life Cycle Consumption Model using Panel Data" *Review of Economics and Statistics*, 75, 48-56.

[112] Markusen, J. and R. Wigle (1990) "Expanding the Volume of North-South Trade" Economic Journal, 100, 1206-15.

[113] Marsh, P. (1979) "Equity Rights Issues and the Efficiency of the UK Stock Market," *Journal of Finance*, 34, 839-862.

[114] Martinez-Zaroso, I. and F. Nowak-Lehmann (2001) "Augmented Gravity Model: An Application to Mercosur-European Union Trade Flows" Unpublished Manuscript DB No.77.

[115] Matyas, L. (1997) "Proper Econometric Specification of the Gravity Model" *The World Economy*, 20, 363-368.

[116] Matyas, L. (1998) " The Gravity Model: Some Econometric Considerations" *The World Economy*, 21, 397-401.

[117] Matyas, L., Konya, L. and M.N. Harris (2002) "Modelling Export Activity of Eleven APEC Countries" Melbourne Institute Working Paper No.5/00.

[118] Mayer, T. (1972) "Permanent Income, Wealth and Consumption: A Critique of the Permanent Income Theory, Life Cycle Hypothesis and Related Theories" University of California Press.

[119] Miller, R.A. (1997) "Estimating Models of Dynamic Optimisation with Microeconomic Data" in Pesaran, M.H. and Smith, P. 1999. Handbook of Applied Econometrics, vol.2.

[120] Miller, R.A. and H. Sieg (1997) "A Microeconometric Comparison of Household Behaviour between Countries" *Journal of Business and Economic Statistics*, 15, 237-253.

[121] Molle, W. (1997) "The economics of European Integration: Theory, Practice and Policy" Ashgate Publishing Limited Eds.

[122] Nelrove, M.(1967) "Experimental Experience on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections" *Economic Studies Quarterly*, 18, 42-74.

[123] Nelrove, M.(1971) "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections" *Econometrica*, 39, 359-382.

[124] Nelson, C.R. (1987) "A Reappraisal of Recent Tests of the Permanent Income Hypothesis" *Journal of Political Economy*, 95, 641-646.

[125] Nickel, S (1981) "Biased in Dynamic Models with Fixed Effects" *Econometrica*, 49, 1417-26.

[126] Oguledo, V.I. and C.R. Macphee (1994) "Gravity Models: a Reformulation and an Application to Discriminatory Trade Arrangements" *Applied Economics*, 26, 107-120.

[127] Oxley, L. and M. McAleer (1993) "Econometric Issues in Macroeconomic Models with Generated Regressors" *Journal of Economic Surveys*, 7, 1-40.

[128] Pagan, A.(1984) "Econometric Issues in the Analysis of Regressions with Generated Regressors" *International Economic Review*, 25 221-247.

[129] Pakko, M.R. and H.J. Wall (2001) "Reconsidering the Effects of a Currency Union" Federal Reserve Bank of St. Louis.

[130] Persson, T. (2001) "Currency Union and Trade: How Large is the Treatment Effect?" *Economy Policy,* 33, 435-448.

[131] Pesaran, M.H (2002) "Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence". University of Cambridge DAE Working Paper No.0305 and CESifo Working Paper Series No.869

[132] Pesaran, M.H., Haque, N.U. and S. Sharma (2000) "Neglected Heterogeneity and Dynamics in Cross-country savings regressions" . In (eds) J. Krishnakumar and E. Ronchetti, *Panel Data Econometrics - Future Direction: Papers in Honor of Professor Pietro Balestra.* In the series, Contributions to Economic Analysis, Elsevier Science, 53-82.

[133] Pesaran, M.H. and Y. Shin (2003) "Small Sample Performance of Panel-based Correlated Common Effects Estimators under Cross Section Dependence" Working Paper.

[134] Pesaran, M.H., Shin, Y. and R. Smith (1999), "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of American Statistical Association,* 49, 621-634.

[135] Pesaran, M.H. and R. Smith (1995) "Estimating Lon-run Relationship from Dynamic Heterogeneous Panels" *Journal of Econometrics,* 68, 79-113.

[136] Phillips, P.C.B. and D. Sul (2002) "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence" Cowles Foundation Discussion Paper No.1362, Yale University.

[137] Quah, D. (1990) "Permanent and Transitory Movements in Income: An Explanation for 'Excess of Smoothness' in Consumption" *Journal of Political Economy*, 98, 449-475.

[138] Reinganum, M.R. "Misspecificaion of Capital Asset Pricing: Empirical Anomalies Based on Earning Yields and Market Value", *Journal of Financial Economics*, 9, 19-46.

[139] Robertson, D. and J. Symons (1992) "Some Strange Properties of Panel Data Estimators" *Journal of Applied Econometrics*, 7, 175-189.

[140] Rose, A.K. (2000) "One Money, One Market: Estimating the Effect of Common Currencies on Trade" *Economic Policy*, 30, 7-46.

[141] Rose, A.K (2002) "The Effect of Common Currency on International Trade: Where Do We Stand?" Unpublished manuscript, Haas School of Business, University of California.

[142] Runkle, D.E. (1991) "Liquidity Constraints and the Permanent Income Hypothesis: Evidence from Panel Data" *Journal of Monetary Economy*,.27, 73-98.

[143] Seater, J.J. (1998) "Testing the PILCH with Aggregate Data" *Macroeconomic Dynamic*, 2, 401-425.

[144] Shanken, J. (1985) "Multivariate Tests of Zero Beta CAPM" *Journal of Finance*, 21, 269-276.

[145] Shanken, J. (1992) "On the Estimation of Beta Pricing Model" *Review of Financial Studies*, 5, 1-33.

[146] Sharpe, W.F. (1964) "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk" *Journal of Finance*, 19, 425-442.

[147] Shin, Y. and A. Snell (2001) "Joint Asymptotic Results for Mean Group Tests in Heterogenous Panels: Application to the KPSS Stationary Test", unpublished manuscript, University of Edinburgh.

[148] Strong, N. and G. Xu (1997) "Explaining the Cross-Section of UK Expected Stock Returns" *British Accounting Review*, 29, 1-23.

[149] West, K.D. (1988) "The Insensitivity of Consumption to News about Income" *Journal of Monetary Economics*, 21, 17-33.

[150] Zeldes, S.P. (1989) "Consumption and Liquidity Constraints: An Empirical Investigation" *Journal of Political Economy*, 97, 305-346.

[151] Zhou, G. (1998) "On Cross-Sectional Stock Returns: Maximum Likelihood Approach" Unpublished Manuscript, University of Washington.