STOCHASTIC COMMON TRENDS AND LONG-RUN RELATIONSHIPS IN HETEROGENEOUS PANELS

by

Stephen Hall
The Management School
Imperial College of Science, Technology and Medicine

Giovanni Urga
City University Business School

May 1996

Financial support from ESRC grant no. W116251003 is gratefully acknowledged.

ABSTRACT

The conditions for the valid aggregation of a set of micro economic relationships to provide a valid macro relationship are stringent and have been known for a considerable time. The conclusion often suggested by this literature is that econometrics should proceed at as micro a level as possible using "panel data". Recent work on "panel data" estimation techniques has suggested that if the micro relationships are dealing with non stationary data then, even if these relationships cointegrate, the properties of a derived aggregate model will be even worse than we previously thought. Robertson and Symons, (1992) and, more recently, Pesaran and Smith (1995) have shown that, with data set of this type, inference often proceeds by imposing equality restrictions on parameters across individuals or through time. This is bad enough in a stationary world but they go on to show that in the presence of non stationary but cointegrated micro relationships aggregation can completely invalidate the macro relationship. In this paper we outline a special case where micro cointegrated relationships with heterogeneous parameter values aggregate to provide valid macro relationships. We further argue that while this is a special case it may often be relevant to real world examples and hence it may provide an explanation of the relative success of aggregate econometrics. We illustrate our argument by demonstrating that the special conditions are applicable to a panel data set of Italian labour demand data and that in this case aggregate estimation provides comparable parameter estimates with explicit micro estimation and aggregation.

JEL Classification number: C2, C5.
Keywords: Heterogeneous Panels, Stochastic common trend, cointegration,

ISSN 0969-6598

CONTENTS

1 INTRODUCTION

The conditions for the valid aggregation of a set of micro economic relationships to provide a valid macro relationship are stringent and have been known for a considerable time(Lovell(1973), Pesaran Pierse and Kumar(1989), Lee, Pesaran and Pierse(1990)) . The conclusion often suggested by this literature is that econometrics should proceed at as micro a level as possible using "panel data". "Panel data" is a very general term. It refers to data sets in which we have repeated observations over time on a sample of individual units, and typical panels consist of a large number of cross-sections of individuals . Thus, econometric estimates utilise both time series and cross-section variation in the data. Recent work on "panel data" estimation techniques has suggested that if the micro relationships are dealing with non stationary data then, even if these relationships cointegrate, the properties of a derived aggregate model will be even worse than we previously thought. Robertson and Symons, (1992) and, more recently, Pesaran and Smith (1995) have shown that, with a data set of this type, inference often proceeds by imposing equality restrictions on parameters across individuals or through time. This is bad enough in a stationary world but they go on to show that in the presence of non stationary but cointegrated micro relationships aggregation can completely invalidate the macro relationship. Pesaran and Smith (1995) in particular state that the common practice of aggregating and pooling by assuming homogeneity in dynamic models is "far from being innocuous"; instead they suggest estimating the individual micro equations and then taking the means of the estimated micro-parameters and relative standard errors.

In this paper we outline a special case where micro cointegrated relationships with heterogeneous parameter values aggregate to provide valid macro relationships. We further argue that while this is a special case it may often be relevant to real world examples and hence it may provide an explanation of the relative success of aggregate econometrics. We illustrate our argument by demonstrating that the special conditions are applicable to a panel data set of Italian labour demand data and that in this case aggregate estimation provides comparable parameter estimates with explicit micro estimation and aggregation.

Section 2 of the paper outlines the basic problem of aggregate estimation when the micro data is cointegrated but heterogeneous. Section 3 outlines the special case which provides a valid aggregate relationship. Section 4 then illustrates how this special case should be tested for and that the estimates of the long run relationship derived on the basis of aggregate data are indeed close to the aggregate of the estimates derived on the micro data. Section 5 concludes.

2 The Pesaran and Smith(1995) case

In a recent paper, Pesaran and Smith (1995) address the problem of estimating the average long run relationship between a set of variables when the micro relationships are made up of I(1) variables which cointegrate but with different cointegrating vectors. They conclude that the micro single equation approach gives consistent estimates of the long-run parameters, whilst the conventional view (Zellner(1962), Malinvaud(1956)) that the pooled and aggregate time-series estimators will also provide consistent estimators of the mean effects, is no longer valid.

In order to demonstrate this we can make use of a very simple example. Let us suppose that x_{it} are I(1) and there is a single cointegrating relationship between y_{it} and x_{it} for each group, with the parameters varying randomly across groups, i.e. suppose that

$$y_{it} = \beta_i x_{it} + \varepsilon_{it} \quad i = 1,\dots,N \quad t = 1,\dots,T \qquad 1$$

where ε_{it} is a stationary process which is integrated of degree 0, I(0), which implies that each of the relationships cointegrate, and that

$$\beta_i \neq \beta_j \quad i \neq j$$
2

Now suppose that we aggregate, then the variables will be

$$\bar{y} = \sum_{i=1}^{n} y_i , \quad \bar{x} = \sum_{i=1}^{n} x_i \qquad 3$$

Pesaran and Smith (1995) argue that the aggregated relationship does not cointegrate as exact aggregation of (1) over n gives

$$y_1 + \dots + y_n = \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon_1 + \dots \varepsilon_n \qquad 4$$

which gives

$$\bar{y} = \beta_1 \bar{x} + \sum_{i=2}^{n} (\beta_1 - \beta_i) x_i + \sum_{i=1}^{n} \varepsilon_i \qquad 5$$

given that x is I(1) we would only therefore expect to find cointegration between $\bar{x}$ *and* $\bar{y}$ 3 when $\beta_1 = \beta_j$ *all j* 4. So if we perform the standard aggregate regression with dynamic terms we will be dealing with non stationary aggregates which do not cointegrate and we would expect the parameter value of the aggregate long run relationship to tend to zero even though all the micro relationships do in fact cointegrate. This is the result pointed out by Robertson and Symons(1992).

3 A Special Case of Valid Aggregation

In this section we argue that, while the basic point made above is quite correct, there is a special case which does allow valid estimation even when the parameters of the micro relationships are heterogeneous. And, moreover, this special case is we believe valid for many real world situations. The basic argument put forward here is that if the exogenous variables in the micro relationships are driven by a common stochastic trend then the simple aggregate relationship can be shown to cointegrate. Moreover in many real world examples we might expect the nonstationary component to be common across a set of micro data, for example wages in different sectors might well be nonstationary because of the general nonstationarity in aggregate wages but relative wages across sectors might well be expected to be stationary. For completeness we will consider a full multivariate case of p regressors x_{jt} ,j=1...p for each of the individual components of the panel(i). Our argument may be seen formally quite simply, suppose that the exogenous variables are all driven by the following common trend model

$$\begin{aligned} x_{ijt} &= \boldsymbol{a}_{ij} x_{jt} + \boldsymbol{m}_{ijt} \\ \Delta x_{jt} &= \boldsymbol{x}_{jt} \end{aligned} \qquad 6$$

where $\boldsymbol{m}_{ijt}$ *and* $\boldsymbol{x}_{jt}$ are stationary ARMA error processes, that is they are integrated of degree 0, I(0).

then x_{jt} becomes the common stochastic trend which drives all the individual x_{ij}'s. We can then express the aggregate relationship as

$$\begin{aligned} \overline{y}_t &= \sum_{j-1}^{p} \sum_{i=1}^{n} \boldsymbol{b}_{ij} x_{ijt} + \sum_{i=1}^{n} \boldsymbol{e}_{it} \\ &= \sum_{j=1}^{p} \sum_{i=1}^{n} \boldsymbol{b}_{ij} \boldsymbol{a}_{ij} x_{jt} + \sum_{i=1}^{n} \boldsymbol{e}_{it} + \sum_{j=1}^{p} \sum_{i=1}^{n} \boldsymbol{b}_{ij} \boldsymbol{m}_{ijt} \end{aligned} \qquad 7$$

and in terms of the aggregates this becomes

$$\begin{aligned} \overline{y}_t &= \sum_{j=1}^{p} \frac{\sum_{i=1}^{n} \boldsymbol{b}_{ij} \boldsymbol{a}_{ij}}{\sum_{i=1}^{n} \boldsymbol{a}_{ij}} \overline{x}_{jt} \\ &+ \sum_{i=1}^{n} \boldsymbol{e}_{it} + \sum_{j=1}^{p} \sum_{i=1}^{n} \boldsymbol{b}_{ij} \boldsymbol{m}_{ijt} + \sum_{j=1}^{p} \sum_{i=1}^{n} \boldsymbol{m}_{ijt} \frac{\sum_{i=1}^{n} \boldsymbol{b}_{ij} \boldsymbol{a}_{ijj}}{\sum_{i=1}^{n} \boldsymbol{a}_{ij}} \end{aligned} \qquad \mathbf{8}$$

so that the aggregate equation cointegrates (the error term consists of a linear combination of weighted stationary ARMA components, which is of course stationary) and moreover the aggregate coefficients are a weighted average of the coefficients in the micro relationships.

The key to this special case is, of course, the validity of the common factor linking the x variables. We can speculate that in many cases the non stationary part of a group of related micro series might well be common, panels of wage data, consumption, prices, etc might well have this property. Indeed it would be surprising if relative wages or prices between sectors were nonstationary and so we might expect that the common factor representation would often be a good one in terms of the main nonstationary component in most data series. We would also suggest that if this is a common property of many data sets then it is a formal explanation of why aggregate econometric estimation works as well as it does, despite the standard conditions for aggregation which are highly implausible.

Formally this suggests that an important stage in analysing a panel of data should be an investigation of the existence of common stochastic trends amongst the individual components of the panel. This can be done in the autoregressive representation by testing for the presence of n-1 cointegrating vectors amongst a set of n series (thus implying one common stochastic trend) or it can be accomplished in the moving average representation by testing for the presence of a single common factor amongst the series following Geweke(1977). In the next section we implement these procedures for a panel of Italian employment data and illustrate that the existence of a single common factor is plausible and that when we find this we get comparable results from both aggregate and micro estimation.

4. An Empirical Example

In this section we demonstrate both that the special case of the last section has empirical relevance and that the predicted finding of valid aggregate estimation seems to be born out in practise. We take a data set of 45 firms belonging to the Italian manufacturing sector over the period 1958-1985. The series come from the CERIS-CNR Research Centre (Turin, Italy) which has been monitoring the accounts of some large industrial firms. Firms may be divided into private, state-owned and foreign firms. In this example we use the 21 privately owned firms. A conventional model of labour demand is used (Nickell, 1986) where, the level of employment(n), measured in annual average number of workers, is explained by the unit real labour cost (w) and real sales (y). All variables are expressed at 1958 prices and natural logs are used. The 21 firms belong to 7 industries (see Data Appendix for more details). A casual inspections of the three variables suggests that all the variables are trended, and these trends seem to be common for each variable.

To test this possibility more formally we adopt both of the two procedures suggested above. However an obvious problem which immediately arises here is one of dimensionality. The panel contains some 27 observations on each of the 3 variables for the 21 firms. To perform a conventional test for the cointegrating rank of the system would involve estimating the number of cointegrating vectors amongst 21 variables based on only 27 observations. This is clearly impossible. We therefore propose testing across the firms in each industry for a single common factor and then testing a sub-group consisting of one firm from each industry. So for example group 2 (chemical and rubber) comprises 6 firms, we would first test these six firms for a single stochastic trend and then take the first of the six to form a group across the seven industries and test this. We would also note an interesting difference in the approach of the two testing procedures. Testing for the number of cointegrating vectors is carried out by performing the Johansen(1988) procedure on each of the sub groups. If there are n variables in the group then we would want to establish that there are n-1 cointegrating vectors. This is done by setting up the null hypothesis that there are n-2 cointegrating vectors and then seeing if we can reject this in favour of the alternative of n-1. So our required statement becomes the alternative and of course failing to reject the null does not reject the alternative. The Geweke(1977) test on the other hand sets up the null that a group of n series has only one common factor and then tries to reject this against the alternative of more than one common factor, so in this case our required result is the null. Failing on this test procedure is then a rejection of our basic requirement of a single common factor while failing to meet the required cointegrating rank is merely a failure to reject an alternative null assumption.

We begin by examining the cointegrating properties of the data in table 1. Here for each group we set out the null hypothesis that the cointegrating rank is n-2 and we hope to reject this in favour of n-1, which would imply a single stochastic trend. The results in table 1 show that only 5 of the 14 groups actually allow us to reject the null in favour of our alternative but even in the other cases the test statistics are relatively large and in almost all cases it is over half the critical value. We must therefore conclude on the basis of this evidence that we can not reject the possibility that there is more than one stochastic trend in some of the groups but the evidence does seem broadly sympathetic to this conclusion.

Table II presents the evidence using the alternative approach of a common factor test (here we have aggregated the smaller groups to produce more uniform group sizes). The null hypothesis in this case is that there is only one common factor and we would reject this against the alternative of more than one common factor if the χ^2 5 test exceeded its 5% critical value. The probability value of the test shows that none of the subgroups reach this critical value and that the test across the groups also is within the critical value. So we can not reject the hypothesis that there is only one stochastic trend in each group and across the groups. This provides stronger support for the conclusion from the cointegrating test reached above.

We now turn to the cointegration properties of the complete panel of data. Following the recommendations of Pesaran and Smith we estimate a cointegrating vector for each of the 21 firms in Table III, almost all the firms pass the test for cointegration and even those which do not pass are close to the critical value. The coefficients on both wages and sales differ widely across the forms however, although all are correctly signed. The Pesaran and Smith result would therefore lead us to expect that the aggregate levels equation would not cointegrate. At the bottom of the table we present the results for an estimate of the cointegrating vector on the aggregate data (as we are dealing with logged data we consider both the sum of the logged data and the more conventional log of the aggregate data). Both measures of the aggregate model convincingly pass the test for cointegration and there coefficients are clearly reasonable, the sum of log coefficient on wages for example is -1.23 while the average over the individual firms is -1.31. So as expected, given our finding of a single common stochastic trend underlying both sales and wages, the aggregate data seems to cointegrate and aggregate estimation seems to provide reasonable estimates of the long run aggregate parameter.

5 CONCLUSIONS

In this paper we have addressed the question of panel data estimation and aggregation when the data is non-stationary. Recent work suggests that aggregate relationships may perform very poorly if the micro relationships are cointegrated but with different cointegrating vectors. We accept this result but argue that there is a relevant special case where the exogenous variables in the micro relationships are all driven by a single stochastic trend. We argue that this is an empirically relevant special case by outlining a testing procedure for this condition and showing that a well known panel data set conforms to this condition. In that case aggregate estimation seems to perform well, as expected.

The existence of single common stochastic trends across a range of micro variables may well explain the relative success of aggregate estimation.

REFERENCES

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series", In D.J. Aigner and A.S. Goldberger eds, Latent Variables in Socio-Economic Models, North Holland, Amsterdam, Ch. 19.

Johansen S.(1988), "Statistical analysis of Cointegrating Vectors" Journal of Economic Dynamics and Control. 12, 231-254.

Lee,K., Pesaran,M.H. and Pierse,R.G. (1990), "Testing for Aggregation Bias in Linear Models", Economic Journal, 100, 137-150.

Lovell, C.A.K.(1973)"A Note on Aggregation Bias and Loss" Journal of Econometrics,1, 310-311.

Malinvaud,E. (1956), "L'aggregation dans les modeles economiques", Cahiers du Seminaire d'Econometrie, 4:69-146.

Nickell,S.J. (1986), "Dynamic Models of Labour Demand", in (eds) O. Ashenfelter and Layard,R., HANDBOOK OF LABOR ECONOMICS, North-Holland.

Pesaran,M.H., Pierse,R.G. and Kumar,M.S. (1989), "Econometric Analysis of Aggregation in the Context of Linear Prediction Models", Econometrica, 57, 861-888.

Pesaran,M.H. and Smith,R.(1995), "Estimating Long-run Relationships from Dynamic Heterogeneous Panels", Journal of Econometrics, 68, 79-113.

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

Zellner,A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of the American Statistical Society, 57, 348-368.