Department of Economics

Determining the Poolability of Individual Series in Panel Datasets

George Kapetanios

Working Paper No. 499 July 2003 ISSN 1473-0278

Queen Mary
University of London

Determining the Poolability of Individual Series in Panel Datasets

George Kapetanios*
Queen Mary, University of London

May 2003

Abstract

Panel datasets have been increasingly used in economics to analyse complex economic phenomena. One of the attractions of panel datasets is the ability to use an extended dataset to obtain information about parameters of interest which are assumed to have common values across panel units. However, the assumption of poolability has not been studied extensively beyond tests that determine whether a given dataset is poolable. We propose a method that enables the distinction of a set of series into a set of poolable series for which the hypothesis of a common parameter subvector cannot be reject and a set of series for which the poolability hypothesis fails. We discuss its theoretical properties and investigate its small sample performance for a particular simple model in a Monte Carlo study.
Keywords: Panel datasets, Poolability, Sequential testing JEL Codes: C12, C15, C23

1 Introduction

Panel datasets have been increasingly used in economics to analyse complex economic phenomena. One of the attractions of panel datasets is the ability to use an extended dataset to obtain information about parameters of interest which are assumed to have common values across panel units.

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk

However, the assumption of poolability, i.e. the validity of the assumption that panel units described by a given model have a common parameter subvector for that model, has not received great attention in the literature. Work in this area has concentrated on whether a given dataset is poolable as a whole, i.e, whether the null hypothesis $H_0 : \beta_j = \beta$, $j = 1, \dots, N$ holds, where β is the assumed common parameter subvector of the N cross-sectional units of the dataset. In that vein a common approach, discussed, in some detail, in Baltagi (2001), is to use an extension of the Chow (1960) parameter stability test on the pooled dataset. Other tests for this null hypothesis have been developped by Ziemer and Wetzstein (1983) and Baltagi, Hidalgo, and Li (1996).

However, if such tests reject the researcher is left with little idea of how to proceed. In other words if we reject this null hypothesis we do not know which series caused the rejection. It would be of some interest if a method were available that would enable the distinction of the set of series into a group of poolable and a group of nonpoolable series. Such methods seem indeed possible, and asymptotically justified under appropriate conditions, and this paper is proposing one. Our method uses a sequence of tests to distinguish between poolable and nonpoolable series. If more than one series are actually poolable then the use of panel methods to investigate the properties of this set of series is indeed more efficient compared to univariate methods.

The method we propose starts by testing the null of all series having a common parameter subvector. In doing that we propose a new test of poolability. Of course, any other poolability test can be used but the new test is useful in that it produces as a by product the means of distinguishing poolable from nonpoolable series. If the test rejects the null hypothesis of poolability, then the series with the maximum difference between the individual estimate of the vector β and its estimate obtained using the pooled dataset, suitably normalised, is considered as non-poolable and is removed from the dataset. We then apply the poolability test to the remaining series and continue in this vein until the poolability test does not reject the null hypothesis for some subset of the original set of series or we are left with a set of one series.

The paper is structured as follows: Section 2 discusses the proposed test of poolability. Section 3 discusses the new method of separating poolable from nonpoolable series. Section 4 provides a Monte Carlo study. Section 5 concludes.

2 The new poolability test

Let us consider the following panel data model

$$y_{j,t} = \alpha_j + \beta_j x_{j,t} + \epsilon_{j,t}, \quad j = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{1}$$

where $x_{j,t}$ is a k-dimensional vector of predetermined variables. This is a standard panel data model where we do not need to specify the nature of the cross sectional individual effect α_j. Our discussion carries through both for fixed and random effect models. The poolability test is concerned with the null hypothesis

$$H_0 : \beta_j = \beta, \quad \forall j \tag{2}$$

We make the following assumptions

Assumption 1 *There exists an efficient, $\sqrt{NT}$-consistent, asymptotically normal estimator for β under the null hypothesis, denoted by $\tilde{\beta}$. There exists an asymptotically normal, $\sqrt{T}$-consistent estimator for the individual β_j using only data on the j-th unit. This estimator is inefficient under the null hypothesis and is denoted by $\hat{\beta}_j$.*

Assumption 2 *$x_{j,t}$ are independent across j. $\epsilon_{j,t}$ are i.i.d across t and independent across j with finite variances σ_j^2.*

Assumption 1 is likely to imply that the processes $y_{j,t}$ and $x_{j,t}$ are stationary. However, the methods we will propose does not rely on stationarity of the data although dealing with nonstationarity will require changes to the derivation of the asymptotic distributions of our test. We do not give more details on these estimators so as to encompass as general a framework as possible, in our discussion. It is clear that the model may be either dynamic (by including lags of $y_{j,t}$) or static. The assumption of model linearity is not essential, either, for the ensuing analysis. Extension to vector $y_{j,t}$ processes is also straightforward as will become obvious from what follows.

A test that $\beta_j = \beta$ for a given j may be based on the test statistic

$$S_{T,j} = (\hat{\beta}_j - \tilde{\beta})' Var(\hat{\beta}_j - \tilde{\beta})^{-1} (\hat{\beta}_j - \tilde{\beta}) \tag{3}$$

This is a Haussman type statistic. Given efficiency of the estimator $\tilde{\beta}$ under the null hypothesis we know from Hausman (1978) that

$$Var(\hat{\beta}_j - \tilde{\beta}) = Var(\hat{\beta}_j) - Var(\tilde{\beta}) \tag{4}$$

A consistent estimate of $Var(\hat{\beta}_j - \tilde{\beta})$ may be based on consistent estimates of the variances on the RHS of (4). Then it follows from our assumption of asymptotic normality of the estimators that as $T \to \infty$

$$S_{T,j} \xrightarrow{d} \chi^2_k \tag{5}$$

Our poolability test will be based on the $S_{T,j}$ statistics. In particular we suggest that $S^s_T = sup_j S_{T,j}$ be used as a test statistic for the test of the null hypothesis H_0. The following theorem discusses the asymptotic properties of our new test

Theorem 1 *Under assumption 1 and as $T \to \infty$ we have the following: Under the null hypothesis: (i) If N is fixed then S^s_T has a nuisance parameter free distribution depending only on N and k. (ii) If $N \to \infty$ then $b_N S^s_T + a_N$ has a cumulative density function given by $e^{e^{-x}}$ where choices for a_N and b_N are given in the appendix. Under the alternative hypothesis, that at least one β_j is not equal to the rest of the β_j, the test is consistent.*

The proof may be found in the Appendix.

Remark 1 *We note a few facts about the distribution with cdf $e^{e^{-x}}$ which is usually referred to as the extreme value distribution. Its probability density function is given by $e^{-x-e^{-x}}$. Its cumulants are given by $\kappa_r = (-1)^r \psi^{(r-1)}(1)$, where $\psi^{(r)}$ is the r-th polygamma function, i.e. $\frac{d^r \ln \Gamma(x)}{dx^r}$. So, $E(X) = 1 + \gamma$ where γ is Euler's constant ($\gamma \simeq 0.57722$) and $Var(X) = 1/6\pi^2$.*

Remark 2 *We note that the assumption of cross sectional independence may be relaxed. However, two things are needed for the N-asymptotic results reported in Theorem 1 to hold. First, we need some structure for the cross sectional dependence resulting in an appropriately sorted sequence, $\{S_{j',T}\}_{j'}$ which is mixing in the sense of definition 3.7.1 of Galambos (1978) (this is similar to standard strong mixing but applied only to the upper tail of the relevant distribution). Secondly we need a bound on the probability that any two $S_{j,T}$ will take large values. A possible bound is given by expression (3.62) of Galambos (1978).*

For finite N, critical values of the nuisance parameter free distribution may be obtained by simulation. 5% critical values for a selection of k and N are given in table 1.

3 The new method of separating poolable from nonpoolable series

For further use define the following. Let $\mathbf{Y_i} = (\mathbf{y}_{j_1}, \ldots, \mathbf{y}_{j_M})$, $\mathbf{i} = \{j_1, \ldots, j_M\}$ and $\mathbf{t_i} = (t_{j_1,T}, \ldots, t_{j_M,T})'$. Also define $\mathbf{i}^j = \{j\}$, $\{1, \ldots, N\} \equiv \mathbf{i}^{1,N}$ and $\mathbf{i}^{-j}$ such that

$$\mathbf{i}^{-j} \cup \mathbf{i}^j = \mathbf{i}$$

We now define the object we wish to estimate. To simplify the analysis we assume that there exists one cluster of series with equal $\beta_j = \beta$. If all series have different β_j then without loss of generality we assume that $\beta_1 \equiv \beta$. For the time being we will assume that there exists just one cluster of series with equal β_j and all the rest of the series have different β_j. The more general case is straightforward to deal with and will be discussed briefly later. For every series $y_{j,t}$ (and associated set of predetermined variables $x_{j,t}$) define the binary object $\mathcal{I}_j$ which takes the value 0 if $\beta_j = \beta$ and 1 if $\beta_j \neq \beta$. Then, $\mathcal{I}_\mathbf{i} = (\mathcal{I}_{j_1}, \ldots, \mathcal{I}_{j_M})'$. We wish to estimate $\mathcal{I}_{\mathbf{i}^{1,N}}$. We denote the estimate by $\hat{\mathcal{I}}_{\mathbf{i}^{1,N}}$.

To do so we consider the following procedure.

1. Set $j = 1$ and $\mathbf{i}_j = \{1, \ldots, N\}$.

2. Calculate the S_T^s-statistic for the set of series $\mathbf{Y}_{\mathbf{i}_j}$. If the test does not reject the null hypothesis $\beta_i = 0$, $i \in \mathbf{i}_j$, stop and set $\hat{\mathcal{I}}_{\mathbf{i}_j} = (0, \ldots, 0)'$. If the test rejects go to step (3).

3. Set $\hat{\mathcal{I}}_{\mathbf{i}^l} = 1$ and $\mathbf{i}_{j+1} = \mathbf{i}_j^{-l}$, where l is the index of the series associated with the minimum $S_{T,s}$ over s. Set $j = j + 1$. Go to step (2).

In other words, we estimate a set of binary objects that indicate whether a series is poolable or not. We do this by carrying out a sequence of poolability tests on a reducing dataset where the reduction is carried out by dropping series for which there is evidence of nonpoolability. A large individual $S_{j,T}$-statistic is used as such evidence.

We will discuss conditions for the consistency of $\hat{\mathcal{I}}_{\mathbf{i}^{1,N}}$ as an estimator of $\mathcal{I}_{\mathbf{i}^{1,N}}$, for infinite N and T. We denote the number of poolable series by N_1 and the number of nonpoolable series $N_2 = N - N_1$. We need the following assumption

Assumption 3 *There exists $0 < \alpha < 1$ such that if the number of series for which $\beta_j \neq \beta$ is at most $O(N^\alpha)$ the panel estimator of β using all series is consistent.*

Remark 3 *Assumption 3 is needed because if the panel estimator is not consistent then it is not possible to guarantee that no poolable series will removed from the dataset before all nonpoolable ones have been removed. For example, it is straightforward to show that for exogenous regressors,* $x_{j,t}$ *and OLS estimation both for panel datasets and individual series if* $N_2/N_1 \rightarrow 0$ *then the panel estimator is consistent.*

Formally, we will show that

Theorem 2 *Under assumptions 1 and 2 and if (i)* $\lim_{T\rightarrow\infty} \alpha_T \rightarrow 0$ *(ii)* $\lim_{T\rightarrow\infty} \ln \alpha_T/T = 0$*, where* α_T *is the significance level used for the poolability test and (iii)*$N \rightarrow \infty$*, and* N_2 *satisfies assumption 3, then*

$$\lim_{N,T\rightarrow\infty} Pr(\sum_{j=1}^{N} |\hat{\mathcal{I}}_{\mathbf{i}^j} - \mathcal{I}_{\mathbf{i}^j}| > 0) = 0 \tag{6}$$

Note the similarities between this setup and the variety of tests of rank where a sequence of tests are needed to determine the rank of a matrix (see e.g., Camba-Mendez and Kapetanios (2001) or Camba-Mendez, Kapetanios, Smith, and Weale (2003)).

A weaker result can be established for a fixed significance level, α.

Theorem 3 *Under assumptions 1 and 2,* $N \rightarrow \infty$*, and if* N_2 *satisfies assumption 3, then*

$$\lim_{N,T\rightarrow\infty} Pr(|\hat{\mathcal{I}}_{\mathbf{i}^j} - \mathcal{I}_{\mathbf{i}^j}| > 0) = 0, \forall j \tag{7}$$

It is clear that our procedure is very general. It can be applied using any poolability test. The main ingredients are a poolability test and a criterion for choosing which series to classify as nonpoolable at each step. Our choice of using the maximum difference between an estimator of the parameter vector from an individual series and one from the pooled dataset seems uncontroversial. The new poolability test is ideally suited for this choice of metric and is therefore adopted.

Before concluding this section, we discuss the possible case of two or more different clusters of series in the panel dataset each with a common value of β within the cluster but different across clusters. Clearly the clusters need to be of different orders of magnitude to satisfy assumption 3. Our procedure will inevitably end with the larger cluster being classified as poolable and the other cluster as nonpoolable. In one sense this is the correct conclusion since the two clusters should not be pooled. On the other hand, the cluster which is not selected as a poolable cluster, should still be analysed using panel methods. As a result it is possible to restart the procedure with only the series which have been classified as nonpoolable from the initial application of the procedure. Then, the second cluster will be identified. The sequential application of the method can be continued until all potential clusters have been identified.

4 Monte Carlo Study

In this section we carry out a Monte Carlo investigation of our new method. We consider the following setup. Let

$$y_{j,t} = \phi_j y_{j,t-1} + \epsilon_{j,t}, j = 1, \ldots, N, \quad t = 1, \ldots, T \tag{8}$$

where $\epsilon_{j,t} \sim N(0,1)$. We investigate the new method along a number of different dimensions for the above model. Namely, we consider variations in N, T and ϕ_j. More specifically, we consider $T \in \{50, 100, 200, 400\}$ and $N \in \{5, 10, 15, 20, 25, 30\}$.

For ϕ_j we consider the following setup: $\phi_j = 0.5$ with probability δ over j and $\phi_j \in (\gamma_1, \gamma_2)$ with probability $1 - \delta$. This is a general setup designed to address a number of issues not widely discussed in the literature. Obviously, the degree of variation in ϕ_j under the alternative hypothesis is of great importance. Further, the choice of δ is likely to affect the performance of the new method. We set $\delta \in \{0.25, 0.5, 0.75\}$.

We choose $\gamma_1 = 0.05$ and $\gamma_2 = 0.95$. The performance measure we use is the estimated probability of classifying a series as nonpoolable. This should tend to zero for poolable series and to one for nonpoolable series. Denote the number of Monte Carlo replications by B. This probability is calculated as follows in our experiments.

$$\hat{P}(\mathcal{I}_{\mathbf{i}^u} = 1 | \mathcal{I}_{\mathbf{i}^u} = s) = \frac{1}{N_s B} \sum_{b=1}^{B} \sum_{\mathcal{I}_{\mathbf{i}^q} = s} \hat{\mathcal{I}}^b_{\mathbf{i}^q} \tag{9}$$

where $N_s = N(1-\delta)s + N\delta(1-s)$ and u denotes a generic series. Results are presented in Table 2. We refer to the new method as Sequential Panel Selection Method (SPSM).

We also carry out a Monte Carlo evaluation of the new test. The setup is the same as above. Tables 3 and 4 present the probability of rejection under the null hypothesis of $\phi_j = 0.5$ and under the alternative hypothesis where $\phi_j = 0.5$ with probability δ over j and $\phi_j \in (\gamma_1, \gamma_2)$ with probability $1-\delta$. We choose $\gamma_1 = 0.25$ and $\gamma_2 = 0.75$ for these experiments. Results for the null hypothesis are presented in Table 3 and for the alternative in Table 4.

A number of conclusions emerge from these Tables. First, we comment on the performance of the SPDM method. We note that the performance of SPSM in terms of classifying poolable series as poolable is in general satisfactory. The probability of misclassification never exceeds 15%. This is to be expected given that the method is based on a test whose null hypothesis is that of a set of series being poolable. On the other hand, as the number of observations increases we see that this probability falls mainly for $\delta = 0.50, 0.75$. This probability also falls with the number of series included in the dataset. This is in line with the asymptotic result in Theorem 3. For example, we see that for $N = 30$, $T = 400$ and $\delta = 0.75$ this probability is only 0.2%.

Moving on to the ability of SPSM to classify nonpoolable series as nonpoolable we see that the probability of that happening increases drastically with T. It does not seem to be affected by N or δ. Given that the test is based on a supremum of a set of statistics this is perhaps to be expected since only the behaviour of one series matters for the test in each sample. We note here the difference between our conclusion and that reached by Kapetanios (2003) where similar techniques are advocated to separate I(0) from I(1) series in a panel. There the test used is based on an average over a set of Dickey Fuller statistics. As a result that procedure is materially affected by the choice of δ and N.

Moving on to the properties of the S_T^s test we see that it is very well behaved under the null hypothesis. The estimated rejection probability never differs from 5% by more than 1.3% except for one case where it is equal to 7.5% but for a sample of 30 observations. The power of the test increases with T, N and decreases with δ.

5 Conclusions

The use of panel datasets for the investigation of a number of economic phenomena has been increasing recently. Both the availability of larger datasets and the development of new estimation methods methods specifically designed for panel datasets can account for this.

An important advantage of panel methods is their ability to improve inference compared to single unit methods. Nevertheless, this implies that the parameter restrictions implied by the panel structure are valid. Poolability tests exists to help with this problem but if they reject the null hypothesis of poolability the researcher is often uncertain about the cause of the rejection, or in particular about the identity of the series that caused this rejection. In other words a method that could distinguish poolable from nonpoolable series within a panel dataset would be of interest to empirical researchers.

This paper has suggested such a method. It is based on the sequential use of a poolability test combined with a criterion for removing series one at a time from the dataset when the test rejects. In our implementation the maximum difference between an individual and a panel estimator has been used as such a criterion. Although, we have developed the formal components of our method using a particular new poolability test it is clear that similar methods can be developed based on other pooolability tests. Our Monte Carlo analysis has clearly shown that both the new test and new method work satisfactorily. Further research can illustrate both the use of the new method in empirical contexts and the potential for alternative poolability tests to give rise to methods that improve upon the results reported here.

Appendix

Proof of Theorem 1

As $T \to \infty$, each of $S_{T,j}$ tends to a χ_k^2 random variable as discussed in the text. For finite N, it is then obvious that the $T-$asymptotic distribution of S_T^s will be nuisance parameter free under the null hypothesis and will depend only on N and k. The above relies on independence of the individual test statistics which follows from the efficiency of the panel estimator under the null hypothesis under Assumption 1 and independence of $x_{j,t}$ and $\epsilon_{j,t}$ over j. Critical values are reported in Table 1.

If $N \to \infty$ as well, we can characterise the asymptotic distribution of S_T^s using sequential asymptotics. In particular, we allow $T \to \infty$ and then $N \to \infty$. For $T \to \infty$, we have a set of N χ_k^2 distributed random variables. As $N \to \infty$ these random variables become independent since the rate of convergence of $\tilde{\beta}$ is $\sqrt{TN}$ whereas the rate of convergence of $\hat{\beta}_j$ is only $\sqrt{T}$ for all j. So we want the asymptotic distribution of the supremum of a set of N independent χ_k^2 random variables as $N \to \infty$.

To obtain this we use results from the asymptotic theory of extreme order statistics. Following Arnold, Balakrishnan, and Nagaraja (1992) or Galambos (1978), there exist only three forms for the asymptotic cumulative distribution function of an appropriate normalisation of this statistic, given by $b_N S_T^s + a_N$. It is not always the case that such an asymptotic representation exists. These cumulative distributions are given by

$$G_1(x, \alpha) = \begin{cases} 0 & \text{if} x \leq 0 \\ e^{-x^{-\alpha}} & x > 0; \alpha > 0 \end{cases} \tag{10}$$

$$G_2(x, \alpha) = \begin{cases} e^{-(-x)^{\alpha}} & x < 0; \alpha > 0 \\ 1 & x \geq 0 \end{cases} \tag{11}$$

and

$$G_3(x) = e^{-e^{-x}} \tag{12}$$

According to Theorem 8.3.2 of Arnold, Balakrishnan, and Nagaraja (1992) the distribution can be of the form G_2 only if $F_k^{-1}(1)$ is finite where $F_k(.)$ is the cdf of a χ_k^2 random variable. Since $F_k^{-1}(1) = \infty$, G_2 is not the form of the asymptotic distribution.

The distribution is G_1 iff the following condition applies

$$\lim_{t \to \infty} \frac{1 - F_k(tx)}{1 - F_k(t)} = x^{-\alpha} \tag{13}$$

But by applying L'Hopital's rule we can easily see that this limit is infinity for $x > 1$ using $F_k(x) = \Gamma_{x/2}(k/2)/\Gamma(k/2)$ and $f_k(x) = \frac{1}{2^{k/2}\Gamma(k/2)} e^{-x/2} x^{(k/2)-1}$ where $\Gamma_a(b) \equiv \int_0^a e^{-t} t^{a-1} dt$ is the incomplete Gamma function. Thus, we need to either verify that G_3 is the appropriate distribution or conclude that no such distribution exists.

To check whether G_3 is the appropriate distribution we use the third von Mises condition given in Theorem 8.3.3 of Arnold, Balakrishnan, and Nagaraja (1992). This condition states that the asymptotic distibution is G_3 iff

$$\lim_{x \to F^{-1}(1)} \frac{d}{dx} \left\{ \frac{1 - F_k(x)}{f_k(x)} \right\} = 0 \tag{14}$$

where $f_k(x)$ is the pdf of a χ^2_k random variable.

The above condition is equivalent to

$$\lim_{x \to \infty} \frac{-f''_k(x)(1 - F_k(x)) + f'_k(x) f(x)}{2 f_k(x) f'_k(x)} = 1 \tag{15}$$

where $f'_k(x)$ and $f''_k(x)$ are the first and second derivatives of $f(x)$. Then, it is easy to see that we need to prove

$$\lim_{x \to \infty} \frac{-f''_k(x)(1 - F_k(x))}{f_k(x) f'_k(x)} = 1 \tag{16}$$

Simple algebra indicates that for the χ^2_k pdf

$$\lim_{x \to \infty} \frac{-f''_k(x)}{f'_k(x)} = 1/2 \tag{17}$$

Further, by a double application of L'Hopital's rule, it follows that

$$\lim_{x \to \infty} \frac{1 - F_k(x)}{f_k(x)} = 2 \tag{18}$$

proving that the required distribution is indeed G_3. Then, by part (iii) of theorem 8.3.4 of Arnold, Balakrishnan, and Nagaraja (1992) we have that possible (but not unique) expressions for a_N and b_N are given by:

$$a_N = F^{-1}(1 - N^{-1}), \;\; b_N = F^{-1}(1 - (Ne)^{-1}) - F^{-1}(1 - N^{-1}) \text{ or } b_N = [N f(a_N)]^{-1} \tag{19}$$

Finally, example 8.3.4 of Arnold, Balakrishnan, and Nagaraja (1992) implies that $b_N \sim \log N$ and that $S^s_T = O_p(\log N)$.

Consistency of the test readily follows since, under the alternative, $S_{T,j} = O_p(T)$ and, therefore, $S^s_T = O_p(T \log N)$.

Proof of Theorem 2

The theorem follows from the following considerations. For all $\hat{\mathcal{I}}_{\mathbf{i}_j}$ such that for some $l \in \mathbf{i}_j$, $\mathcal{I}_{\mathbf{i}^l} = 1$ we know that the poolability test on the set of series $\mathbf{Y}_{\mathbf{i}_j}$ will reject with probability 1 by the consistency of the poolability test and condition (ii) of Theorem 2 combined with standard arguments on sequences of tests as discussed in , e.g., Hosoya (1989). Consistency of the poolability test follows from Theorem 1. This implies that S_T^s is at least $O_p(T^{1/2})$ even for one nonpoolable series in the panel. Further, we know using assumption 3 that with probability 1, $S_{l,T} > S_{m,T}$ asymptotically if $\mathcal{I}_{\mathbf{i}^l} = 1$ and $\mathcal{I}_{\mathbf{i}^m} = 0$. As a result, all series for which $\mathcal{I}_{\mathbf{i}^l} = 1$ will be identified as such, by the sequential approach with probability approaching 1. By condition (i) of Theorem 2 we know that if $\mathcal{I}_{\mathbf{i}^l} = 0$ for all j in $\mathbf{i}_j$ then the poolability test will reject with probability equal to $\alpha_T \to 0$.

Proof of Theorem 3

We start by noting that, by assumption 3, with probability 1 all series for which $\mathcal{I}_{\mathbf{i}^l} = 1$ will be detected as nonpoolable by the sequential test before any series for which $\mathcal{I}_{\mathbf{i}^l} = 0$. This is because the individual $S_{j,T}$ tests for nonpoolable series are $O_p(T^{1/2})$ whereas they are $O_p(1)$ for all nonpoolable series. When all nonpoolable series have been removed from the dataset, a poolability test will be carried out on a set of poolable series. With probability α this test will reject. In general, with, at most, probability $\tilde{\alpha}^k$, k or more redundant poolability tests will be carried out. Note that $\tilde{\alpha}$ may be different from α as the sequence of tests is not made up of independent tests. However, it is guaranteed that $\tilde{\alpha} < 1$. Therefore, the probability that k poolable series are missclassified as nonpoolable is $O(\tilde{\alpha}^k)$ and tends to zero exponentially with k. Thus, for any given series, out of the N_1 poolable series, the probability that it will be missclassified as nonpoolable tends to zero.

References

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992): *A First Course in Order Statistics*. Wiley.

Baltagi, B. H. (2001): *Econometric Analysis of Panel Data*. Wiley.

Baltagi, B. H., J. Hidalgo, and Q. Li (1996): "A Nonparametric Test for Poolability Using Panel Data," *Journal ofEconometrics*, 75(1).

Camba-Mendez, G., and G. Kapetanios (2001): "Testing the Rank of the Hankel Covariance Matrix: A Statistical Approach," *IEEE Transactions on Automatic Control*, 46(2), 331–336.

Camba-Mendez, G., G. Kapetanios, R. J. Smith, and M. Weale (2003): "Tests of Rank in Reduced Rank Regression Models," *Journal of Business and Economic Statistics*, 23(1).

Chow, A. (1960): "Tests of Equality Between Sets of Coefficients in two Linear Regressions," *Econometrica*, 28, 591–605.

Galambos, J. (1978): *The Asymptotic Theory of Extreme Order Statistics*. Wiley.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(1).

Hosoya, Y. (1989): "Hierarchical Statistical Models and a Generalised Likelihood Ratio Test," *Journal of the Royal Statistical Society B*, 51(3), 435–448.

Kapetanios, G. (2003): "Determining the Stationarity Properties of Individual Series in Panel Data," Mimeo, Queen Mary, University of London.

Ziemer, R. F., and M. E. Wetzstein (1983): "A Stein Rule Method for Pooling Data," *Economics Letter*, 11(1).

Table 1: Simulated 5% critical values for the S_T^s test

	S_T^s									
N/k	1	2	3	4	5	6	7	8	9	10
1	4.021	5.943	7.765	9.324	11.157	12.513	14.174	15.574	16.696	18.414
2	5.048	7.391	9.299	11.195	12.870	14.495	16.158	17.636	19.033	20.367
3	5.633	8.236	10.230	12.007	14.007	15.396	17.100	18.766	20.022	21.873
4	6.330	8.790	10.771	12.665	14.525	15.973	17.868	19.518	21.083	22.383
5	6.585	9.242	11.189	13.329	14.927	16.742	18.445	19.944	21.638	23.295
6	6.801	9.469	11.668	13.577	15.282	17.135	19.021	20.654	22.112	23.747
7	7.152	9.830	12.010	13.959	15.941	17.717	19.299	20.990	22.876	23.838
8	7.496	10.005	12.163	14.413	16.297	17.893	19.762	21.211	22.870	24.635
9	7.675	10.265	12.518	14.419	16.514	18.311	20.120	21.468	23.221	24.757
10	7.794	10.681	12.870	14.801	16.708	18.278	20.122	21.691	23.684	25.158
15	8.481	11.372	13.509	15.665	17.664	19.452	21.262	22.832	24.644	26.472
20	9.129	11.948	14.078	16.494	18.038	20.318	21.916	23.608	25.473	26.904
25	9.534	12.328	14.809	16.945	18.859	20.926	22.420	24.367	25.957	27.662
30	9.751	12.785	15.147	17.212	19.437	21.146	22.967	24.551	26.576	28.096
35	10.041	13.147	15.482	17.776	19.674	21.537	23.342	25.173	26.835	28.575
40	10.458	13.317	15.864	18.111	19.989	21.848	23.826	25.470	27.247	28.982
45	10.593	13.561	16.069	18.239	20.166	22.213	23.944	25.774	27.680	29.262
50	10.782	13.851	16.114	18.427	20.360	22.337	24.271	26.063	27.760	29.365

Table 2: SPSM[a]

%Poolab	(N,T)	50	100	200	400
	5	$\binom{0.062}{0.231}$	$\binom{0.116}{0.349}$	$\binom{0.150}{0.490}$	$\binom{0.150}{0.611}$
	10	$\binom{0.038}{0.233}$	$\binom{0.081}{0.376}$	$\binom{0.128}{0.516}$	$\binom{0.152}{0.640}$
0.25	15	$\binom{0.029}{0.235}$	$\binom{0.056}{0.384}$	$\binom{0.102}{0.546}$	$\binom{0.107}{0.661}$
	20	$\binom{0.014}{0.220}$	$\binom{0.039}{0.392}$	$\binom{0.074}{0.552}$	$\binom{0.080}{0.672}$
	25	$\binom{0.009}{0.221}$	$\binom{0.016}{0.389}$	$\binom{0.031}{0.559}$	$\binom{0.035}{0.689}$
	30	$\binom{0.009}{0.218}$	$\binom{0.017}{0.383}$	$\binom{0.026}{0.557}$	$\binom{0.023}{0.685}$
	5	$\binom{0.036}{0.228}$	$\binom{0.034}{0.417}$	$\binom{0.031}{0.574}$	$\binom{0.025}{0.695}$
	10	$\binom{0.018}{0.246}$	$\binom{0.025}{0.406}$	$\binom{0.024}{0.573}$	$\binom{0.015}{0.693}$
0.50	15	$\binom{0.011}{0.228}$	$\binom{0.010}{0.408}$	$\binom{0.010}{0.577}$	$\binom{0.007}{0.704}$
	20	$\binom{0.007}{0.221}$	$\binom{0.007}{0.396}$	$\binom{0.006}{0.573}$	$\binom{0.004}{0.701}$
	25	$\binom{0.004}{0.222}$	$\binom{0.004}{0.396}$	$\binom{0.005}{0.569}$	$\binom{0.004}{0.699}$
	30	$\binom{0.004}{0.220}$	$\binom{0.005}{0.390}$	$\binom{0.003}{0.567}$	$\binom{0.003}{0.690}$
	5	$\binom{0.021}{0.241}$	$\binom{0.017}{0.425}$	$\binom{0.013}{0.601}$	$\binom{0.013}{0.726}$
	10	$\binom{0.010}{0.245}$	$\binom{0.009}{0.446}$	$\binom{0.006}{0.601}$	$\binom{0.005}{0.716}$
0.75	15	$\binom{0.006}{0.254}$	$\binom{0.006}{0.416}$	$\binom{0.004}{0.593}$	$\binom{0.004}{0.708}$
	20	$\binom{0.004}{0.231}$	$\binom{0.003}{0.403}$	$\binom{0.003}{0.569}$	$\binom{0.002}{0.701}$
	25	$\binom{0.004}{0.224}$	$\binom{0.003}{0.397}$	$\binom{0.002}{0.567}$	$\binom{0.002}{0.697}$
	30	$\binom{0.003}{0.218}$	$\binom{0.002}{0.392}$	$\binom{0.002}{0.559}$	$\binom{0.002}{0.695}$

[a] %Poolab denotes the proportion of series which are poolable. For the notation $\binom{a}{b}$ we have that a gives the probability that an poolable series will be classified as nonpoolable, whereas b gives the probability that an nonpoolable series will be classified as nonpoolable.

Table 3: Rejection Probability Under the Null[a]

		S_T^s			
%Poolab	(N,T)	50	100	200	400
1	5	0.052	0.047	0.054	0.052
	10	0.063	0.049	0.049	0.046
	15	0.066	0.052	0.047	0.054
	20	0.065	0.057	0.047	0.058
	25	0.062	0.063	0.046	0.057
	30	0.074	0.059	0.049	0.059

[a]%Poolab denotes the proportion of series which are poolable.

Table 4: Rejection Probability Under the Alternative[a]

		S_T^s			
%Poolab	(N,T)	50	100	200	400
0.25	5	0.250	0.473	0.689	0.867
	10	0.385	0.693	0.899	0.991
	15	0.493	0.812	0.973	0.999
	20	0.543	0.860	0.991	1.000
	25	0.547	0.878	0.995	0.999
	30	0.627	0.913	0.999	1.000
0.50	5	0.182	0.345	0.556	0.762
	10	0.322	0.569	0.852	0.965
	15	0.333	0.637	0.905	0.987
	20	0.410	0.728	0.952	0.997
	25	0.430	0.784	0.967	1.000
	30	0.509	0.828	0.988	0.999
0.75	5	0.127	0.217	0.355	0.528
	10	0.182	0.322	0.521	0.749
	15	0.219	0.360	0.659	0.853
	20	0.277	0.525	0.794	0.957
	25	0.269	0.525	0.845	0.977
	30	0.306	0.546	0.898	0.989

[a]%Poolab denotes the proportion of series which are poolable.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm