Are Panel Unit Root Tests Useful for Real-Time Data Sets

Christian GENGENBACH[1] **Alain HECQ**

Jean-Pierre URBAIN

Maastricht University, Department of Quantitative Economics

February 26, 2010

[1]Corresponding author: Christian Gengenbach, Maastricht University, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands. E-mail: c.gengenbach@maastrichtuniversity.nl.
The authors would like to thank Stephan Smeekes for provinding us the Gauss codes for the boostraped unit root tests.

Abstract

With the development of real-time databases, N vintages are available for T observations instead of a single realization of the time series process. Therefore, the implementation of panel unit root tests with the aim to gain in efficiency seems obvious. Several practical problems emerge however: *(i)* the selection of the vintage variables, *(ii)* the redefinition of the series, *(iii)* the N and T usable observations, *(iv)* the strong heterogeneity and cross-correlation of the disturbance terms or *(v)* the presence of cross-member cointegration. A set of well know panel unit root tests, from the first generation to most recent boostraped procedures, are considered in this paper. Both the outcome empirical analyses and a Monte Carlo investigation, heavily mitigate the first euphoric perspective.

Keywords: Real-time data, panel unit root tests

1 Introduction, notations and caution notes

Before being considered definitive, many data produced and published by Statistical Offices or Central Banks are initially provisional, as they are partly based on estimates. Then they are subject to later revisions when new information is available and hence figures for the same reporting period may change over time resulting in different releases of the same phenomenon. The collection of all these vintages is referred to as a real-time data set. Exploiting this type of data in the context of panel unit root tests is intuitively obvious because N vintages are available for T observations each. Patterson and Heravi (2004a) consider this idea but do not investigate the possibilities further. In this paper we illustrates to what extent several panel unit root tests provide new insights for such collection of data.

In order to fix notations, let us consider the Jacobs and van Norden (2007) triangular representation (which is a trapezoid here as often in practice) of the real-time data matrix for $N = 6$ vintages, i.e. releases of a complete sequence of historical data. Table 1 provides the notation for a generic variable y. Each column of that matrix is referred to as a vintage series and the point of time of the referred data are in the rows. Consequently, y^t_{t-1} is the estimate, i.e. one scalar, available or published at time t of the value of y for time $t-1$. In that example the data is available and published with a delay of one period. The first difference operator $\Delta = (1 - L)$ runs over both indexes with for instance $\Delta y^t_{t-1} = (1 - L)y^t_{t-1} = y^t_{t-1} - y^{t-1}_{t-2}$.

Table 1: Real-time data matrix:

$\frac{\textit{vintages}}{\textit{calendar time}}$	$t-5$	$t-4$	$t-3$	$t-2$	$t-1$	t
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t-7$	y^{t-5}_{t-7}	y^{t-4}_{t-7}	y^{t-3}_{t-7}	y^{t-2}_{t-7}	y^{t-1}_{t-7}	y^{t}_{t-7}
$t-6$	y^{t-5}_{t-6}	y^{t-4}_{t-6}	y^{t-3}_{t-6}	y^{t-2}_{t-6}	y^{t-1}_{t-6}	y^{t}_{t-6}
$t-5$	$-$	y^{t-4}_{t-5}	y^{t-3}_{t-5}	y^{t-2}_{t-5}	y^{t-1}_{t-5}	y^{t}_{t-5}
$t-4$	$-$	$-$	y^{t-3}_{t-4}	y^{t-2}_{t-4}	y^{t-1}_{t-4}	y^{t}_{t-4}
$t-3$	$-$	$-$	$-$	y^{t-2}_{t-3}	y^{t-1}_{t-3}	y^{t}_{t-3}
$t-2$	$-$	$-$	$-$	$-$	y^{t-1}_{t-2}	y^{t}_{t-2}
$t-1$	$-$	$-$	$-$	$-$	$-$	y^{t}_{t-1}
	$-$	$-$	$-$	$-$	$-$	$-$

When we collect data, on some websites for instance, for the gross domestic product at some point in time t, only the last column of Table 1 is downloaded. This is the time series people use for e.g. unit root testing. It is well known that the properties of unit root tests can be weak when the sample has a small T dimension. In order to improve size properties and the power of such test statistics we can think to rely on more information about these series and carry out panel unit root tests using additional information from previous vintages. This means that we cam exploit not only the last column of Table 1 but also the columns before as different realizations of the same process.

Before doing this, however, several practical issues must be analyzed carefully.

(i) The choice of the vintage series

We think that it is not relevant to work by vintages, namely on the verticals of the real-time data matrix. Indeed when we look at the historical series by verticals, most of observations in the past are identical and only the last few quarters or years say are different. This would mean to apply panel unit root tests on N times the same variable. It is consequently more interesting to work with the diagonals. The last diagonal provides the first estimate that is available for the whole time series. In the scheme of Table 1, it is one period after the end of the corresponding calendar time. The second diagonal gives the first revision arising two periods after the calendar time, the third the second revision, etc. In the following we will then work with these diagonals, i.e. what Patterson and Heravi (2004) call revision-stage vintages. One can also call them releases: first release, second release,...

(ii) The choice of the diagonals and the N/T trade-off

When testing for unit roots on the diagonals it is necessary to make the distinction between two approaches. In the bivariate example, one can consider working either with the process $(y_{t-2}^{t-1}, y_{t-2}^{t})'$, $t = 1 \ldots T$ or $(y_{t-1}^{t}, y_{t-2}^{t})'$, $t = 1 \ldots T$. Hecq and Jacobs (2009) call the first approach *OBS* (for *Observations Balanced System)* and the second way, to take the diagonals, *VBS,* where the acronym now stands for *Vintage Balanced System*. The pros and cons of these two approaches are detailed in Hecq and Jacobs (2009); for example, the VBS approach is robust to the introduction of a new base year (see next point). As far as the trade-off between the number of vintages and the number of observations is concerned, it should also be noticed that for the OBS approach, a part of the information is lost when comparing the diagonals for the same time dimension. For instance if we analyze two series the observation y_{t-1}^t is not included in the OBS analysis because the last observation common to the two series is for the calendar period $t-2$. In general, with N vintages one looses the triangle whose height is $N-1$. We do not have that problem in the VBS approach taking the diagonals. However, what we extract from the data is a bit different. At least the presence of cointegrating vectors is not altered though. Finally, for both approaches remains the problem of the choice of N, i.e. the number of variables to be considered. In the OBS approach, increasing N decreases T in a balanced panel. But to what extent it is better to consider 10 or 20 series is an open question.

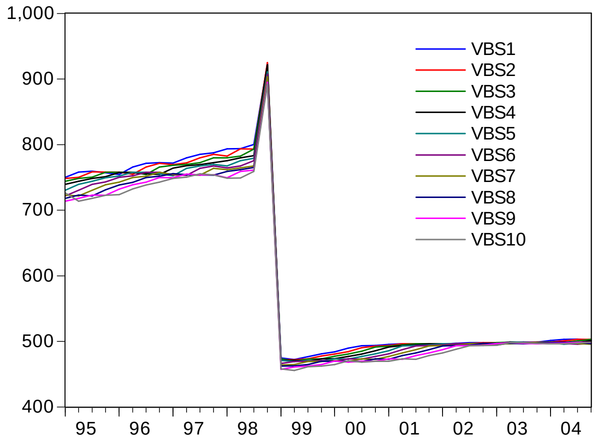
(iii) New definition of the series

One important problem for econometricians working on this type of data is that the definition of the series changes through time. For instance in our empirical application on the German real gross domestic product there is a first modification of the series occurring in 1999Q1 and a second one in 1999Q2. The first one is the introduction of a new base year, changing from 1991=100 to 1995=100. The second modification is the publication of the series in Euros and not in Deutsche Marks anymore. If we work on the diagonals we will see two level shifts around these two quarters. The advantage of the VBS is to see that they occur at the same time period (see Figure 1 for the 10 most recent diagonals), in the OBS there exists a time shift in the breaks. The treatment of these problems can be relatively easy for statistical officers but a mess for econometricians if there is no overlapping period to change the base year for instance. Moreover, in our case this also implies looking at the Euro/DM exchange rate to retropolate the series. We consider a rather ad-hoc way in this paper. The effect of more advanced techniques (see Knetsch and Reimers, 2006) is left for further investigations.

(iv) Typos and mistakes

The construction of real-time databases is a relatively new research area in the program agenda of many institutions and a new field into which almost every country jumps in. But this is a huge task (see Croushore and Stark, 2001) and some manual research must be undertaken to sometimes copy the numbers from newspapers or monthly bulletins to spread sheets. Although there exist some automatic double checks, inevitably some typos remain. For instance, we noticed one typo in one of our vintage

Figure 1: First 10 VBS diagonals for the German real GDP



and corrected it, because it seems that a 900 instead of a 800 has been typed in in the Euro/DM transformation. When we discover such a problem however (and some of them are not detectable) we should in theory contact the institution who releases the data.

(v) Cross-member cointegrating vectors

Because of the presence of many interactions, the diagonal vintages series cannot be simply pooled to get NT observations. Out of N diagonal series there likely exist $N-1$ cross-member cointegrating vectors, if we assume that the revision process is stationary, an hypothesis that is highly plausible. This is what is empirically found in most studies using Johansen's cointegration test statistics on a small number of series (see inter alia Patterson (2000, 2002a, 2002b, 2003), Patterson and Heravi (1991a, 1991b, 2004a, 2004b)). Although only $N-2$ cointegrating vectors are sometimes detected, it is not clear whether this does not reflect the problem of level shifts that we raise in point (iii). But beyond the issue of $N-1$ or $N-2$ cointegrating vectors, the presence of cross-member long-run relationships will likely lead to distortions when applying first generation panel unit root tests (those of e.g. Levin, Lin and Chu (2002) or Im, Pesaran and Shin (2003)) as well as problems for some procedures of the second generation (Bai and Ng, 2004). Moreover, the presence of a single common stochastic trend of time dimension T when $N-1$ cointegrating vectors, do not improve the performance of the unit root tests over univariate procedures.

(vi) The common stochastic trend factor

If it emerges that the cointegrating rank is $N-1$, every cointegrating vector between two successive diagonals is probably $\beta_i = (1 : -1)'$, $\forall i = 1...N-1$. Before looking at the presence of a unit root in y_t versus the alternative of stationarity around a linear deterministic trend, we propose to analyze that null hypothesis. However if this his the case, this implies that $\beta_\perp = (1 : 1 : .. : 1)'$. Now if we use the simple average to obtain the common stochastic trend, that is to say $\beta_\perp y_t$, we have what is called the Kasa (1992) decomposition that is well known not to be a Permanent-Transitory decomposition (see e.g. Mills (1998)).

(vii) To log or not to log

On a single time series for the real gross domestic product for instance, we often take the logarithms before we test for a unit root. In real-time data, it might be more sensible to consider the level of the variables. Indeed, let us consider a combination of the vintages to extract a common component. It is more intuitive to add levels than log levels.

(viii) Seasonal adjustment

A point that would deserve more attention, but that is out of the scope of this paper, is the impact of the seasonal adjustment. Indeed, once a new observation becomes available, statistical offices apply seasonal filters to the whole historical sequence. This operation can influence the comparison between vintages.

(ix) The covariance matrix of the disturbances

The estimated variance-covariance matrix of the disturbance terms for N vintages shows a very particular pattern. First, on the diagonal, the variance of the first release is quite high and might rapidly decrease for latter revisions. Second, the correlation between two vintages is very low between for first releases and subsequent vintages but that correlation is very high between revised vintages. This is indeed obvious because "older" vintages are less revised and the series are then more similar. Overall this strange structure has got in nature some problematic influence on the results.

The rest of the paper is organized as follows. Section 2 describes the data we use and the rebasement method we have considered. We apply ADF univariate tests on the German quarterly real output as well as on the German monthly industrial production index. Section 3 compares the outcome of different panel unit root tests, accounting for the presence of cross-member cointegrating vectors or not on both VBS and OBS diagonal variables. Sections 4 provides a Monte Carlo investigation that allows to shed some light on the empirical results. Section 5 concludes and provides some comments on the use of panel unit root tests for real time data sets.

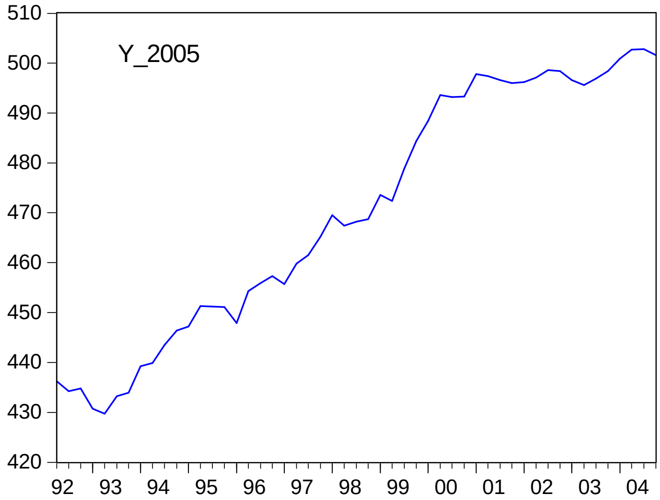
2 Data and individual unit root tests

We take two economic variables from the GERDA database. This is a comprehensive real-time database of the German economy available on the Bundesbank's website.[1] GERDA currently contains historical vintages of some 280 economic indicators from the national accounts, the monthly business and labour market surveys as well as price statistics.

The first series we take is the seasonally adjusted quarterly gross domestic product at constant price (code Q.DE.Y.A.AG1.CA010.C.A) we denote Y_t. Looking at the diagonals of that variable we build 17 usable diagonals without holes. There are quarterly vintages and there is a new base introduced in the release 1999Q2. Between 1999Q2 and 1999Q1 in April, the Bundesbank also published a vintage for the base 1995=100. We use it to convert the series before in 1995=100. Next there is a difference between vintages releases in Euros from 1999Q2 onwards and the previous releases. We simply convert the series in Euros using the fix exchange rate of 1.95583. This is of course again a rough approximation. We apply the usual Dickey-Fuller unit root tests on the last available vertical vintage (published 2005Q1). It turns out that we cannot reject the unit root null hypothesis for both the levels and the log-levels. As an example, the $p-value$ for the ADF test in the model with an intercept and a time trend is 0.59 for a value of the test of -1.99 (sample 1991Q1-2004Q1). If we now restrict the sample to 1995Q1-2004Q1 to match the sample we have for the panel unit root tests, the $p-value$ is 0.94. Moreover we reject the null of a unit root in first difference. Y_t as well as $\ln(Y_t)$ are consequently best summarized as I(1)

[1] http://www.bundesbank.de/statistik/statistik_realtime.en.php

Figure 2: Level of the quarterly German real GDP



series. When we look at a plot of the data, the question that we can raise however concerns the power of the unit root test versus the trend stationary alternative. Indeed it is well known that for such small number of observations, ADF tests lack power against trend stationary alternative hypotheses. The first issue would then be to look at panel unit root tests to empirically investigate whether they are helpful in this case.

Next we look at a second example from the same data base: the industrial production index . We have monthly data and monthly vintages for that variable. In order to avoid the impact of rebasing we had for real GDP, we have only worked with vintages from 2003M12 to 2008m12 that are on the same base year 2000=100. We denote this series IIP_t. Figures 3 and 4 show the series in levels and in first differences.

We have 61 observations for IIP_t. Recent events heavily impact the end of the series and these few points influence the results of unit root tests. Indeed, one does not reject the null of a unit root neither for the levels ($p - value$: 0.52 for ADF with intercept and trend) nor for the first differences where the $p - value$ of the ADF(2) in a model with a constant is 0.94. This implies that the level of the production index is I(2). Whether panel unit root tests can improve our understanding here is the

Figure 3: Last vertical vintage for the IIP

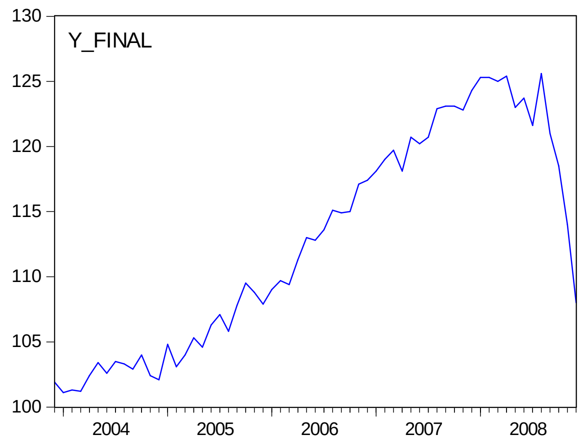
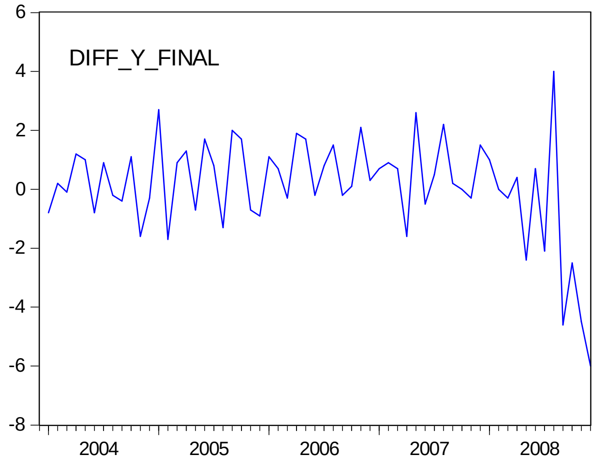


Figure 4: Last vertical vintage for the IIP in first differences



second issue we want to look at.

In summary, these two cases illustrate two situations we would like to investigate using panel unit root tests, namely (i) the power issue between trend stationary/difference stationary processes and (ii) the I(2) vs. I(1).

3 Panel unit root tests

Given the fact that a N vintages are available for the same variable, it might seem appropriate to apply panel unit root tests to the data to investigate whether it is non-stationary or not. Standard statistical software packages as for example EViews include early panel unit root tests, such as those proposed by Levin *et al.* (2002) or Im *et al.* (2003). These tests assume cross-sectional independence, but allow for heterogeneity in the form of individual deterministic effects (constant and/or linear time trend) and heterogenous serial correlation structure of the error terms. Both methods test the same null hypothesis of non-stationarity, but differ in terms of the considered alternative and hence, in the way information is pooled. Levin *et at.* (2002) study balanced panels with N cross-sectional units and T time series observations. They assume a homogenous first order autoregressive parameter and their test is based on the pooled t-statistic of the estimator. Im *et al.* (2003) allow unbalanced panels with N cross sectional units and T_i time series observations for each $i = 1, \ldots, N$. They propose a standardized average of individual ADF statistics to test the pooled unit root null hypothesis against a heterogenous alternative. Both methods assume cross-sectional independence among panel units except for a common time effect. In that case, the derived results remain valid if cross-sectional averages are subtracted from the data.

Table 2 shows the values of the LLC and IPS test statistics as well as the associated $p - values$ for German quarterly GDP. We consider 17 vintages with 40 observations each for the VBS and a minimum of 23 observations in the last OBS series. We can see that if we include the time trend, the unit root null hypothesis is not rejected. The null is rejected in the regression with an intercept only (which does not make sense given the figures) for the VBS framework.

Table 2: First generation unit root tests: Quarterly GDP

		levels		Logs	
		VBS	OBS	VBS	OBS
		$T = 646$	$T = 510$	$T = 646$	$T = 510$
LLC	α	-1.81 (0.03)	-0.41 (0.33)	-2.48 (0.006)	-0.66 (0.25)
IPS	α	3.71 (0.99)	4.95 (1.00)	3.09 (0.99)	4.70 (1)
LLC	α, t	1.56 (0.94)	1.89 (0.97)	1.34 (0.91)	1.87 (0.96)
IPS	α, t	0.71 (0.76)	0.94 (0.82)	0.33 (0.63)	0.83 (0.79)

Table 3: First generation unit root tests: IIP

		levels		Logs	
		VBS	OBS	VBS	OBS
		$T = 1169$	$T = 977$	$T = 1169$	$T = 977$
LLC	α	9.36 (1)	7.64 (1.00)	7.40 (1)	5.98 (1)
IPS	α	13.27 (1)	11.99 (1.00)	11.69 (1)	10.66 (1)
LLC	α, t	2.70 (0.96)	1.29 (0.9)	4.62 (1)	1.48 (0.93)
IPS	α, t	2.76 (0.92)	2.30 (0.98)	2.59 (0.99)	2.11 (0.98)

For the industrial production index we consider 20 series. Table 3 reports the results of LLC and IPS panel unit root tests. We see that we never reject the null of a unit root in the levels or the log-levels. When we take the first differences however we reject the null and the series is seen as I(1), contrarily to the individual unit root tests. For instance the value of the LLC is -27.33 for ΔIIP_t in the OBS vintages and -29.23 for the VBS diagonals.

The assumption of cross-sectional independence, on which the asymptotic results of both the LLC and the IPS procedures rely, has received some attention in the literature. In simulation studies, Banerjee *et al.* (2004, 2005) assess the finite sample performance of panel unit root and cointegration tests when panel members are cross-correlated or even cross-sectionally cointegrated.Their finding is, that all methods experience size distortions when panel members are cointegrated. This means that procedures such as the LLC or IPS test would over-reject the non-stationarity null when there are

Figure 5: IIP: 20 VBS diagonal series

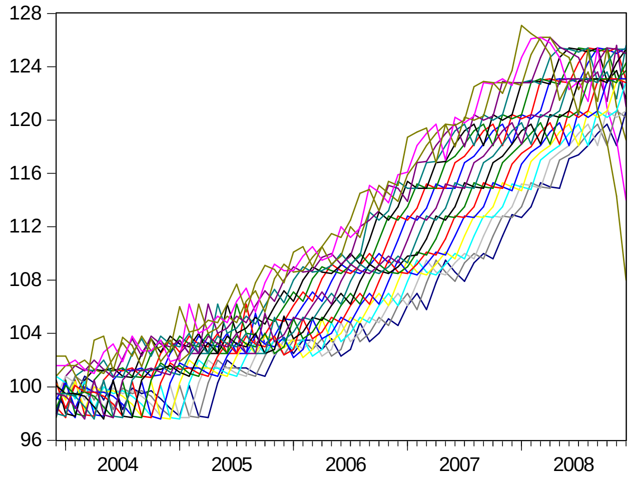
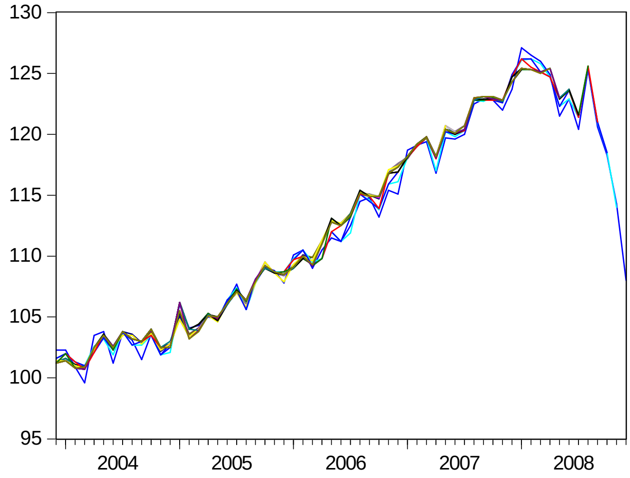


Figure 6: IIP: 20 OBS diagonal series



common sources of non-stationarity. This is analytically confirmed by Lyhagen (2000).

However, due to the nature of the data we should assume the different vintages to not only be cross-sectionally dependent but cross-sectionally cointegrated. Several panel unit root tests have been proposed recently that allow for persistent cross-sectional dependence in the form on non-stationary common factors. Gengenbach, Palm and Urbain (2010) provide a comparison of several tests and assess their small sample properties via Monte Carlo simulations. They find that the method proposed by Bai and Ng (2004) is best equipped to handle the case of cross-member cointegration, while tests based on defactored data tend to reject the null hypothesis of a unit root in this case.

Bai and Ng (2004) propose a procedure to test for a unit root in the common factor and the idiosyncratic component separately. They consider the following model,

$$y_{it} = d_{it} + \sum_{l=1}^{k} \lambda_{il} f_{lt} + v_{it}, \quad (1)$$

where y_{it} is the observed data, d_{it} is a deterministic component, f_{lt} $l = 1, \ldots, k$ are the common factors with corresponding factor loadings λ_{il} and v_{it} is the idiosyncratic component. The deterministic component d_{it} contains either a constant α_i or a linear trend $\alpha_i + \beta_i t$. The vector of common factors $f_t = (f_{1t}, \ldots, f_{kt})'$ is assumed to follow a process

$$\Delta f_t = \Phi(L)\eta_t,$$

where $\eta_t \sim i.i.d.(0, \Sigma_f)$ with finite fourth-order moment and $\Phi(L) = \sum_{s=0}^{\infty} \Phi_s L^s$ is a possibly infinite lag polynomial with rank$(\Phi(1)) = k_1$. So, f_t contains $k_1 \leq k$ stochastic trend and $k - k_1$ stationary components. The idiosyncratic components are modelled as an AR(1) process

$$v_{it} = \delta_i v_{it-1} + e_{it},$$

where e_{it} follows a mean zero, stationary and invertible MA process, such that $e_{it} = \Gamma_i(L)\varepsilon_{it}$ with $\varepsilon_{it} \sim i.i.d.(0, \sigma^2_{\varepsilon_i})$ with finite eighth-order moment. Bai and Ng (2004) initially allow for some weak dependence between the shocks driving the idiosyncratic components, but impose cross-sectional independence to validate pooled testing. Redundant factors, i.e. factors that asymptotically only influence a finite number of panel members, are excluded by assuming $N^{-1}\sum_{i=1}^{N}\lambda_i\lambda_i' \longrightarrow \Sigma_\lambda > 0$, where $\lambda_i = (\lambda_{i1},\ldots,\lambda_{ik})'$.

In this setup, the goal of PANIC is to determine the number of non-stationary common factors k_1 and to test whether $\delta_i = 1$ for $i = 1,\ldots,N$. Bai and Ng (2004) suggest using principal components to obtain consistent estimates of the unobserved components f_t and v_{it}. However, since y_{it} is I(1) principal components cannot be applied to the level but have to be applied to either first differences Δy_{it}, if d_{it} only contains a constant, or demeaned first differences $\Delta y_{it} - \Delta\bar{y}_i$, where $\Delta\bar{y}_i = (T-1)^{-1}\sum_{t=2}^{T}\Delta y_{it}$, if the data contains a linear trend. This method yields estimates $\Delta\hat{f}_t$ and $\Delta\hat{v}_{it}$ which are then accumulated to remove the effect of possible over-differencing. Estimates of the common factors and idiosyncratic components are now

$$\hat{f}_t = \sum_{s=2}^{t}\Delta\hat{f}_s, \tag{2}$$

and

$$\hat{v}_{it} = \sum_{s=2}^{t}\Delta\hat{v}_{is}. \tag{3}$$

For the estimated idiosyncratic component, Bai and Ng (2004) suggest to use an ADF statistic to test for a unit root in an individual series. Denote the t-statistic used to test the unit root null hypothesis for a given $\hat{v}_{it}$ as $ADF_{\hat{v}_i}$. $ADF_{\hat{v}_i}$ has a nonstandard distribution depending on the deterministic components specified. Critical values and a look-up table for p-values are obtained via stochastic simulations and provided by Bai and Ng. To test the pooled hypothesis that all idiosyncratic series are non-stationary, Bai and Ng (2004) propose a Fisher-type test given by

$$P_{\hat{v}_i} = \frac{-2\sum_{i=1}^{N}\log\pi_i - 2N}{\sqrt{4N}}, \tag{4}$$

where π_i is the p-value for the i-th cross-section. Provided the idiosyncratic shocks are cross-sectionally independent, $P_{\hat{v}_i}$ has a standard normal limiting distribution.

To test for unit roots in the extracted common factors, Bai and Ng (2004) suggest an ADF test if only a single common factor has been estimated. The t-statistic used to test the unit root null hypothesis, denoted as $ADF_{\hat{f}}$, has a limiting distribution which corresponds to the usual Dickey-Fuller distribution, depending on the deterministic components specified. If more than one common factor has been estimated, Bai and Ng (2004) suggest an iterative procedure to determine the number of common stochastic trends k_1. Starting with $m = k$, Bai and Ng (2004) suggest two modified Q statistics, MQ_c and MQ_f, to test the null hypothesis that $\hat{f}_t$ contains m unit roots. If the null hypothesis $k_1 = m$ is rejected, the procedure is repeated with $m = m - 1$. Otherwise, $\hat{k}_1 = m$. Bai and Ng (2004) derive the limiting distributions for MQ_c and MQ_f which are again non-standard and provide 1%, 5% and 10% critical values for various values of m[2].

When applying the PANIC method in practice, once has to determine the number of common factors k. Bai and Ng (2002) propose several consistent criteria to select k. However, in small sample the criteria may be unreliable, often selecting the minimum or maximum number of possible factors. Therefore, we perform the tests for k from 1 to 4 factors to investigate whether the results are robust to a misspecification of k, as well as selecting k using the different information criteria allowing for a maximum or 4 factors[3]. We apply the PANIC method to the several panels available. Denote the panels obtained using for 17 vintages of quarterly German GDP as OBS_{gdp} and VBS_{gdp}, respectively. OBS_{gdp} has $N = 17$ and $T = 23$ observations while VBS_{gdp} has $N = 17$ and $T = 39$. The OBS_{iip} and VBS_{iip} panels are obtained using 20 vintages of monthly industrial production index. For the OBS_{iip} panel we have $N = 20$ and $T = 41$ observations, while the VBS_{iip} panel has $N = 20$ and $T = 61$ observations. We consider models with and without a linear deterministic trend.

For OBS_{gdp} all criteria except the IC_2 select $\hat{k}$ = number of common factors. IC_2 select one common factor if we include a deterministic linear trend in the model and two factors if we omit the trend. Including a linear trend, we can reject the unit root null hypothesis for the idiosyncratic data component if we include more than 2 common factors in the model. When we omit the deterministic

[2]For details of the procedure we refer the interested reader to Bai and Ng (2004) p.1133f.

[3]We consider the PC_i and IC_i, $i = 1, 2, 3$, as well as the BIC_3 criteria of Bai and Ng (2002). The later one is not asymptotically consistent but performs well in small samples.

trend, we can reject the unit root null for all k. MQ_c and MQ_f usually select the maximum number of non-stationary factors. The only exception is MQ_c selecting only 3 non-stationary factors in the model with linear deterministic trend and $k = 4$.

For the VBS_{gdp} panel, the PC criteria select the maximum number of possible factors, $\hat{k} = 4$, while the IC criteria select the minimum number, $\hat{k} = 1$. In the model without deterministic trend it is possible to reject the unit root null hypothesis for the idiosyncratic component, but we cannot reject the null if we include a deterministic trend. For a single common factor, the unit root null is rejected for the deterministic trend case but not without a trend. The MQ_f statistic selects the maximum number of possible non-stationary common factors. MQ_c only selects 3 non-stationary factors if $\hat{k} = 4$ and 2 non-stationary factors if $\hat{k} = 3$ for the deterministic trend model.

For the OBS_{iip} panel all considered information criteria select the the maximum number of common factors, $\hat{k} = 4$. We can reject the unit root null hypothesis for the idiosyncratic data component for all k with or without deterministic trend, indicating a stationary revision process. If a linear trend is included, the MQ_c criterion indicates that the common factors are stationary. If a deterministic trend is omitted from the model, MQ_c indicates 2 non-stationary common factors at $\hat{k} = 4$ and 1 common stochastic trend otherwise. MQ_f tends to select the maximum possible number of non-stationary factors.

Using the VBS_{iip} panel, the PC criteria of Bai and Ng (2002) also select the maximum number of common factors. The IC criteria select $\hat{k} = 1$ common factors if no deterministic trend is included and 2, 1 and 4 common factors in the model with trend for IC_1, IC_2 and IC_3, respectively. For the VBS_{iip} panel, we cannot reject the unit root null hypothesis for the idiosyncratic data component. The MQ_c criteria can only reject non-stationarity in the common factors in the model with deterministic trend and $\hat{k} = 2$. Otherwise, it selects 1 or 2 common stochastic trends. The MQ_f statistic once again tends to select the maximum possible number of non-stationary factors.

Finally we apply the bootstrapped panel unit root tests proposed by Palm, Smeekes and Urbain (2008). They consider pooled Levin *et* al. (2002) type tests and group mean Im *et* al. (2003) type tests, respectively based on the OLS or group mean estimate of ρ in the following regression:

$$\Delta y_{it} = \rho y_{it-1} + u_{it}.$$

In particular, the pooled LLC_{bs} statistic is defined as

$$LLC_{bs} = T(\hat{\rho}_{pols}). \tag{5}$$

The group mean IPS_{bs} statistic is given by the following equation,

$$IPS_{bs} = N^{-1} \sum_{i=1}^{N} T(\hat{\rho}_i), \tag{6}$$

where

$$\rho_i = \left(\sum_{t=2}^{T} y_{it-1}^2 \right)^{-1} \left(\sum_{t=2}^{T} y_{it-1} y_{it} \right).$$

Additionally, Palm et al. (2008) also consider a test statistic which is given by T times the median of $(\hat{\rho}_i)$,as the median might be more robust to outliers. Palm et al. (2008) propose a block bootstrap and show that it is asymptotically valid for a number of cross-sectional correlation models, including cross-sectional cointegration.

Using the LLC and IPS bootstrap tests we cannot reject the null hypothesis that the data is I(1) for both GDP and IIP. This is the case for all considered specifications, using OBS and VBS panels, logarithms or level data, and models with and without deterministic trend. Bootstrap p-values range from 0.149 to 1.

4 A Monte Carlo Simulation

In order to simplify the presentation, let us start with the modeling of the last three diagonals, i.e. with $N = 3$ diagonals in a VBS framework. Let us further assume that the variables are $I(1)$ and that there are cointegrated with $r = N - 1$ cointegrating vectors with linear combination, namely cointegrating vectors $(1 - 1)$.

The VECM representation for $p = 1$ can be written as follows

$$\begin{pmatrix} \Delta y_{t-1}^{t} \\ \Delta y_{t-2}^{t} \\ \Delta y_{t-3}^{t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-2}^{t-1} \\ y_{t-3}^{t-1} \\ y_{t-4}^{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

We have estimated that model on several systems of real-time data and it emerges that we cannot jointly reject the following set of null hypothesis

$$H_0 : \alpha_{11} = \alpha_{12} = 0, \alpha_{22} = \alpha_{31} = 0, \alpha_{21} = \alpha_{32} = 1, c_2 = c_3 = 0.$$

Imposing the previous set of restrictions leads to

$$\begin{pmatrix} \Delta y_{t-1}^{t} \\ \Delta y_{t-2}^{t} \\ \Delta y_{t-3}^{t} \end{pmatrix} = \begin{pmatrix} c_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y_{t-2}^{t-1} - y_{t-3}^{t-1} \\ y_{t-3}^{t-1} - y_{t-4}^{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

or in the levels

$$\begin{cases} y_{t-1}^{t} = c_1 + y_{t-2}^{t-1} + \varepsilon_{1t} \\ y_{t-2}^{t} = y_{t-2}^{t-1} + \varepsilon_{2t} \\ y_{t-3}^{t} = y_{t-3}^{t-1} + \varepsilon_{3t} \end{cases}$$

namely a system with a set of $N - 1$ unbiased revisions and the first release that follows a random walk. The variance-covariance Σ_ε matrix of the disturbances has a very particular shape. First the variance of ε_{1t} is very high, then variances of successive diagonals decrease. Next the correlation of older diagonal vintages is quite high while that covariance is very small for new releases. For instance here is the covariance matrix for 3 series

$$\Sigma_\varepsilon = \begin{bmatrix} 4.11 & 0.74 & 0.67 \\ & 2.71 & 1.10 \\ & & 1 \end{bmatrix}.$$

For the intercept we choose $c_1 \tilde{} U[-3,3]$.

We consider $M = 1000$ replications. To save space we only report the outcome for $T = 15$ and $T = 50$ time observations. We added 50 observations to initialize the processes. For the number of vintages we consider $N = \{2, 5, 10, 20\}$. We compute panel unit root tests on both OBS and VBS systems. We can of course only report OBS results for $T > N$. We use for the lags in the estimated models either $k = 0$ and $k = 2$.

First consider result for the first generation panel unit root tests reported in Table 4. None of the procedure gives decent results for our DGP. Size is above the nominal 5% for almost all cases. For the IPS statistics, size seems to increase with the cross-section dimension N, leading to sever size distortions except if $N = 2$. The size of the LLC statistics improves in N, however the tests are still over-sized for all considered panel dimensions. For $T = 50$ since is close to 1, except if $N = 20$.

Next we explore the behavior of the Bai and Ng (2004) tests statistics, presented in Table 5. For the unit root test on the common factor, $ADF_{\hat{f}}$, we find rejection frequencies which are in general above the 5% nominal size. Furthermore, these size distortions are increasing in both N and T. The test on the idiosyncratic component, $P_{\hat{v}_i}$, which should give a high rejection frequency, also show disappointing results. The power of this statistic ranges from 13.4% to 98.8%.

Finally we consider the bootstrap tests statistics proposed by Palm *et al.* (2008). Table 6 shows that these tests are undersized for small N. However, rejection frequencies increase in N leading to some severe size distortions in larger panels. This finding is in line with results reported by Palm *et al.* (2008) for the cross-sectional cointegration case. However, size distortions are more severe using our

Table 4: First generation unit root tests: Monte Carlo simulation

		T = 15				**T = 50**			
		$k=0$		$k=2$		$k=0$		$k=2$	
		VBS	OBS	VBS	OBS	VBS	OBS	VBS	OBS
N = 2	IPS_c	3.7	3.7	5.2	5.9	2.5	2.3	3.1	3.4
	$IPS_{c,t}$	11.4	11.2	10.8	10.8	12.6	12.4	11.0	11.8
	LL_c	92.8	88.1	43.0	40.8	99.7	99.9	95.3	96.7
	$LL_{c,t}$	87.3	81.6	34.9	32.4	99.7	99.9	91.7	94.4
N = 5	IPS_c	15.5	18.7	7.1	12.0	11.4	12.5	5.8	7.7
	$IPS_{c,t}$	43.9	35.7	13.3	17.2	68.5	65.2	26.6	26.0
	LL_c	84.4	64.3	54.8	33.4	99.0	99.8	98.7	99.3
	$LL_{c,t}$	78.7	53.6	40.0	24.3	99.0	99.7	98.5	98.1
N = 10	IPS_c	48.9	28.9	14.1	38.2	34.5	38.8	15.7	18.0
	$IPS_{c,t}$	74.3	12.1	19.1	38.1	94.1	90.2	41.6	39.8
	LL_c	76.0	19.3	48.9	0	99.6	99.4	98.3	98.1
	$LL_{c,t}$	66.7	8.3	26.4	0.1	99.6	99.2	96.2	94.9
N = 20	IPS_c	93.0	-	27.2	-	84.8	83.1	36.9	42.5
	$IPS_{c,t}$	95.4	-	26.3	-	99.8	97.6	66.1	56.7
	LL_c	63.2	-	44.5	-	99.9	93.2	95.6	86.9
	$LL_{c,t}$	49.9	-	21.4	-	99.9	91.7	90.0	76.7

Table 5: Bai and Ng second generation unit root tests: Monte Carlo simulation

		T = 15				**T = 50**			
		$k=0$		$k=2$		$k=0$		$k=2$	
		VBS	OBS	VBS	OBS	VBS	OBS	VBS	OBS
N = 2	$ADF_{\hat{f}}$	5.8	5.5	12.8	13.2	2.8	1.7	3.4	2.4
	$P_{\hat{v}_i}$	51.6	49.9	39.0	39.0	60.2	60.5	28.4	30.2
N = 5	$ADF_{\hat{f}}$	45.5	16.5	35.1	28.7	42.0	6.0	15.5	4.8
	$P_{\hat{v}_i}$	26.9	45.7	26.4	48.7	27.4	45.9	17.8	28.2
N = 10	$ADF_{\hat{f}}$	85.9	40.2	56.7	77.5	89.8	20.0	34.4	12.4
	$P_{\hat{v}_i}$	34.3	48.8	33.5	81.7	20.9	80.5	13.4	56.6
N = 20	$ADF_{\hat{f}}$	-	-	-	-	100	57.6	59.3	37.7
	$P_{\hat{v}_i}$	-	-	-	-	61.7	98.8	35.0	87.3

Table 6: Smeekes, Palm and Urbain boostraped unit root tests: Monte Carlo simulation

		$T = 15; k = 0$		$T = 50; k = 0$	
		VBS	OBS	VBS	OBS
N = 2	LLC_{bs}	0.9	0.4	0.9	0.5
	IPS_{bs}	2.0	1.9	0.3	0.2
N = 5	LLC_{bs}	9.5	6.0	5.3	2.2
	IPS_{bs}	12.3	10.2	2.7	1.3
N = 10	LLC_{bs}	32.6	6.3	18.7	10.7
	IPS_{bs}	40.0	6.5	12.7	8.1
N = 20	LLC_{bs}	-	-	66.3	51.3
	IPS_{bs}	-	-	48.1	40.5

DGP. Size can be improved by increasing the block length for the bootstrap, which raises the question of optimal block length selection in applications.

5 Conclusion

It is intuitively an interesting idea to look at real-time data with the aim to improve the size and the power of unit root tests by using several vintages in a panel, as e.g. suggested by Patterson and Heravi (2004a). However there are many new issues that this type of observation imply. Even assuming that the data a free of typos and have been correctly rebased, the large number of cross-member cointegrating relationships makes the first generation of panel unit root tests invalid.

Because of the potential existence of $N-1$ cross-member cointegrating relationships we are left with a unique common stochastic trend of dimension T. While the panel unit root test of Bai and Ng (2004) is theoretically equipped to handle a situation where the non-stationarity is due to a single common factor, the power of the unit root test on the common factor in that situation does not improve over univariate tests on a single vintage. Additionally, we obtain a panel unit root test for the idiosyncratic data component, which we view as a proxy for the measurement error. This statistic enables us to gain insight in the consistency of the revision process. However, in simulations the power of this statistics is quite sensitive to the actual specifications of the sample. It seems that working with OBS diagonals, namely for the same calendar period, gives more plausible results even though the T dimension is reduced in balanced panels.

The bootstrap panel unit root tests of Palm *et al.* (2008) are also theoretically able to handle cross-member cointegration. However, in the extreme case of a single common stochastic trend driving the non-stationarity of the data they suffer from size distortions and reduced power. Our simulation results are complementary to the ones optained by Palm *et al.* (2008) in that regard.

While it seems intuitive to use panel techniques to improve unit root tests for real-time data, it appears that existing methods are not yet equipped to handle the strong dependence between individual vintages, induced by the peculiar DGP underlying such data sets.

References

[1] Bai, J. and S. Ng (2004), A PANIC Attack on Unit Roots and Cointegration, *Econometrica* 72, 1127-1177.

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

[3] Banerjee, A., M. Marcellino and C. Osbat (2004), Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data, *Econometrics Journal* 7, 322-340.

[4] Banerjee, A., M. Marcellino and C. Osbat (2005), Testing for PPP: Should We Use Panel Methods?, *Empirical Economics* 30, 77-91.

[5] Croushore, D. and T. Stark (2001), A real-time data set for macroeconomists, *Journal of Econometrics*, 105, 111-130.

[6] Gengenbach, C., F.C. Palm and J.-P. Urbain (2010), Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling, *Econometric Reviews*, 29, 111-145.

[7] Hecq, A. and J.P.A.M. Jacobs (2009), On the VAR-VECM Representation of Real Time Data, mimeo, Maastricht University.

[8] Im, K.S., M.H. Pesaran and Y. Shin (2003), Testing for unit roots in heterogenous panels, *Journal of Econometrics* 115, 53-74.

[9] Jacobs, J.P.A.M. and S. van Norden (2007), Modeling Data Revisions: Measurement Error and Dynamics of "True" Values, Technical Report CREF 07-09, HEC, Montréal.

[10] Kasa, K. (1992), Common Stochastic Trends in International Stock Markets, *Journal of Monetary Economics*, 29, 95-124.

[11] Knetsch, T. and H.-E. Reimers (2006), How to treat benchmark revisions? The case of German production and order statistics, Deutsche Bundesbank, DP 38/2006.

[12] Levin, A., C.F. Lin and C.S.J. Chu (2002), Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of Econometrics* 108, 1-24.

[13] Lyhagen, J. (2000), Why not use standard panel unit root test for testing PPP, mimeo, Stockholm School of Economics.

[14] Mills, T.C. (1998), Recent Developments in Modeling Nonstationary Vector Autoregressions, *Journal of Economic Surveys*, 12, 279-312.

[15] Patterson, K.D. (2000), Which Vintage of Data to use When There are Multiple Vintages of Data? Cointegration, Weak Exogeneity and Common factors, *Economics Letters*, 69, 115-121.

[16] Patterson, K.D. (2002a), Modelling the Data Measurement Process for the Index of Production, *Journal of the Royal Statistical Society, Series A*, 165, 279-296.

[17] Patterson, K.D. (2002b), The Data Measurement Process for UK GNP: Stochastic Trends, Long Memory, and Unit Roots, *Journal of Forecasting*, 21, 245-264.

[18] Patterson, K.D. (2003), Exploiting Information in Vintages of Time-Series Ddata, *International Journal of Forecasting*, 19, 177-197.

[19] Patterson, K.D. and S.M. Heravi (1991a), Are Different Vintages of Data on the Components of GDP Co-Integrated?, *Economics Letters*, 35, 409-413.

[20] Patterson, K.D. and S.M. Heravi (1991b), Data Revisions and the Expenditure Components of GDP, *Economic Journal*, 101, 887-901.

[21] Patterson, K.D. and S.M. Heravi (2004a), Revisions to Official Data on U.S. GDP: A Multivariate Assessment of Different Vintages, *Journal of Official Statistics*, 20, 573-602.

[22] Patterson, K.D. and S.M. Heravi (2004b), Rejoinder: Rebasing, Common Cycles, and Some Practical Implications of Modelling Data Revisions, *Journal of Official Statistics*, 20, 631-644.