Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: http://www.tandfonline.com/loi/raec20

Can you do the wrong thing and still be right? Hypothesis testing in I(2) and near-I(2) cointegrated VARs

Francesca Di Iorio, Stefano Fachin & Riccardo Lucchetti

To cite this article: Francesca Di Iorio, Stefano Fachin & Riccardo Lucchetti (2016): Can you do the wrong thing and still be right? Hypothesis testing in I(2) and near-I(2) cointegrated VARs, Applied Economics

To link to this article: http://dx.doi.org/10.1080/00036846.2016.1142660

Published online: 10 Feb 2016.

APPLIED ECONOMICS, 2016
http://dx.doi.org/10.1080/00036846.2016.1142660

Routledge
Taylor & Francis Group

Can you do the wrong thing and still be right? Hypothesis testing in I(2) and near-I(2) cointegrated VARs

Francesca Di Iorio[a], Stefano Fachin[b] and Riccardo Lucchetti[c]

[a]Università di Napoli Federico II, Naples, Italy; [b]'Sapienza' Università di Roma, Roma, Italy; [c]Università Politecnica delle Marche, Ancona, Italy

ABSTRACT
In this paper, we investigate the small-sample performance of LR tests on long-run coefficients in the $I(2)$ model; we focus on a comparison between $I(2)$ and near-$I(2)$ data, i.e. $I(1)$ data with a second root very close to unity, and report the results of some Monte Carlo experiments. With near-$I(2)$ data, the finite-sample properties of the tests are (i) similar to those found with genuine $I(2)$ data, (ii) systematically superior to those of the analogous tests constructed in the $I(1)$ model, even if the latter is, in principle, correctly specified and the former is not. Therefore, there seems to be strong support to the idea that, in practice, modelling near-$I(2)$ data using the $I(2)$ model may be a good idea, despite the inherent misspecification.

KEYWORDS
Cointegration; $I(2)$; near-$I(2)$; hypothesis testing

JEL CLASSIFICATION
C12; C13

I. Introduction

Since its introduction, the cointegrated $I(1)$ VAR has proven to be an invaluable tool to investigate the long-run dynamics of economic systems in countless applications. The driving factor behind its success is that $I(1)$ dynamics is often consistent both with economic theory and with observed evidence for macroeconomic variables. However, a view sometimes held among macro practitioners is that incorporating the restrictions from cointegration into VARs used for empirical analysis is a rather futile exercise. In fact, a statement often heard at seminars and conferences is that if all you need to do is analyse the dynamic properties of the economy via tools such as impulse response analysis, then worrying about cointegration is a waste of time, since 'unit roots will take care of themselves'.

This attitude is arguably less justified when it comes to testing hypotheses which can be given economic interpretation, especially for the long-run properties of the system. In this case, one would surmise that specifying correctly the long-run properties of the estimated system is a matter of utmost importance that deserves special care.

The issue becomes even more complex when the observed persistence of the data series is such that higher orders of integration should be considered: for instance, models investigating purchasing power parity (PPP) for exchange rates have largely failed to reach clear results (see, e.g., Rogoff 1996). Bacchiocchi and Fanelli (2005) argued that this may be due to inadequate specification: modelling exchange rates may require allowing for $I(2)$, rather than $I(1)$, dynamics. Analogously, Johansen et al. (2010) and Juselius (2013) pointed out that the cointegrating relations estimated in $I(1)$ VARs involving exchanges rates typically show long swings away from equilibrium. Inspecting the estimates reveals that, in those cases, the largest unrestricted roots are very close to unity: in other words, these systems, although strictly speaking $I(1)$, are in practice very close to be $I(2)$. Johansen et al. (2010) label these systems as *near-*$I(2)$.

If correct specification of the order of integration was an absolute must, hypothesis testing on the long-run properties of the system would be considerably complicated when $I(1)$ and $I(2)$ models are nearly observationally equivalent. Since the cointegrated $I(2)$ model is basically an $I(1)$ model with some extra rank restrictions, the safe option of applying as few restrictions as possible would lead to adopting the $I(1)$ model. This, however, would leave unresolved the problem of explicitly modelling long swings as $I(1)$ processes, which may be inadequate in finite samples. In this paper, we check if the

CONTACT Stefano Fachin stefano.fachin@uniroma1.it

reverse option, that is, employing an $I(2)$ statistical framework and the restrictions it entails, even when data are not genuinely $I(2)$ (but nearly so), is a justifiable empirical strategy, given the need for modelling extremely persistent data paths.

It has already been argued that allowing for $I(2)$ or near-$I(2)$ dynamics in empirical modelling may be a very reasonable choice: according to Johansen et al. (2010), $I(2)$ methods should also be used for data generated by near-I (2) data generating processes (DGPs). This would avoid the shortcomings of the I (1) approach discussed earlier, while yielding the additional advantage of exploiting the ability of the I (2) model to allow for short-, medium- and long-run error correction behaviour separately.[1] In other terms, Johansen et al. (2010) suggest that a (theoretically misspecified) $I(2)$ VAR is more suited than a (theoretically well-specified) $I(1)$ VAR for analysing near-I (2) data.

The idea of deliberately applying a misspecified model may appear puzzling at first sight. However, it may be argued that even if one knew the true DGP, that may not be the optimal choice for empirically modelling the data, because the statistical model implied by the DGP could be too demanding to apply in a given empirical situation. To exemplify this concept, consider a situation in which, $E(y_i|x_i) = g(x'_i;\, \theta)$ where the $g(\cdot)$ function exhibits potentially problematic features (discontinuities, non-linearity, and so on). If one had a practical situation where the dataset was small, or where the observed values for x_i were all concentrated in a small interval, a simple linear model may well perform better at some tasks (such as prediction or hypothesis testing) than the true DGP. This would amount to imposing restrictions that are not, strictly speaking, satisfied in the population model but make the empirical model more focused on the features of interest by the practitioner, and therefore more effective in finite samples. As the famous quote goes: 'Essentially, all models are wrong, but some are useful.' (Box and Draper 1987, 424).

The key question is not which of the two options ($I(1)$ versus $I(2)$) is 'less wrong' for near-$I(2)$ datasets, but which one is the more useful in a given empirical setting.[2] Going back to the $I(2)$ model, a first open question is the following: although the statistical foundations of the $I(2)$ model are now over two decades old (Johansen 1992, 1997) and fairly well developed, very little is known on the applicability of asymptotic results to finite samples. A second, even more important point, concerns the possible adverse consequences of the use of a misspecified model, as Elliot (1998) showed that tests on the cointegrating coefficients of near-I (1) variables in $I(1)$ cointegrated VARs can be severely biased. Summing up, there are two empirically relevant open questions on the $I(2)$ model: first, what are its small sample performances? Second, what are its properties when estimated on near-$I(2)$ datasets?

In order to investigate these issues we will concentrate, like Elliot (1998), on tests on long-run coefficients. We deliberately avoid to tackle the issue analytically and concentrate on a simulation-based study. This choice will be fully motivated in the body of the article, but the gist of the argument is that the technical details involved in tackling the problem via an abstract and general approach are extremely complex, and may obfuscate unnecessarily our message to the practitioner: *finite-sample bias may outweigh misspecification issues in real-life situations*, so a practitioner may end up with more reasonable results by employing the 'wrong' models than the 'true' one. Or, as the title goes, you may be right in doing the wrong thing.

After a brief review of the $I(2)$ model (used mainly to establish notation) in Section II, in Section III we illustrate the basic idea that will be developed in the rest of the paper, that is a *thought experiment*, in which we perform a comparison between the same statistic computed on $I(2)$ and near-$I(2)$ data. A Monte Carlo experiment is then presented in Section IV; Section V summarises the results and concludes.

II. I(2) and near-I (2) VARS

The diffusion of $I(2)$ VARs has so far been considerably hampered by the common view of double-unit roots as economically implausible[3]. However, the

[1] See, e.g. Juselius (2006), p. 333. From the empirical point of view this may be important, e.g. in asset markets, where error-increasing medium-run relations may be associated with error-correcting long-run ones.

[2] A similar question has been addressed in the $I(1)$ case by (Johansen 2006a) in his critical appraisal of the standard methods used to estimate empirical DSGE models.

[3] Applications of the $I(2)$ model are rather scant: besides Bacchiocchi and Fanelli (2005), some examples are Kongsted (2003), Nielsen and Bowdler (2006), and more recently the empirical section in Kurita, Nielsen, and Rahbek (2011).

Table 1. ADF tests on real GDP growth rates.

Countries	Available obs.	ADF test	Asymptotic *p*-value (%)
Australia	125	−3.87	0.22
Belgium	73	−4.64	0.01
Canada	125	−5.16	0.00
Chile	41	−3.17	2.15
Czech Republic	69	−2.74	6.72
Denmark	89	−3.66	0.47
Estonia	73	−2.52	11.03
Finland	93	−3.24	1.75
France	125	−4.15	0.08
Germany	89	−5.48	0.00
Hungary	73	−1.89	33.72
Iceland	65	−2.21	20.13
Ireland	65	−1.67	44.89
Italy	89	−3.35	1.27
Japan	77	−5.59	0.00
Korea	125	−2.82	5.57
Luxembourg	73	−3.03	3.18
Mexico	81	−4.78	0.01
Netherlands	101	−1.95	30.81
New Zealand	101	−2.52	11.10
Norway	125	−2.57	9.93
Poland	73	−3.89	0.21
Portugal	73	−1.40	58.61
Slovak Republic	65	−2.92	4.28
Slovenia	69	−1.94	31.31
Spain	53	−0.93	77.99
Sweden	81	−3.18	2.09
Switzerland	125	−5.43	0.00
Turkey	61	−4.49	0.01
United Kingdom	65	−2.94	4.08
United States	125	−2.71	7.19
Euro area	73	−3.52	0.75
European Union	73	−3.39	1.14
G7	125	−2.48	12.02
NAFTA	125	−2.79	5.90

potential usefulness of the $I(2)$ model has probably been somehow downplayed.

Take, for example, real GDP growth: if one looks at actual historical data, then that the standard assumption of stationarity should not be taken for granted. Table 1 reports the details of ADF tests on quarterly GDP growth rates of 35 countries and international aggregates, as supplied by the online OECD database, for the sample 1982:1–2013:2. Of course, this should not be taken as a serious modelling exercise; these results may be spurious for a multitude of reasons (structural breaks, misspecification of the ADF equation and low power are just examples), and therefore should not be taken at face value. However, even with all these caveats in mind, the number of cases in which the null hypothesis of a unit root in GDP growth cannot be rejected at the customary significance levels is striking. In almost half of the cases (15 out of 35) the p-values are greater than 5%, and in nearly one-third of the panel (10 cases) even greater than 10%.

Even under the interpretation of Table 1 as nothing more than a collection of descriptive statistics, we believe that its message is quite clear: in the sample sizes typically available to practitioners it may not be easy to reconcile the standard view of GDP growth as stationary with observed facts.[4]

Moreover, the case for $I(2)$ models has been forcefully put forward, from a theoretical point of view, by including agents' forecasting revision rules based on Frydman and Goldberg (2007) 'Imperfect Knowledge Economics' (IKE): Frydman et al. (2012) showed that an IKE-based model of exchange rates will generate $I(2)$ or near-$I(2)$ dynamics.

The distinction between $I(2)$ and near-$I(2)$ DGPs can be succinctly illustrated as follows: consider a p-variate VAR(2) in VECM form

$$\Delta^2\mathbf{X}_t = \Pi_1\mathbf{X}_{t-1} + \Gamma\Delta\mathbf{X}_{t-1} + \varepsilon_t \qquad (1)$$

where $\varepsilon_t = (\varepsilon_{1t},\ldots, \varepsilon_{pt})' \sim IID(0, \Sigma)$, and $t = 1,\ldots, T$; the matrix Π_1 is assumed to satisfy the reduced rank restriction $\Pi_1 = \alpha\beta'$, where α and β are $p \times r$ matrices, with $p > r$. Now define a second matrix

$$\Pi_2 = \alpha'_\perp \Gamma \beta_\perp \qquad (2)$$

If this $p - r$ square matrix has reduced rank $s < p - r$, then the order of integration of $\mathbf{X}_t$ is (at least) 2;[5] in this case, a parallel factorisation $\Pi_2 = \varphi\eta'$ holds.[6]

Much of our analysis will involve *near-$I(2)$* processes, that is, processes that strictly speaking are not $I(2)$, but exhibit features such that in finite samples they are almost impossible to distinguish from true $I(2)$ systems. More formally, following Johansen et al. (2010), a near-$I(2)$ cointegrated VAR is defined as a system satisfying the following conditions: (i) Π_1 has reduced rank r; (ii) Π_2 is invertible; (iii) the largest unrestricted roots are very close to unity.[7]

[4]Note also that univariate unit root tests may have a low ability to detect a double unit root when the shocks to the drift term of the differenced process (which generate the second unit root) are actually small compared to those to the differenced process (see Juselius 2013). Hence, these tests are likely to be somehow biased *against* the $I(2)$ hypothesis.

[5]To exclude orders of integration higher than two, an additional rank condition has to hold; see Johansen (1992).

[6]Clearly, greater generality can easily be achieved by adding deterministic terms and/or lags of $\Delta^2\mathbf{X}_t$ to the right-hand side of Equation (1), but this is totally unnecessary in the present context.

[7]The phrase 'unrestricted roots' is meant as a short form of 'roots of the autoregressive polynomial $I(1 - z)^2 - \Pi_1 z - \Gamma z(1 - z)$ when estimated unrestrictedly'.

In a recent working paper, Mosconi and Paruolo (2013) showed that the following reparametrisation of (1) holds:

$$\Delta^2\mathbf{X}_t = \alpha(\beta'\mathbf{X}_{t-1} + \upsilon'\Delta\mathbf{X}_{t-1}) + (\xi\gamma' + \varsigma\beta')\Delta\mathbf{X}_{t-1} + \varepsilon_t \quad (3)$$

where the matrices υ and ς are $p \times r$, while $\boldsymbol{\xi}$ and γ are $p \times s$. The terms $(\beta'\mathbf{X}_{t-1} + \upsilon'\Delta\mathbf{X}_{t-1})$ and $[\gamma \mid \beta]'\Delta\mathbf{X}_{t-1}$ are both stationary; the former is known as multicointegration relation, or integral control term, the latter as medium-run relation, or proportional control term. Model (3) can be conveniently rewritten as

$$\Delta^2\mathbf{X}_t = \alpha[\beta' : \upsilon']\begin{bmatrix}\mathbf{X}_{t-1}\\ \Delta\mathbf{X}_{t-1}\end{bmatrix} + [\xi : \varsigma]\begin{bmatrix}\gamma'\\ \beta'\end{bmatrix}\Delta\mathbf{X}_{t-1} + \varepsilon_t \quad (4a)$$

$$= [\alpha : \xi : \varsigma]\begin{bmatrix}\beta' & \upsilon'\\ 0 & \gamma'\\ 0 & \beta'\end{bmatrix}\begin{bmatrix}\mathbf{X}_{t-1}\\ \Delta\mathbf{X}_{t-1}\end{bmatrix} + \varepsilon_t \quad (4b)$$

$$= \eta\zeta'\begin{bmatrix}\mathbf{X}_{t-1}\\ \Delta\mathbf{X}_{t-1}\end{bmatrix} + \varepsilon_t$$

The system in this parametrisation can be estimated using a two-step generalised least squares (GLS) switching algorithm first devised in Boswijk (2000), similar to the one proposed by Johansen (1997). Further, hypotheses on linear restrictions on the cointegration parameters are easily expressed as

$$H_0 : \mathrm{vec}(\zeta) = S\Phi_\zeta + s,$$

where s is a normalisation vector and Φ_ζ is the vector of unrestricted coefficients in $\boldsymbol{\zeta}$.

As anticipated in the introduction, in our experiments we will concentrate on likelihood ratio (LR) tests on β, the most important parameter as far as the interest is on the long-run properties of the system. For instance, the so-called nominal-to-real transformation, routinely applied in empirical work when analysing variables in real terms, implicitly imposes a restriction on the β vector (Kongsted 2005). Clearly, from the economic point of view, hypotheses on all elements of the multicointegration relation are of even higher interest. However, Johansen (2006b) showed that while, analogously to the $I(1)$ case, in the $I(2)$ VAR LR tests on β are asymptotically χ^2, the statistics for hypothesis on the entire vector of multicointegration coefficients are not LAMN, so that the asymptotic distribution of the LR test is unknown.

III. The thought experiment

Before proceeding any further, let us state precisely our questions of interest. What happens, in finite samples, when the matrix Π_2 in Equation (2) is invertible, but very badly conditioned (that is, very close to being singular)? Clearly, a rigorous approach to this matter would involve analysing the issue via a local-to-unity approach, in a similar way as the literature on unit root testing has developed over the past 30 years.

This issue, however, is considerably more complicated than in the simple unit-root test case: first, the multivariate nature of the models make it far from obvious to indicate precisely what kind of 'deviation from unity' should be adopted. Moreover, given a question of interest (say, hypothesis testing on a certain sub-vector of parameters), a cointegrated system contains, even in the simplest cases, a high number of ancillary parameters to consider, which could potentially make the issue quite murky. Therefore, we decided to adopt a more experimental approach and consider Monte Carlo evidence in some specific cases so as to steer the research question to a qualitative appraisal of the issue. A fully developed analytical approach will form the object of future research.

Our object of attention here will be the properties of LR tests on the matrix β, which are by far the most important subset of parameters in terms of economic theory, when the order of integration is unknown and may be misspecified. In other terms, our purpose is to assess the properties of a certain testing procedure which assumes $I(2)$-ness, with both genuine $I(2)$ and near-$I(2)$ data.

To check the effects on LR tests of misspecification of the order of integration, these must be disentangled from other sources of finite-sample problems. To achieve this, we would need two datasets as close as possible and such that Π_2 is singular in one case and invertible in the other, with the second eigenvalue of the companion matrix of the system (1) being close to 1. Then, the statistics of interest could be computed for both datasets and compared. Our conjecture is that the tests from the

$I(2)$-VECM's should perform (i) in roughly the same way with genuine $I(2)$ data and with near- $I(2)$ ones; (ii) with near-$I(2)$ data, not worse, and hopefully better, than the tests computed from the $I(1)$-VECM.

Note that, in order for the aforementioned strategy to work, it is not sufficient to have two datasets generated by slightly different DGPs. If it were so, the actual datapoints may (and, as a rule, will) be very different from one another. The requirement we need is much stricter: we need two datasets in which *the difference between each and every datapoint is attributable only to the difference in the order of integration of the respective DGPs*.

Clearly, no real dataset can ever satisfy this requirement. On the contrary, a Monte Carlo experiment allowing such a comparison is readily designed. Assume the data are generated as per the VAR(2) formula (1) using $\mathbf{X}_0$ as starting (pre-sample) values, an artificially generated $T \times p$ matrix of disturbances E_n and a given set of parameter values $\boldsymbol{\theta}_i$ (where $i = 1,2$). More precisely, define $\boldsymbol{\theta}_2$ as a set of parameters under which the DGP is $I(2)$ and $\boldsymbol{\theta}_1$ as a set such that the DGP is $I(1)$, but nearly $I(2)$. Further, let $\mathbf{X}_n^2$ and $\mathbf{X}_n^1$, respectively, be the resulting datasets obtained in the n-th simulation, and τ_2 and τ_1 the LR tests for a given hypothesis on β computed in the $I(2)$ model estimated, respectively, on $\mathbf{X}_n^2$ and $\mathbf{X}_n^1$. Since the other two elements of the DGP (the initial values, $\mathbf{X}_0$, and the shocks, E_n) are the same, any difference in the behaviour of the LR statistics τ_2 and τ_1 can be safely ascribed to model misspecification alone: the former is obtained under $\boldsymbol{\theta}_i = \boldsymbol{\theta}_2$, so that the data are indeed $I(2)$ and the $I(2)$ model is correct. The latter, on the other hand, is computed on data obtained under $\boldsymbol{\theta}_i = \boldsymbol{\theta}_1$, which are $I(1)$: the assumption of two unit roots underlying the $I(2)$ model is wrong.

Therefore, we first consider the $I(2)$-VECM model estimated on $I(2)$ data to evaluate the finite sample performances of the test. Next, the $I(2)$-VECM is estimated again on the corresponding near-$I(2)$ data to assess the consequences of misspecification of the integration order. Finally, these latter results are compared to their obvious benchmark, that is, the performances of the analogous tests computed from the correctly specified $I(1)$ VAR. Note that the problem of assessing the order of integration is totally irrelevant here. In other words, the $I(2)$ model is not assumed to be applied to near-$I(2)$ data as a consequence of, for example, inadequate pretesting. This model is purposely used because it is believed to bring some distinctive advantages *in spite* of misspecification.

Table 2. Acceptance/rejection decisions for the LR test in the I (2) model, H_0 true.

LR test τ_1 on near-$I(2)$ data				
LR test τ_2 on $I(2)$ data		no rejection	rejection	
	no rejection	A	B	$A + B$
	rejection	C	D	$C + D$
		$A + C$	$B + D$	N

For a given sample size and significance level, the performances of the test under the two DGPs over the N Monte Carlo replications can be usefully summarised in Table 2.

Since under H_0 the asymptotic distribution of the LR test on $I(2)$ data, τ_2 is χ^2, then the ratio $(C + D)/N$ must converge asymptotically to the nominal size, although its small sample behaviour is unknown. On the other hand, nothing is known about the LR test on near-$I(2)$ data, τ_1. Is its performance noticeably different from τ_2's? And if so, how?

A couple of remarks are in order here. First of all, recall that we are not comparing (as customary in the literature) two different test statistics on the same dataset. In fact, we are doing precisely the opposite: we are comparing the behaviour of the same test statistic (LR) on datasets generated by two (slightly) different DGPs; the main point is to understand how large the differences are. This is why the rejection rates of the tests, on the marginal row and column of Table 2, are not the most important items to consider. Examining the non-diagonal cells in the table is crucial for the present purpose, as those are the cases when the LR tests in models with correctly specified and misspecified number of unit roots deliver contrasting results. More specifically, B is the number of cases when τ_2 leads to the correct decision and τ_1 does not, whereas C is the frequency of the opposite case of τ_2 leading to the wrong decision and τ_1 to the correct one. The combination B, when τ_2 leads to the correct decision and τ_1 does not, is relatively unsurprising (the test in the misspecified model performs worse than in the right model); the other, counted by C, is quite counterintuitive, as the test in the misspecified model outperforms the 'correct' one. However, in finite samples, this can be the simple consequence of misspecification and sample biases of opposite sign compensating each other.

Hence, to summarise the differences in the outcomes delivered by the tests with $I(2)$ and near-$I(2)$ data, the key statistic to compute is the proportion of simulations in which τ_2 and τ_1 yield opposite conclusions, i.e. $\nabla = (B + C)/N$.

IV. The Monte Carlo experiments

As argued in the previous section, in order to perform the thought experiment, we must simulate several pairs of datasets as similar as possible to one another, with one being truly $I(2)$ and the associated one being near-$I(2)$. In order to achieve this goal, there are various conditions to be controlled. From Section II, a first requirement which must be always satisfied is that the matrix Π_1 as defined in (1) must be of the same reduced rank (say, r). The second requirement is that in the $I(2)$ datasets the rank of the matrix Π_2 must be smaller than $p - r$, where p is the number of variables. Instead, in near-$I(2)$ systems, this matrix must be invertible, and the largest unrestricted root close to unity.

Hence, to generate the near-$I(2)$ dataset closest to a given $I(2)$ one, it is important not only that the rank conditions apply, but also that the same number of $I(1)$ cointegrating relationships holds for the two datasets. Given this rather complex set of conditions, the task of reconciling the contrasting requirements of control and empirical relevance is more demanding than usual in Monte Carlo work.

For this reason, we will use different DGPs, characterized by opposite advantages and disadvantages. The first two DGPs are highly stylized, hence, on the one hand, easily controlled, but, on the other, admittedly of limited relevance for applied work. The first one, inspired by Hendry and von Ungern-Stenberg's (1981) famous specification of the consumption function, has been used in the previous studies on $I(2)$ systems by Johansen (1992) and by Paruolo and Rahbek (1999), and is thus the natural first step of our simulation experiments. The second DGP, taken from Kongsted (2005), mimics a test of the validity of the nominal-to-real transform (henceforth NtR), thus providing an example of a theoretically important question with empirically relevant consequences[8]. Finally, the third DGP is one of the models estimated by Bacchiocchi and Fanelli (2005) in their study of PPP. It is thus by definition of the maximum possible relevance for applied work (the DGP parameters are set at the values estimated empirically), but, on the other hand, more difficult to control.[9] Should the simulations using these different DGPs deliver consistent results, these may provide some insight into the practical issues of empirical work. The set-up and results from each experiment will be presented separately first, and then an overall assessment will be drawn.

The Hendry and von Ungern-Stenberg DGP

Set-up

The Hendry and von Ungern-Stenberg (briefly H-US) DGP mimics a system with three variables, naturally interpreted as logs of consumption (c_t), liquid assets (l_t) and income (y_t). The system is driven by one stochastic trend and includes two error correction mechanisms:

$$\begin{aligned}\Delta c_t &= v\Delta y_{t-1} - a_{11}(c_{t-1} - b_1 y_{t-1}) + \\ &\quad - a_{12}(l_{t-1} - b_2 y_{t-1}) + \varepsilon_{1t}\end{aligned} \tag{5a}$$

$$\Delta l_t = -a_{21}(c_{t-1} - b_1 y_{t-1}) + \varepsilon_{2t} \tag{5b}$$

$$\Delta y_t = \rho \Delta y_{t-1} + \varepsilon_{3t} \tag{5c}$$

Equation (5a) is a consumption function, Equation (5b) describes asset accumulation through saving, and (5c) is a simple law of motion for GDP. In this simple formulation, it is immediately obvious that the order of integration of the system is determined by the parameter ρ: for $\rho = 1$ it is $I(2)$ and for $\rho < 1$ it is $I(1)$, while near-$I(2)$ systems are characterized by ρ close to 1.

Defining $\mathbf{X}_t = (c_t, l_t, y_t)'$, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$, and imposing, for convenience, the homogeneity condition $b_1 = b_2 = 1$ the DGP can be compactly written in levels as

[8]We thank an anonymous referee for drawing our attention to this point

[9]For instance, since any change of the long-run parameters would affect in an unknown way the adjustment coefficients, in this case we will not be able to evaluate the power of the test.

$$\mathbf{X}_t = \begin{bmatrix} 1-a_{11} & -a_{12} & \nu + a_{11} + a_{12} \\ -a_{21} & 1 & a_{21} \\ 0 & 0 & 1+\rho \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} 0 & 0 & -\nu \\ 0 & 0 & 0 \\ 0 & 0 & -\rho \end{bmatrix} \mathbf{X}_{t-2} + \varepsilon_t. \tag{6}$$

and in a manner compatible with Equation (4a) as

$$\Delta^2 \mathbf{X}_t = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} -1 & 0 & \nu \\ 0 & -1 & 0 \\ 0 & 0 & \rho - 1 \end{bmatrix} \Delta \mathbf{X}_{t-1} + \varepsilon_t \tag{7}$$

Consistently with the remarks made previously on the order of integration of the system, the matrix Π_2 (which equals $\rho - 1$) has a reduced rank, and therefore the system is truly $I(2)$ (with $r = 2$ and $s = 1$) only if $\rho = 1$.

In the simulation experiments, the loadings a_{ij} have been chosen in such a way to control the roots of the VAR, so as to generate series with a wide range of diverse dynamic features. Here we report detailed results for the three combinations of loadings, given in Table 3; other combinations give essentially the same results. As explained earlier, using the same loadings and ρ close to 1 yields near-$I(2)$ systems; the simulations reported here were obtained with $\rho = 0.80$ and $\rho = 0.90$. The lagged effect of income growth on consumption growth, ν, is always fixed at 0.5. Disturbance terms are IID Gaussian with covariance matrix diag(1,1,0.1).[10] These choices were made so as to generate simulated series whose relative order of magnitude resembles the ones observed in real-world data. The sample sizes considered are $T = 64,128, 256$. The first two are representative of standard macroeconomic applications (e.g. $T = 128$ corresponds to 32 years of quarterly data), while the last one is useful to check asymptotic behaviour.

Table 3. H-US DGP: VAR dynamic properties in different Monte Carlo settings – $\rho = 1$.

Model	a_{11}	a_{12}	a_{21}			λ	
DGP_1	0.20	0.20	−0.20	1	1	0.92	0.92
DGP_2	0.80	0.20	−0.20	1	1	0.95	0.25
DGP_3	0.80	0.40	−0.40	1	1	0.60	0.60
DGP_4	0.50	0.50	−0.50	1	1	0.87	0.87
DGP_5	0.80	0.80	−0.80	1	1	0.92	0.92
DGP_6	0.40	0.20	−0.20	1	1	0.80	0.80
DGP_7	0.80	0.60	−0.60	1	1	0.75	0.75

λ: non-zero eigenvalues of the companion matrix for (1).

Some remarks on the estimation strategy are now in order. Of course, these choices are hardly representative of a real-world problem. However, in the experiments with this DGP we aim at maximum streamlining and ease of control, so the form of Equations (5a)–(5c) is assumed to be known at the expense of realism. In other terms, we assume lag length, type of deterministic terms (constant only, no trend), cointegration rank and shape of loadings matrix to be known. In this way elements of the estimation process, like the selection of lag length or (more critically) cointegrating rank, which are irrelevant for our purposes will have no effect on the results. We will thus be able to evaluate as precisely as possible the effect of the misspecification of the number of unit roots on the performances of the LR statistic. Note that, in doing so, we are following to a somewhat wider scale the standard practice of simulation experiments, as some aspects of the DGP are always taken as known in order to concentrate on the point of interest.[11] The analysis of a much more realistic DGP will be provided in section *The Bacchiocchi–Fanelli DGP*.

Moreover, the iterative two-step GLS switching algorithm is, unfortunately, quite cumbersome to apply here, because the system of constraints to be applied to the parameters involves non-linear restrictions which are complicated to handle by the iterative two-step technique. Therefore, we use the BFGS algorithm, as implemented in the gretl software library. Finally, the number of replications is 5000 in all cases.

Under each setting, we computed the Type I errors of asymptotic LR tests of H_0: $\mathrm{vec}(\zeta) = S\Phi_\zeta +$

[10]In Elliot (1998), the size distortions in the tests on the cointegration vectors are dependent on the cross correlation in the error term as this leads to endogeneity bias. However, inference on the cointegration vectors in the VAR is invariant to multiplication by full rank matrices.

[11]For instance, to the best of our knowledge, all simulation studies of tests on the cointegrating coefficients in the $I(1)$ VAR take the cointegrating rank as known, and this is also typically the case for the type of deterministic kernel.

s, with S and s defined so as to (i) have the cointegration vectors normalized, respectively, on c_t and l_t, and (ii) impose homogeneity (denoting as usual by β the 3×2 matrix of cointegrating coefficients, $\beta_{31} = -\beta_{11}$, $\beta_{32} = -\beta_{22}$). H_0 is thus imposed only on the coefficients of the level terms of the multicointegration relation.

Results

The results for this DGP are reported in Tables 4 and 5, respectively, for $I(2)$ and near-$I(2)$ data, and are easily summarized.

As Table 4 shows, the LR tests computed from the $I(2)$ model applied to genuine $I(2)$ data are somehow oversized in small samples. However, the size bias rapidly disappears as T increases, as dictated by the asymptotic distribution, but is still quite sizeable even with 256 observations. In Table 5, an extremely interesting result emerges: the Type I errors of the tests in the, theoretically misspecified, $I(2)$ model (headings $MI(2)$) are always closer to nominal values than the Type I errors of the tests in the, theoretically well-specified, $I(1)$ model (headings $MI(1)$). The differences shrink with sample size, but are still noticeable at $T = 256$. Remarkably, qualitatively, the same results hold even for other values of ρ quite far from unity (results not reported here, available on request).[12]

Table 4. H-US DGP: empirical size of LR tests on long-run coefficients – I(2) data (ρ = 1), I(2)-VECM model.

T	64		128		256	
nominal size	5.0	10.0	5.0	10.0	5.0	10.0
DGP_1	8.7	14.8	6.3	12.0	5.7	11.1
DGP_2	7.5	13.6	6.5	11.9	5.5	10.9
DGP_3	7.3	13.4	6.3	11.6	5.2	10.4
DGP_4	6.9	12.0	5.6	11.3	5.3	10.5
DGP_5	6.1	12.0	5.4	11.2	5.1	10.1
DGP_6	8.3	14.9	7.0	12.8	5.9	11.4
DGP_7	6.6	12.1	5.7	11.2	5.1	10.2

Table 5. H-US DGP: empirical size of LR tests on long-run coefficients – near-$I(2)$ data (ρ = 0.80,0.90).

	64		128		256	
T	MI (1)	MI(2)	MI (1)	MI (2)	MI(1)	MI(2)
α = 0.05, ρ = 0.90						
DGP_1	13.5	8.2	8.7	6.5	6.4	5.6
DGP_2	18.1	7.6	9.9	6.6	6.8	5.1
DGP_3	12.2	7.4	8.3	6.3	6.4	5.5
DGP_4	9.9	6.8	7.3	5.9	6.0	5.6
DGP_5	8.3	6.1	6.3	5.6	5.7	5.2
dgp_6	15.5	8.7	9.3	7.1	7.0	5.8
DGP_7	10.1	6.4	7.2	5.8	5.9	5.3
α = 0.05, ρ = 0.80						
DGP_1	13.9	8.1	8.6	6.6	6.7	5.6
DGP_2	19.5	7.2	11.5	6.5	7.7	5.6
DGP_3	11.9	7.2	8.2	6.4	8.2	6.4
DGP_4	10.1	6.6	7.0	5.6	6.2	5.5
DGP_5	8.7	6.2	6.8	5.3	5.6	5.2
DGP_6	16.3	8.5	10.0	6.9	7.0	6.0
DGP_7	10.2	7.1	6.7	5.7	6.0	5.3

where MI(1) is I(1)-VECM model and MI(2) is I(2)-VECM model

Table 6. H-US DGP: ∇ index of divergence between LR tests from the I(2)-VECM model on I(2) and near-I(2) data (ρ = 0.80, 0.90).

	ρ = 1 vs ρ = 0.80			ρ = 1 vs ρ = 0.90		
	T			T		
DGP	64	128	256	64	128	256
DGP_1	0.13	0.13	0.12	0.11	0.11	0.10
DGP_2	0.12	0.11	0.11	0.09	0.10	0.10
DGP_3	0.10	0.10	0.10	0.11	0.10	0.10
DGP_4	0.11	0.10	0.10	0.10	0.09	0.10
DGP_5	0.10	0.09	0.09	0.08	0.09	0.09
dgp_6	0.13	0.12	0.11	0.11	0.11	0.10
DGP_7	0.11	0.10	0.10	0.10	0.09	0.10

Therefore, on the basis of the simulation evidence, there are no doubts that if one focuses specifically on the task of testing hypothesis on the β's, $I(1)$ data with a second root very close to unity are best modelled as if they were $I(2)$. This conclusion is confirmed by the divergence measure ∇, the share of simulations in which the LR test yields opposite conclusions with $I(2)$ data and the closest near-$I(2)$ dataset; results are reported in Table 6. The ∇ index is generally very small, always in the order of 10%. Curiously enough, it seems to be rather insensitive to both ρ and T, which suggests that near-observational equivalence happens for this DGP in a rather large range of settings.

Finally, for the sake of completeness, we also checked power for a few selected alternatives. More precisely, we considered DGPs deviating from homogeneity by letting b_1 and b_2 take values different from 1. To save space, in Table 7 we report only the results for the case $b_1 = b_2 = 0.8$ with $\rho = 0.90$.

Remarkably, although rejection rates are always extremely high for both models, those from the $I(2)$ VECM are consistently, albeit marginally, higher. Considering that this model has also a smaller size bias, we can conclude that it definitely outperforms the theoretically well-specified $I(1)$ VECM (Table 8) in all respects.

[12]We thank Gunnar Bårdsen for his suggestion to try this.

Table 7. H-US DGP: power of LR tests on long-run coefficients – I(2) and near-*I*(2) data (ρ = 0.90), I(2)-VECM model.

	I(2)				*near-I(2)*			
	T = 64		T = 128		T = 64		T = 128	
DGP	5.0	10.0	5.0	10.0	5.0	10.0	5.0	10.0
DGP_1	97.1	97.9	99.9	100.0	90.1	92.5	99.6	99.8
DGP_2	98.4	98.8	100.0	100.0	96.6	97.8	99.9	99.9
DGP_3	99.9	99.9	99.9	100.0	99.7	99.8	100.0	100.0
dgp_4	99.9	100.0	100.0	100.0	99.7	98.8	100.0	100.0
DGP_5	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0
DGP_6	97.0	98.1	100.0	100.0	93.1	95.1	99.9	100.0
DGP_7	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0

DGP settings: b_1 = 0.8, b_2 = 0.8; H_0: $\beta_{31} = -\beta_{11}$, $\beta_{32} = -\beta_{22}$ T = 256: empirical power is always 100%, omitted.

Table 8. H-US DGP: Power of LR tests on long-run coefficients – near-*I*(2) data (ρ = 0.90), I(1)-VECM model.

	near-I(2)			
	T = 64		T = 128	
DGP	5.0	10.0	5.0	10.0
DGP_1	87.6	91.0	99.5	99.6
DGP_2	96.1	97.9	100.0	100.0
DGP_3	99.6	99.8	100.0	100.0
dgp_4	99.7	99.8	100.0	100.0
DGP_5	100.0	100.0	100.0	100.0
DGP_6	91.3	94.7	99.8	99.9
DGP_7	100.0	100.0	100.0	100.0

DGP settings: b_1 = 0.8, b_2 = 0.8; H_0:$\beta_{31} = -\beta_{11}$, $\beta_{32} = -\beta_{22}$ T = 256: empirical power is always 100%, omitted.

The Kongsted DGP

Set-up

The Kongsted DGP is a system with three variables, naturally interpreted as logs of money (m_t), nominal income (y_t) and prices (p_t). Using for convenience the same symbols as in Kongsted (2005),

$$m_t - y_t = (\delta - 1)y_{t-1} + [\gamma(\delta - 1) + \kappa\theta]\,\Delta p_{t-1} + \varepsilon_{1t} - \varepsilon_{2t} \tag{8a}$$

$$\Delta y_t = \gamma \Delta p_{t-1} + \varepsilon_{2t} \tag{8b}$$

$$\Delta p_t = \rho \Delta p_{t-1} + \varepsilon_{3t} \tag{8c}$$

Equation (8a) is a generalized velocity, Equation (8b) is the equation for nominal income, and (8c) is the price equation. The scenario here is that of a study of money and income data in which prices, and thus the variables in nominal terms, may be either pure $I(2)$ (hence, with non-stationary inflation) or near-$I(2)$ (hence, with stationary but long-swinging inflation). Since the integration order of inflation is often a subject of debate, it is interesting to examine with this DGP the properties of the test on β when ρ is smaller or equal to 1 with both the $I(1)$ and $I(2)$ VECMs. Letting $\mathbf{X}_t = (m_t, y_t, p_t)'$, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$, the VAR(2) representation is

$$\mathbf{X}_t = \begin{bmatrix} 0 & \delta & (\delta\gamma + \kappa\theta) \\ 1 & 1 & \gamma \\ 0 & 0 & 1+\rho \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} 0 & 0 & (\delta\gamma + \kappa\theta) \\ 0 & 0 & -\gamma \\ 0 & 0 & -\rho \end{bmatrix} \mathbf{X}_{t-2+\varepsilon_t}. \tag{9}$$

Like in the previous DGP, the value of ρ determines the order of integration of the system. It is implicitly set at 1 in Kongsted (2005), who studies a pure $I(2)$ case (which implies $r = 1$ and $s = 1$). There is a single-level cointegrating vector (from $I(2)$ to $I(1)$ when $\rho = 1$ and from $I(1)$ to $I(0)$ when $\rho < 1$), given by $\beta = (1, -\delta, 0)'$. Hence, H_0: $\beta = (1, -1, 0)'$ is the first hypothesis to be considered when testing the NtR transform[13].

As in Kongsted (2005) we set the DGP parameters as follows: $\kappa = -2$, $\delta = 1$ and $\gamma = 1$. We also take advantage of the close similarity of H-US and NtR GDPs to perform an informal (but still suggestive) robustness check of our results to the correlation structure of the disturbances, setting for this DGP the error term covariance matrix as

$$\Sigma = \begin{bmatrix} 1 & 0.1 & 0.025 \\ 0.1 & 0.9 & 0.02 \\ 0.025 & 0.02 & 0.075 \end{bmatrix}.$$

Results

The results for $I(2)$ and near-$I(2)$ DGPs are reported in Tables 10 and 12, respectively. All the simulation settings (errors, sample sizes, etc.) are as in the previous set of experiments. The near-$I(2)$ data are obtained by setting $\rho = 0.90$ and $\rho = 0.80$, and the dynamic properties of the companion form are reported in Table 9. For the sake of simplicity, power has been checked only for $\delta = 0.980$, the

Table 9. NtR DGP: Dynamic properties: non-zero eigenvalues of the companion matrix for I(2) and near-*I*(2) data (ρ = 0.90, 0.80).

I(2)	1.00	1.00	1.00
near-*I*(2)	1.00	1.00	0.90
near-*I* (2)	1.00	1.00	0.80

[13]The second set of restrictions, on the polynomial cointegrating vector, is required in the *I*(2) case to ensures that the (reduced) rank of the Π matrix in the *I*(2) system is preserved in the real-transformed system.

Table 10. NtR DGP: Empirical size and power of LR tests on long-run coefficients, I(2) data, I(2) – VECM model.

T	64		128		256	
nominal size	5.0	10.0	5.0	10.0	5.0	10.0
emp. size	6.04	11.48	5.56	10.80	5.86	11.14
emp. power	59.56	66.72	97.14	98.14	100	100

Power settings: $\beta = (1, -0.98, 0)$; H_0: $\beta_2 = 1$, $\beta_3 = 0$

most distant value from the null among those considered in Kongsted's 'Case 1'.

Let us discuss size first. As Table 10 shows, the Type I errors of the LR tests computed from the $I(2)$ model applied to genuine $I(2)$ data are close to the nominal level for all sample sizes. From Table 11 we can see that very much the same holds for the tests computed on $I(1)$ data, even with $\rho = 0.8$. In fact, although with near-$I(2)$ data the Type I errors of the tests from the two models are generally quite similar, those from the $I(1)$ one are always higher. The difference is especially noticeable for $T = 64$.

Therefore, simulation evidence seems to indicate that if one focuses specifically on the task of testing hypothesis on the β's, $I(1)$ data with a second root very close to unity are best modelled as if they were $I(2)$. Besides, the results we obtain seem to be remarkably robust with respect to moderate variations in the correlation in the reduced-form disturbance covariance matrix, as long as the relative orders of magnitude of the generated artificial data series remain comparable. This result, however, should be taken as only suggestive, given that our experimental setup is far from being comprehensive in this direction. As in the previous case, this conclusion is fully confirmed by the ∇ index (see Table 11), which takes always rather small values, with no evident link with ρ or sample size.

This conclusion is fully supported by simulation results on power. With $I(2)$ data (bottom row of Table 10), the ability to reject the false null examined is just acceptable for $T = 64$, but very high already for $T = 128$. With near-$I(2)$ data (Table 12) for both the $I(1)$ and the $I(2)$ models, the rejection rates are consistently lower: definitely disappointing for $T = 64$, good but not as in the $I(2)$ case for the larger sample sizes. The problem with this DGP is clearly due to the inadequate signal/noise ratio in small sample sizes, which affects adversely the ability of both models to reject false null hypothesis. The important message for our research question is again that adopting the $I(2)$ model with $I(1)$ data has no adverse consequences.

Table 11. NtR DGP:Empirical size and divergence measures of LR tests on long-run coefficients, near-I(2) data ($\rho = 0.80, 0.90$).

T / Nom. Size	64 *MI*(1)	64 *MI*(2)	128 *MI* (1)	128 *MI* (2)	256 *MI*(1)	256 *MI*(2)
$p = 0.90$						
0.05	7.94	5.96	6.32	5.82	5.02	5.12
0.10	14.44	11.44	12.80	11.14	10.06	10.66
Divergence ∇	0.09		0.09		0.10	
$p = 0.80$						
0.05	7.28	6.00	5.94	6.06	5.14	6.02
0.10	13.88	11.76	11.96	11.64	10.10	12.14
Divergence ∇	0.10		0.10		0.11	

where *MI*(1) is *I*(1)-VECM model and *MI*(2) is *I*(2)-VECM model

Table 12. NtR DGP: Empirical Power of LR tests on long-run coefficients, near-*I*(2) data ($\rho = 0.90,0.80$).

T / Nom. Size	64 *MI*(1)	64 *MI*(2)	128 *MI*(1)	128 *MI* (2)	256 *MI*(1)	256 *MI*(2)
$\rho = 0.90$						
0.05	34.62	35.94	77.06	79.26	97.40	97.72
0.10	44.60	45.64	82.64	84.36	98.44	98.48
$\rho = 0.80$						
0.05	23.00	22.76	56.48	57.62	88.32	88.78
0.10	31.96	31.28	64.32	65.88	91.38	92.08

where *MI*(1) is *I*(1)-VECM model and *MI*(2) is *I*(2)-VECM model
Power settings: $\beta = (1, -0.98, 0)$; H_0: $\beta_2 = 1$, $\beta_3 = 0$

The Bacchiocchi–Fanelli DGP

Set-up

Bacchiocchi and Fanelli (2005) studied PPP for a small panel of advanced economies over the post–Bretton Woods period and provided a time-series-based interpretation of the controversial evidence characterizing the dynamics of real exchange rates. More precisely, they showed that the persistent deviations from the PPP of the exchange rates between a set of European countries (France, Germany and UK) and the United States may be empirically attributed to the presence of $I(2)$ stochastic trends in prices measured using Consumer Price Indices. Their model in the integral-proportional equilibrium correction representation can be written as (Bacchiocchi and Fanelli (2005), Equation (17)):

$$\Delta^2 X_t = \alpha[\beta' X_{t-1} + \tilde{\delta} g' \Delta X_{t-1}] + \zeta_1^*(\beta' \Delta X_{t-1}) + \zeta_2^*(\beta'_1 \Delta X_{t-1}) + \epsilon_t \qquad (10)$$

where, $X_t = (e_t, p_t, p_t^*)$, e_t is the bilateral exchange rate (domestic/foreign) at time t, p_t and p_t^* are, respectively, the domestic and the foreign currency prices at time t.

The model estimated for the case USA vs UK appears particularly convenient as a simulation DGP for $I(2)$ datasets, as there are no short-run terms, $r = 1$ and $s = 1$. It will be referred to as B-F DGP and its parameters are as follows (source: Table V in Bacchiocchi and Fanelli 2005):

$$\alpha = \{-0.0298;\ -0.0025;\ -0.0049\}$$
$$\beta = \{1;\ 0.95;\ -0.42\}$$
$$\beta_1 = \{1;\ -0.75;\ 0.55\}$$
$$g = \{0;\ 0;\ 1\}$$
$$\tilde{\delta} = 66.7$$
$$\zeta_1^* = \{-1.71;\ -0.24;\ 0.17\}$$
$$\zeta_2^* = \{0.83;\ 0.24;\ -0.16\}$$
$$\Sigma = 0.001 \cdot \begin{bmatrix} 287.0 & -0.5013 & -5.2997 \\ -0.5013 & 1.5459 & 0.14 \\ -5.2997 & 0.14 & 9.6618 \end{bmatrix}$$

To generate the near-$I(2)$ artificial dataset closest to a given $I(2)$ model, not only that the matrix Π_2 must be of full rank, but the two models must also have the same number of cointegrating relationships. Unfortunately, contrary to the case of the H-US DGP, it is now not obvious how the set of parameters of the $I(2)$ DGP (see above) should be modified in order to satisfy these requirements. This problem is solved by using the VAR(2) representation in the following levels:

$$\mathbf{X}_t = A_1\mathbf{X}_{t-1} + A_2\mathbf{X}_{t-2} + \varepsilon_t$$

where the correspondence with Equation (1) is given by $A_1 = 2I + \Pi_1 + \Gamma$ and $A_2 = -(I + \Gamma)$.

Clearly, the rank restrictions of the $I(2)$ VECM imply restrictions on A_1 and A_2. We therefore aim for two transformations of A_1 and A_2, say, B_1 and B_2, which satisfy the rank restrictions implied by the $I(1)$ VECM. By fine-tuning these transformations, the desired near-$I(2)$ DGP is obtained. Our method is based on re-calibration of the original VAR matrices via two scalars h_1 and h_2, with two distinct purposes: (a) scaling the α matrix so as to control the speed of adjustment to the long-run equilibria without altering the rank of Π_1 and (b) altering the Γ matrix in such a way that Π_2 is closer/further away from being singular.

Table 13. B-F DGP: Dynamic properties, non-unit eigenvalues of the companion matrix.

$I(2)$	0.1426	0.4099	0.6432	1.0000
near-$I(2)$	0.3926	0.0885	0.1727	0.9740

In terms of Equation (1), $\boldsymbol{\alpha}$ is multiplied by some scalar $(1 + h_1)$ to increase the speed of adjustment to disequilibrium for $h_1 > 0$ without affecting the cointegration matrix $\boldsymbol{\beta}$; then, a scalar matrix $h_2 \cdot I$ is added to Γ in order to 'push Π_2 away' from singularity. Hence

$$B_1 = (1 + h_1)A_1 + h_1A_2 + (h_2 - h_1)I \qquad (11)$$

$$B_2 = A_2 - h_2I \qquad (12)$$

The relationship between h_1, h_2 and the roots of the companion matrix of the system proved to be highly non-linear and thus difficult to control, but a careful calibration of Equations (11–12) showed that setting $h_1 = 1$ and $h_2 = -0.5$ the rank conditions of the $I(1)$ VECM are satisfied and the largest unrestricted root is close to 1, as desired (see Table 13). We thus generated near-$I(2)$ data from the VAR(2) in levels with coefficient matrices B_1 and B_2 derived as in (11–12) and the same initial conditions and errors of the experiment with $I(2)$ data.

Sample sizes are $T = 64, 128, 256$ as before. Since this DGP is designed to mimic actual empirical work as closely as possible, the cointegrated $I(2)$ VECM is estimated using the switching algorithm assuming known-only cointegrating ranks ($s = 1$, $r = 1$) and lag length ($k = 2$), as would realistically happen in a true empirical exercise. Following point (iii) of the list of the DGP parameters, the null hypothesis of the LR test is that the second element of the β vector is equal to 0.95.

The number of Monte Carlo replication is 5000 as in the first set of experiments, but in a extremely small number of cases (always less than 1%) the reduced rank estimation could not be completed and the replication discarded.[14] The exact number of replications used to compute the rejection rates is reported in Table 14.

Results

The results are summarized in Tables 14 and 15, respectively, for $I(2)$ and near-$I(2)$ data. Unsurprisingly,

[14]Note: we do not estimate the model exactly as Equation (11); instead of the β_1 vector (which is orthogonal to β) what we estimate is some other γ vector (not necessarily orthogonal to β). For the purposes of testing the hypothesis of our interest, though, this is irrelevant.

Table 14. B-F DGP: Empirical size of LR tests on long-run coefficients – I(2) data, I(2)-VECM model.

T	64		128		256	
nominal size	5.0	10.0	5.0	10.0	5.0	10.0
	13.65	21.25	9.26	16.16	7.10	12.91
Completed replications	4974		4969		4958	

Table 15. B-F DGP: Empirical size of LR tests on long-run coefficients – near-$I(2)$ data.

T	64		128		256	
Nom. Size	$MI(1)$	$MI(2)$	MI (1)	MI (2)	$MI(1)$	$MI(2)$
0.05	23.44	11.60	14.49	8.53	9.06	7.28
0.10	32.95	19.22	22.86	14.61	15.65	12.69
Divergence ∇	0.20		0.15		0.11	

where $MI(1)$ is $I(1)$-VECM model and $MI(2)$ is $I(2)$-VECM model

performance is always inferior to the previous experiments 4.1.2, when estimation was carried out on a much more streamlined DGP, containing a much lower number of extra parameters.

With true $I(2)$ data the size bias of the test is much higher, but it shrinks quite rapidly as T increases, as expected. The striking result is that performance is essentially the same with near-$I(2)$ data, and much superior to those of the tests from the $I(1)$ VECM. Remarkably, the ∇ index, which measures the consequences of misspecification on the LR test in the $I(2)$ VECM, is higher than in the H-US DGP only for $T = 64$, 128, but it falls rapidly with sample size and for $T = 256$ it is practically the same for the two DGPs. This means that, even for sample sizes as large as 256, small-sample effects still dominate the adverse effects of misspecification.

Although surprising, this behaviour is entirely consistent with the Type I errors discussed earlier. With smaller sample sizes, both tests are oversized, but that from the $I(1)$ model more heavily so; as T grows, the rejection rates of both tests approach nominal sizes, and (consistently) the fraction of contrasting outcomes falls. In other terms, small sample bias always dominates model misspecification bias. We can then conclude[15] that even when simulating a much more realistic DGP two remarkable results emerge: first, the test from the $I(2)$ model does not seem to be adversely affected by the misspecification of the number of unit roots; second, it outperforms the analogous test computed in the $I(1)$-VECM even at sample sizes that an applied macro-economist would ordinarily consider perfectly adequate.

V. Conclusions

The main conclusion of our study is that, in $I(1)$ systems characterized by extremely high persistence in their stationary components (so-called near- $I(2)$ data), the best empirical strategy for the purpose of running hypothesis tests on the cointegration parameters may be to treat them as if they were $I(2)$. In other terms, it may make little sense to worry about telling genuine $I(2)$ processes apart from near-$I(2)$ ones, as finite-sample bias dominates the effects of misspecification. More specifically, using a number of observations in the order of magnitude of most macro-econometric applications, we found that the size properties of the LR test on the long-run coefficients computed in the $I(2)$ VECM estimated on $I(2)$ data are consistent with asymptotic results, and with near-$I(2)$ data are superior to those of the corresponding test computed in the $I(1)$ VECM. This happens in spite of the fact that the number of unit roots is correctly specified in the latter but not in the former. Power of the test from the $I(2)$ VECM also seems to be quite satisfactory.

These findings are of direct relevance for tests that are, in most applied problems including tests of the nominal-to-real transform, the most important application of hypothesis testing for linking the statistical model to the underlying economic theory apparatus. Further, it may be conjectured that they may extend to several other phases of inference.

A natural objection to these statements is that the special modelling requirements of near-$I(2)$ data would be better met using a well-specified $I(1)$ VECM with appropriate corrections if required, rather than the misspecified $I(2)$ VECM. However, this strategy would, in turn, open two new issues: first, it would still leave the 'long swings' in the data emphasized by Johansen et al. (2010) and Juselius (2013) poorly modelled. Second, we would forsake the rich modelling potential of the $I(2)$ model. In conclusion, our results provide considerable support to the proposal of using $I(2)$ VARs to model both $I(2)$ and near-$I(2)$ data.

[15]Power was not evaluated for this DGP: changing the long-run parameters would of course require a corresponding modification in the adjustment coefficients; the issue is quite complex and was left for future research.

The message to the practitioner therefore is: no, unit roots will not always take care of themselves. Actually, when testing for long-run properties you may have to choose between small-sample adverse effects and mis-specification from imposing extra constraints; and in this case, you had better go for the lesser of two evils, which could well be not so evil after all.

Acknowledgement

We would like to thank Gunnar Bårdsen, Luca Fanelli, Massimo Franchi, Søren Johansen, Katarina Juselius and two anonymous referees for their suggestions and criticisms. The usual disclaimers, of course, apply.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research supported by 'La Sapienza' [grant n. 2011C26A1145RM] and MIUR [PRIN grant 2010J3LZEN].

References

Bacchiocchi, E., and L. Fanelli. 2005. "Testing the Purchasing Power Parity through I(2) Cointegration Techniques." *Journal of Applied Econometrics* 20 (6): 749–770. doi:10.1002/(ISSN)1099-1255.

Boswijk, H. P. 2000. "Mixed Normality and Ancillarity in I (2) Systems." *Econometric Theory* 16: 878–904. doi:10.1017/S0266466600166046.

Box, G., and N. Draper. 1987. *Empirical Model-Building and Response Surfaces*. New York: Wiley.

Elliot, G. 1998. "On the Robustness of Cointegration Methods When Regressors Almost Have Unit Roots." *Econometrica* 66 (1): 149–158. doi:10.2307/2998544.

Frydman, R., and M. Goldberg. 2007. *Imperfect Knowledge Economics: Exchange rates and Risk*. Princeton: Princeton University Press.

Frydman, R., M. Goldberg, S. Johansen, and K. Juselius. 2012. *Long Swings in Currency Markets: Imperfect Knowledge and I(2) Trends*. White Plains: New York University.

Hendry, D., and T. Von Ungern-Stenberg. 1981. "Liquidity and Inflation Effects of Consumere Expenditure." In *Essays in the Theory and Measurement of Consumers' Behaviour*, edited by A. S. Deaton. Cambridge (UK): Cambridge University Press.

Johansen, S. 1992. "A Representation of Vector Autoregressive Processes Integrated of Order 2." *Econometric Theory* 8 (02): 188–202. doi:10.1017/S0266466600012755.

Johansen, S. 1997. "Likelihood Analysis of the I(2) Model." *Scandinavian Journal of Statistics* 24: 433–462. doi:10.1111/sjos.1997.24.issue-4.

Johansen, S. 2006a. "Confronting the Economic Model with the Data." In *Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, edited by D. Colander. Cambridge: Cambridge University Press.

Johansen, S. 2006b. "Statistical Analysis of Hypotheses on the Cointegrating Relations in the I(2) Model." *Journal of Econometrics* 132 (1): 81–115. doi:10.1016/j.jeconom.2005.01.024.

Johansen, S., K. Juselius, R. Frydman, and M. Goldberg. 2010. "Testing Hypotheses in an I(2) Model with Piecewise Linear Trends. an Analysis of the Persistent Long Swings in the Dmk/$ Rate." *Journal of Econometrics* 158 (1): 117–129. doi:10.1016/j.jeconom.2010.03.018.

Juselius, K. 2006. *The Cointegrated VAR Model*. Oxford: Oxford University Press.

Juselius, K. (2013). Testing for near I(2) trends when the signal to noise ratio is small. mimeo, University of Copenhagen.

Kongsted, H. 2003. "An I(2) Cointegration Analysis of Small-Country Import Price Determination." *The Econometrics Journal* 6 (1): 53–71. doi:10.1111/ectj.2003.6.issue-1.

Kongsted, H. 2005. "Testing the Nominal-To-Real Transformation." *Journal of Econometrics* 124 (2): 205–225. doi:10.1016/j.jeconom.2004.02.016.

Kurita, T., H. Nielsen, and A. Rahbek. 2011. "An I(2) Cointegration Model with Piecewise Linear Trends." *The Econometrics Journal* 14 (2): 131–155. doi:10.1111/ectj.2011.14.issue-2.

MacKinnon, J. G. 1996. "Numerical Distribution Functions for Unit Root and Cointegration Tests." *Journal of Applied Econometrics* 11: 601–618. doi:10.1002/(ISSN)1099-1255.

Mosconi, R., and P. Paruolo. 2013. *Identification of Cointegrating Relations in I(2) Vector Autoregressive Models*. White Plains, NY: Università dell'Insubria.

Ng, S., and P. Perron. 2001. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica* 69 (6): 1519–1554. doi:10.1111/ecta.2001.69.issue-6.

Nielsen, H., and C. Bowdler. 2006. "Inflation Adjustment in the Open Economy: An I(2) Analysis of UK Prices." *Empirical Economics* 31 (3): 569–586. doi:10.1007/s00181-005-0030-9.

Paruolo, P., and A. Rahbek. 1999. "Weak Exogeneity in I(2) Var Systems." *Journal of Econometrics* 93 (2): 281–308. doi:10.1016/S0304-4076(99)00012-3.

Rogoff, K. 1996. 'The Purchasing Power Parity Puzzle.' *Economic Literature* 34: 647–668.

Appendix

Dependent variable is the log difference of seasonally adjusted GDP volume index (OECD code: VIXOBSA). For uniformity, we used the 'constant only' variant throughout; the lag length was chosen via the modified Akaike criterion, as per Ng and Perron (2001); the second column reports the actual number of observation used, as data for some of the countries start later than 1982. *P*-values as per MacKinnon (1996).