Università Politecnica delle Marche

Dipartimento di Scienze Economiche e Sociali

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

Abstract

We review the $I(2)$ model in the perspective of its application to near-$I(2)$ data, and report the results of some Monte Carlo simulations on the small sample performance of asymptotic tests on the long-run coefficients in both $I(2)$ and near-$I(2)$ systems. Our findings suggest that these tests suffer from some finite-sample issues, such as size bias. However, the behaviour of these statistics is not markedly different in the $I(2)$ and near-$I(2)$ case at ordinary sample sizes, so the usage of the $I(2)$ model with near-$I(2)$ data is perfectly defensible in finite samples.

Indirizzo: Francesca Di Iorio: Università di Napoli "Federico II"
   Stefano Fachin: Università di Roma "La Sapienza"
   Riccardo Lucchetti: Università Politecnica delle Marche

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

1   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 for macroeconomic variables is often consistent both with economic theory and with observed evidence.

However, Johansen *et al.* (2010) and Juselius (2013) pointed out that in some empirically important cases (*e.g.*, exchange rates) the cointegrating relations estimated in $I(1)$ VARs show very 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-$I(2)$*.

Although the hypothesis of two unit roots was traditionally considered economically implausible, Frydman and Goldberg (2007) have convincingly argued that in systems involving interaction between the real and the financial sector (*e.g.*, those including exchange rates, assets and house prices) expectations based on imperfect knowledge will generate highly persistent dynamics resulting in near-$I(2)$ trends.

The key empirical question is then how such near-$I(2)$ systems should be modelled in practice. If the DGP is strictly speaking $I(1)$, obviously the choice of an $I(2)$ model is wrong *in theory*. However, *in practice* an $I(2)$ model could yield more sensible results if some the stable roots of the VAR polynomial are very close to unity. Johansen *et al.* (2010) argue for this approach, which yields the additional advantage of exploiting the ability of the $I(2)$ model of estimating separately short, medium and long-run error-correction behavior.[1]

[1]This could be very important for empirical studies, as in asset markets error-increasing

Although empirical applications of the $I(2)$ model are scant[2], its statistical foundations are now over two decades old (Johansen, 1992, 1997), and fairly well developed. However, before its the application to near-$I(2)$ data may be recommended we need to explore two open points. First, what are the finite-sample properties of test statistics in the $I(2)$ model? In order to investigate this issue we will concentrate on Likelihood Ratio tests on long-run coefficients, which are the key parameters of the system. Johansen (2006) showed that these tests are asymptotically χ^2, but very little is known on the applicability of this result to finite samples. Second, what are the consequences of misspecification of the number of unit roots, i. e. of the application of the $I(2)$ model to near-$I(2)$ data?[3] Although our main goal will be to shed light on the second question, we clearly need to explore the first one as well.

In order to frame the problem correctly, we first review the $I(2)$ model briefly (mainly, to establish notation) in section 2; then, in Section 3 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 4; section 5 summarises the results and concludes.

2 The I(2) and near-I(2) VAR

To establish notation, consider a p-variate VAR(2)

$$\Delta^2 \mathbf{X}_t = \Pi \mathbf{X}_{t-1} + \Gamma \Delta \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{pt})' \sim IID(0, \Sigma)$, and $t = 1, \ldots, T$; the matrix Π is assumed to satisfy the reduced rank restriction $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $p \times r$ matrices, with $p > r$. If, further, $\boldsymbol{\alpha}'_\perp \Gamma \boldsymbol{\beta}_\perp = \phi \boldsymbol{\eta}'$ has reduced rank s (with $s < p - r$), then $\mathbf{X}_t$ is an $I(2)$ process. Of course, 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 our context.

Mosconi and Paruolo (2013) recently showed that the following reparametrisation of (1) holds:

$$\Delta^2 \mathbf{X}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}' \mathbf{X}_{t-1} + \boldsymbol{v}' \Delta \mathbf{X}_{t-1}) + (\boldsymbol{\xi}\boldsymbol{\gamma}' + \boldsymbol{\varsigma}\boldsymbol{\beta}')\Delta \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2}$$

where the matrices $\boldsymbol{v}$ and $\boldsymbol{\gamma}$ are respectively $p \times r$ and $p \times s$. The terms $(\boldsymbol{\beta}' \mathbf{X}_{t-1} + \boldsymbol{v}' \Delta \mathbf{X}_{t-1})$ and $(\boldsymbol{\gamma}' + \boldsymbol{\beta}')\Delta \mathbf{X}_{t-1}$ are both stationary; the former

---

medium run relations may be associated with error-correcting long-run ones.

[2] *Inter alia*: Fanelli and Bacchiocchi (2005), Kongsted (2003), Nielsen and Bowdler (2006), and more recently the empirical section in Kurita, Nielsen and Rahbek (2011).

[3] In the following we will sometimes use for convenience "$I(2)$ data" as a shorthand for "datasets generated by $I(2)$ DGPs".

is known as multicointegration relation, or integral control term, the latter as medium run relation, or proportional control term. As we will be clear shortly, it is convenient to rewrite (2) as

$$\Delta^2 \mathbf{X}_t = \boldsymbol{\alpha}[\boldsymbol{\beta}' : \boldsymbol{v}'] \begin{bmatrix} \mathbf{X}_{t-1} \\ \Delta \mathbf{X}_{t-1} \end{bmatrix} + [\boldsymbol{\xi} : \boldsymbol{\varsigma}] \begin{bmatrix} \boldsymbol{\gamma}' \\ \boldsymbol{\beta}' \end{bmatrix} \Delta \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{3a}$$

$$= [\boldsymbol{\alpha} : \boldsymbol{\xi} : \boldsymbol{\varsigma}] \begin{bmatrix} \boldsymbol{\beta}' & \boldsymbol{v}' \\ \mathbf{0} & \boldsymbol{\gamma}' \\ \mathbf{0} & \boldsymbol{\beta}' \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1} \\ \Delta \mathbf{X}_{t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t \tag{3b}$$

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

Hypotheses on the cointegration parameters can be easily expressed as

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

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 tests on $\boldsymbol{\beta}$, because this matrix is by far the most important parameter as far as the interest is on the long-run properties of the system. Johansen (2006) showed that, in the case of hypothesis on all elements of the multicointegration relation, inference is not LAMN; however, likelihood ratio tests on the long-run structure parameters $\boldsymbol{\beta}$ are asymptotically χ^2, but their finite-sample properties may be very different.

3 The thought experiment

Before proceeding any further, let us restate our main question of interest. Assume that a cointegrated $I(2)$ VAR is estimated, applying all the standard $I(2)$ techniques; then, what are the properties of the LR test on $\boldsymbol{\beta}$ with data that are near-$I(2)$, as opposed to genuinely $I(2)$? Note that we are not interested at all in the issue of telling $I(2)$ and near-$I(2)$ datasets apart. Rather, the point of interest is assessing the properties of a certain testing procedure which assumes $I(2)$-ness, with both type of data.

When the data are actually $I(2)$, the test is applied within a correctly specified model, whereas when they are near-$I(2)$ the statistical model is misspecified. To check what the effects of misspefication are, we need to disentangle these effects from other sources of finite-sample problems so to be able to focus on the former. In order to do this, one should ideally have two datasets, which have exactly the same characteristics in every way, except for the second eigenvalue of the companion form of system (1); then the statistics of interest could be computed for both datasets and compared.

Clearly, this issue is impossible to settle by using real-world data: even if the data were actually generated by a process exactly like (1) the "counterfactual" dataset would not be available. On the other hand, a Monte Carlo

experiment allowing such a comparison is easily designed. Let us represent the simulated data in the n-th Monte Carlo replication (where $n = 1 \ldots N$ and $i = 1, 2$) as

$$\mathbf{X}_n^i = MC(\mathbf{X}_0, E_n, \boldsymbol{\theta}_i),$$

where $\mathbf{X}_n^i$ is the simulated dataset we obtain from formula (1) by using $\mathbf{X}_0$ as starting (pre-sample) values, E_n as the artificially-generated $T \times p$ matrix of disturbances and $\boldsymbol{\theta}_i$ a given set of parameter values. Define $\boldsymbol{\theta}_2$ as a set of parameters under which the DGP is $I(2)$ and $\mathbf{X}_n^2$ the resulting dataset obtained in the n-th simulation. Further, let τ_2 be the LR test for a given hypothesis on $\boldsymbol{\beta}$ computed in the $I(2)$ model estimated on $\mathbf{X}_n^2$. In the same way let $\boldsymbol{\theta}_1$ the parameter set such that the DGP is $I(1)$ but nearly $I(2)$, $\mathbf{X}_n^1$ and τ_1 respectively the resulting dataset and LR test in the $I(2)$ model estimated on $\mathbf{X}_n^1$. Since the other two elements of the DGP (the initial values, $\mathbf{X}_0$, and the shocks, E_n) are the same, we can safely ascribe any difference in the behaviour of the LR statistics τ_2 and τ_1 to model misspecification alone. 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 1.

Table 1: Acceptance/rejection decisions for the LR test

LR test on near-$I(2)$ data

		τ_1 accepts H_0	τ_1 rejects H_0	
LR test on	τ_2 accepts H_0	A	B	$A + B$
$I(2)$ data	τ_2 rejects H_0	C	D	$C + D$
		$A + C$	$B + D$	N

In order to interpret Table 1 correctly, it is essential to realise that, by performing this experiment, 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, comparing the behaviour of the same test statistic (LR) on datasets generated by two (slightly) different DGPs. Our aim is to understand how large the differences are. More precisely, since under H_0 the asymptotic distribution of τ_2 (the LR test on $I(2)$ data) is χ^2, we know that 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 τ_1 (the LR test on near-$I(2)$ data). Is its performance noticeably different from τ_2's? And if so, how?

To answer these questions, our attention will focus on B and C, which are the cases when the LR tests in models with correctly specified and misspecified number of unit roots deliver contrasting results: 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[4]. The key point is that the difference (B-C) measures the impact of model misspecification. To see why, consider an α-level test. The empirical Type I error of the τ_2 test ($I(2)$ data) is $(C+D)/N = \alpha$, while the proportion of non-rejections of the true null hypothesis is $(A+B)/N = 1-\alpha$. Next, consider τ_1 (near-$I(2)$ data). If model misspecification had no effects we would find its Type I error and proportion of non-rejections to be respectively $(B+D)/N = \alpha$ and $(A+C)/N = 1-\alpha$. It is easily seen that therefore $B=C$.

To summarise the differences in the outcomes delivered by the tests with $I(2)$ and near-$I(2)$ data we adapt to our case the following two indicators:

$$\nabla_a = (B + C)/N,$$

which is simply the proportion of simulations in which τ_2 and τ_1 yield opposite conclusions, and

$$\nabla_r = B \ln \left[\frac{2B}{B + C} \right] + C \ln \left[\frac{2C}{B + C} \right].$$

The ∇_r statistic can be seen either as the Kullbalck-Leibler distance of the table above from one with same marginals and $B = C$, and as an LR test for the equality of $P(B)$ and $P(C)$. It is then a measure of the relevance of misspecification effects: should we find that these are dominated by small sample bias we could conclude that for practical purposes the use of the LR tests in $I(2)$ models applied to near-$I(2)$ data is acceptable.

4 The experiment setup

If misspecification effects should turn out to be negligible, then using the LR test in the $I(2)$ model with near-$I(2)$ data would be justified, in practice, by the fact that the test statistic has optimal asymptotic properties in the "nearby" $I(2)$ DGP. Therefore, the first issue we will investigate is the appropriateness of asymptotic distributions in finite samples if the data are $I(2)$, as postulated. Secondly, we will use the same statistics on near-$I(2)$ data to ascertain how detrimental misspecification is for those properties.

Our experiments will be based on a Data Generating Process (DGP) derived from Johansen (1992), in turn inspired by Hendry and von Ungern-Stenberg's (1981) famous specification of the consumption function, also used by Paruolo and Rahbek (1999).

---

[4]The combination when τ_2 leads to the correct decision and τ_1 does not is relatively unsurprising (the test in the misspecified model works worse than in the right model); the opposite one (the test in the misspecified model outperforms the other one) is quite counterintuitive. However, in finite samples this can be the simple consequence of misspecification and sample biases of opposite sign compensating each other.

More precisely, we consider a system with three variables: consumption, liquid assets and income (all in natural logarithms: c_t, l_t and y_t respectively); the system is driven by one stochastic trend and linked by an Error Correction Mechanism (ECM):

$$\Delta c_t = \nu \Delta y_{t-1} - a_{11}(c_{t-1} - b_1 y_{t-1}) - a_{12}(l_{t-1} - b_2 y_{t-1}) + \varepsilon_{1t} \qquad (4a)$$

$$\Delta l_t = -a_{21}(c_{t-1} - b_1 y_{t-1}) + \varepsilon_{2t} \qquad (4b)$$

$$\Delta y_t = \rho \Delta y_{t-1} + \varepsilon_{3t} \qquad (4c)$$

Equation (4a) is a consumption function, equation (4b) describes asset accumulation through saving, and (4c) is a simple law of motion for GDP. The restrictions from economic theory one would want to test are the homogeneity conditions

$$H_0 : b_1 = b_2 = 1. \qquad (5)$$

For $\rho = 1$ the stochastic trend is $I(2)$, whereas for $\rho = 1 - \delta$ (where δ is some small positive number), it is near-$I(2)$. In the above model, $\rho = 1$ would imply that the growth rate of GDP is $I(1)$: while this is clearly implausible from an economic point of view, it is very possible that the growth rate is so persistent in a finite sample that ρ is observationally indistinguishable from unity.

In order to show that this is not merely a theoretical possibility, we computed 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; these are shown in Table 2.[5] Of course, we do not claim that this should be taken as a serious modelling exercise: we are fully aware that the results it contains may be spurious for a multitude of reasons (structural breaks, misspecification of the ADF equation, low power are the first that spring to mind) and therefore should not be taken at face value.

However, even with this caveat in mind, the number of cases in which we would be unable to reject the null hypothesis of a unit root in GDP growth 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 if we chose to interpret Table 2 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 economic intuition (which would treat GDP growth as stationary) with observed facts and rule out the possibility that GDP growth rates are in fact $I(1)$.[6]

---

[5]Dependent variable is the log difference of seasonally adjusted GDP volume index (OECD code: VIXOBSA). We use the variant with a constant only; 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).

[6]In fact, univariate unit root tests may have low power to detect a double unit root

Table 2: ADF tests on GDP growth rates

Country	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

After defining $\mathbf{X}_t = (c_t, l_t, y_t)'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$, under H_0 as per (5), the DGP can be compactly written in levels as

$$
\mathbf{X}_t = \begin{bmatrix} 1 - a_{11} & -a_{12} & \nu + a_{11} + a_{12} \\ -a_{21} & 1 & a_{21} \\ 0 & 0 & 2 - \delta \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} 0 & 0 & -\nu \\ 0 & 0 & 0 \\ 0 & 0 & \delta - 1 \end{bmatrix} \mathbf{X}_{t-2} + \boldsymbol{\varepsilon}_t. \quad (6)
$$

and in a manner compatible with equation (3a) as:

$$
\Delta^2 \mathbf{X}_t = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} -1 & 0 & \nu \\ 0 & -1 & 0 \\ 0 & 0 & -\delta \end{bmatrix} \Delta \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (7)
$$

where the matrix $\boldsymbol{\alpha}'_\perp \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp$ has reduced rank (and therefore the system is truly $I(2)$) only if $\delta = 0$. After imposing the $I(2)$ assumption, using (3b) we obtain the Mosconi-Paruolo representation, in which the structure of the integral control term and the medium run relationship involved by this model may be clearly appreciated:

$$
\Delta^2 \mathbf{X}_t = [\boldsymbol{\alpha} : \boldsymbol{\varsigma}] \begin{bmatrix} \boldsymbol{\beta}' & \boldsymbol{v}' \\ \mathbf{0} & \boldsymbol{\beta}' \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1} \\ \Delta \mathbf{X}_{t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t.
$$

where

$$
\boldsymbol{\alpha} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \\ 0 & 0 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \quad \boldsymbol{\varsigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \boldsymbol{v} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ \nu & 0 \end{bmatrix}.
$$

and $\boldsymbol{\gamma} = 0$.

In the simulation experiments the loadings a_{ij} have been chosen in such a way to control of the roots of the VAR, so to generate series with a wide range of diverse dynamic features; details of the seven combinations of loadings considered in the various experiments are given in Table 3 below. For all our Monte Carlo experiments, we set the number of replications to 5000.

To evaluate the properties of the tests in near-$I(2)$ systems, we ran a second set of experiments with the same loadings and $\rho = 0.90$, 0.95 and 0.99. The lagged effect of income growth on consumption growth, ν, is always fixed at 0.5. Disturbance terms are IID Gaussian with covariance matrix $\mathrm{diag}(1, 1, 0.1)$. These choices were made so to obtain simulated series whose relative order of magnitude resemble the ones observed in real-world data.

The sample sizes considered are $T = 64, 128, 256$; we consider these representative of the order of magnitude usually available in macroeconomic

---

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 the differenced process (see Juselius, 2013), so that these tests are likely to be somehow biased *against* the $I(2)$ hypothesis.

applications. It should be noted that we do not perform unconstrained estimation of all the parameters in the model, but we assume that the form of the three equations (4a)–(4c) is known. Therefore, several aspects of the estimated DGP are not derived from the data, but imposed *a priori*. More precisely, we assume lag length, type of deterministic terms and cointegration rank to be known, as routinely done in simulation experiments of cointegrated systems. Further, we also assume known the shape of loading matrix, so that we may constrain the three parameters a_{22}, a_{32} and a_{33} to zero.

Table 3: VAR Dynamic properties in different Monte Carlo settings - $\rho = 1$

Model	a_{11}	a_{12}	a_{21}	*Roots*			
M_1	0.20	0.20	-0.20	1	1	0.92	0.92
M_2	0.50	0.50	-0.50	1	1	0.87	0.87
M_3	0.80	0.80	-0.80	1	1	0.92	0.92
M_4	0.40	0.20	-0.20	1	1	0.80	0.80
M_5	0.80	0.20	-0.20	1	1	0.95	0.25
M_6	0.80	0.40	-0.40	1	1	0.60	0.60
M_7	0.80	0.60	-0.60	1	1	0.75	0.75

Roots: roots of the companion form of (1); two zero roots omitted

As argued in Section 3, we made this choice so to focus on our present aim, which is to evaluate if and how tests on the cointegration parameters exhibit worse properties because of the misspecification of the integration order, without blurring the results with the possible inaccuracies stemming from the fact that we estimate everything unrestrictedly. Moreover, this procedure makes it possible to use a simple one-step optimisation algorithm for estimation[7] instead of Mosconi and Paruolo's iterative two-step GLS switching algorithm, inevitably more complex and computationally demanding. This, in turn, makes it possible in practice to run a much larger number of Monte Carlo simulations, which is essential in this type of exercise. Under each setting, we computed the Type I errors of asymptotic Likelihood ratio tests of $H_0 : \text{vec}(\boldsymbol{\zeta}) = S\Phi_\zeta + s$, with S and s defined so to (*i*) have the cointegration vectors normalised respectively on c_t and l_t and (*ii*) impose homogeneity ($\beta_{31} = -\beta_{11}, \beta_{32} = -\beta_{22}$). H_0 is thus imposed only on the coefficients of the level terms of the multicointegration relation.

---

[7]We use the BFGS algorithm, as implemented in the `gretl` software library.

5 Experiment results and conclusions

The results of the simulations for the empirical size of τ_2 and τ_1 are reported in Tables 4 and 5, and easily summarised.[8] First of all, in the correctly specified case the rejection frequency is slightly higher than the nominal size, but seems to converge steadily to the nominal value as the sample size increases. In fact, even with $T = 64$ the size bias is often quite small, and decreases uniformly for all models considered. Moreover, and most notably, there seems to be little difference between the $I(2)$ and near-$I(2)$ cases. Even if the largest root considered here is 0.90, quite distant from unity, this is evident from the rejection rates and fully confirmed by the divergence measures reported in Table 6 and plotted in Fig. 1 and 2.

Table 4: Size of LR tests on long-run coefficients – $I(2)$ case, $\rho = 1$

	T	64		128		256	
	α	5.0	10.0	5.0	10.0	5.0	10.0
$Model$	M_1	7.1	13.0	5.6	10.3	5.5	10.2
	M_2	6.0	11.8	5.4	10.2	5.7	10.7
	M_3	6.1	11.5	5.4	10.3	5.3	10.6
	M_4	8.5	14.7	6.2	11.9	5.8	11.4
	M_5	7.2	13.2	5.5	11.3	5.7	10.9
	M_6	7.2	13.3	5.6	11.1	5.3	10.4
	M_7	6.2	12.6	5.3	10.6	5.5	10.6

Table 5: Size of LR tests on long-run coefficients – near-$I(2)$ case, $\rho = 0.90$

	T	64		128		256	
	α	5.0	10.0	5.0	10.0	5.0	10.0
$Model$	M_1	7.2	13.2	5.6	11.3	5.7	11.0
	M_2	6.4	11.8	5.5	11.3	5.8	11.2
	M_3	5.8	11.5	5.2	10.9	5.3	10.3
	M_4	8.1	15.1	6.7	12.3	6.2	11.6
	M_5	7.0	13.0	6.3	11.7	5.7	10.6
	M_6	7.9	13.6	6.6	12.1	6.3	11.8
	M_7	6.1	12.6	5.6	10.9	5.7	11.6

First of all, let us examine the values of ∇_a in Fig. 1, which is the proportion of simulations in which the LR tests with $I(2)$ data and the closest near-$I(2)$ dataset yield opposite conclusions. This proportion behavies

---

[8]For the sake of conciseness, we report, under the "near-$I(2)$" heading, only the results for $\rho = 0.90$; the other tables are, of course, available on request.

Table 6: Distances ∇_a and ∇_r between the test performances with $I(2)$ and near-$I(2)$ data

| ρ | Model | ∇_a | | | ∇_r | | |
| | | T | | | T | | |
		64	128	256	64	128	256
0.99	M1	0.01	0.02	0.03	0.04	0.04	0.99
	M2	0.01	0.02	0.03	0.04	0.53	0.44
	M3	0.01	0.02	0.03	0.16	2.24	2.22
	M4	0.01	0.02	0.03	0.04	0.01	0.39
	M5	0.01	0.02	0.04	0.05	0.59	0.17
	M6	0.01	0.02	0.03	0.12	0.45	2.91
	M7	0.01	0.02	0.04	1.31	0.59	0.63
0.95	M1	0.05	0.06	0.07	0.03	0.43	1.28
	M2	0.05	0.05	0.07	0.57	0.19	0.00
	M3	0.04	0.03	0.06	0.20	0.06	0.65
	M4	0.05	0.05	0.06	1.92	0.33	0.05
	M5	0.05	0.06	0.08	1.36	0.72	0.02
	M6	0.05	0.06	0.06	0	0.99	0.95
	M7	0.04	0.05	0.07	2.66	0.99	0.10
0.90	M1	0.08	0.08	0.07	0.02	0.29	0.75
	M2	0.06	0.07	0.08	0.53	0.10	0.24
	M3	0.06	0.06	0.07	0.00	0.02	0.34
	M4	0.08	0.07	0.07	1.41	0.86	0.02
	M5	0.08	0.08	0.09	0.01	1.05	0.03
	M6	0.07	0.07	0.08	1.22	2.99	1.36
	M7	0.06	0.06	0.07	2.41	3.82	1.27

Figure 1: Type I errors in $I(2)$ and near$-I(2)$ DGPs: ∇_a distance

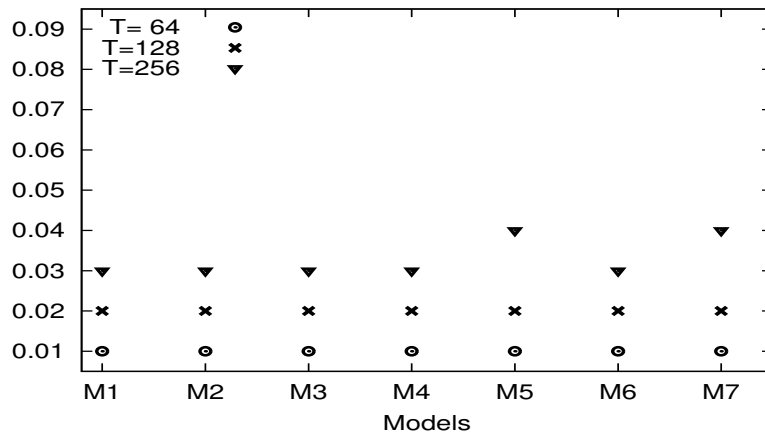$\rho = 0.90$
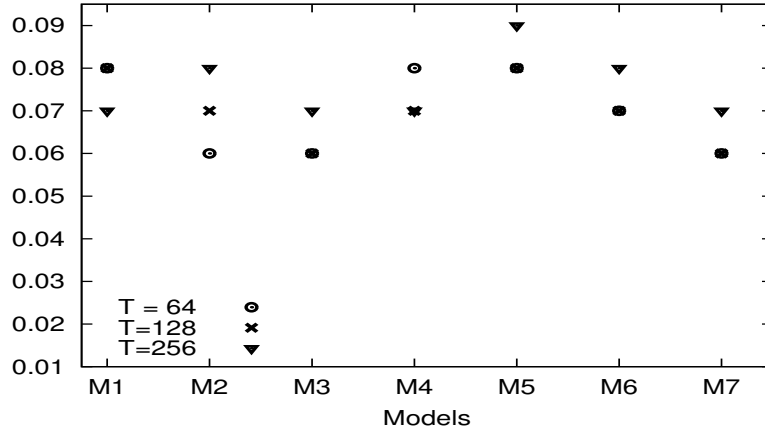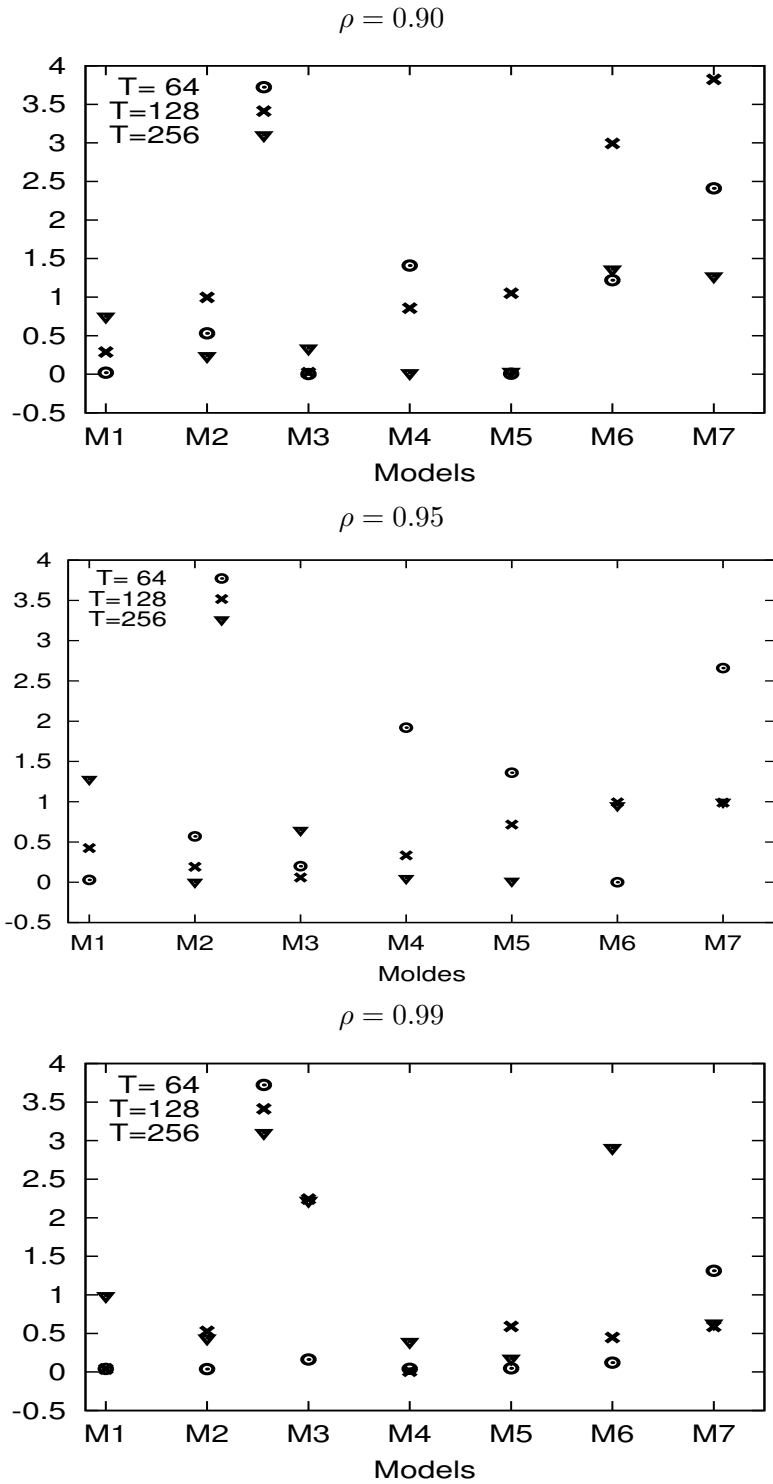

$\rho = 0.95$
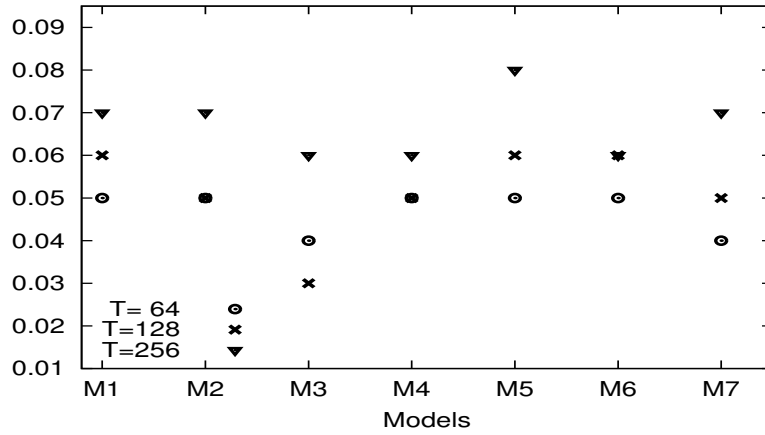
$\rho = 0.99$

Figure 2: Type I errors in $I(2)$ and near$-I(2)$ DGPs: ∇_r distance

$\rho = 0.90$

$\rho = 0.95$

$\rho = 0.99$

largely as expected: first of all, it is generally very small (always less than 10%). Second, it decreases as ρ approaches unity, and, for given ρ, with T.

The values of ∇_r are more difficult to assess. From Fig. 2, the distances appear to be approximately randomly distributed, with no clear effects of sample size, ρ and DGP. M7 may appear as possible exception, but a closer inspection shows that the conclusion that ∇_r does not diverge with T holds for this DGP also. In fact, the distance for $T = 256$ is much smaller than those for $T = 64$ and 128 and entirely in line with those found for other combinations of DGP and ρ. Simulations with $T = 512$, not reported here, confirmed for all combinations of ρ and DGP that ∇_r does not diverge with T. Hence, in our simulations the misspecification bias is indeed dominated by the small sample bias.

This finding suggests that with a number of observations in the order of magnitude of most macro-econometric applications, it makes little sense to worry about telling "real" $I(2)$ processes from near-$I(2)$, as finite-sample bias dominates the effects of misspecification. Tables 7 and 8 suggest that similar results hold for power against a trivial alternative. It may be surmised that some bootstrap version of the test may offer even better properties.

The answer to our original question, "Can you do the wrong thing and still be right?" is, then, absolutely yes. Clearly, one could conceivably argue that it would be better still to "do the right thing" from the outset, that is, use the $I(1)$ model with appropriate corrections for the near-$I(2)$-ness of the data. For instance, since in this set-up many statistics of the $I(1)$ model may have distributions poorly approximated by asymptotic ones, these may replaced by bootstrap distributions. However, this strategy would, in turn, open two new issues: first, it would still leave the "long swings" in the data poorly modelled. Second, we would forsake the rich modelling potential of the $I(2)$ model, for instance the ability to estimate separate error correction behaviour for different time horizons.

In conclusion, our results provide strong support to the use of asymptotic LR tests on the long-run coefficients of the $I(2)$ model with both $I(2)$ and near-$I(2)$ data. However, it should be kept in mind that our results are somewhat partial, as we assume full knowledge of the structure of the VAR (lag length, cointegration ranks and exogeneity status of the different variables). A more thorough investigation in a more realistic simulation setting is the subject of ongoing work.

Table 7: Power of LR tests on long-run coefficients – $I(2)$ and near-$I(2)$ case, $\rho = 0.90$

| | $I(2)$ | | | | $near\text{-}I(2)$ | | | |
| | $T = 64$ | | $T = 128$ | | $T = 64$ | | $T = 128$ | |
$Model$	5.0	10.0	5.0	10.0	5.0	10.0	5.0	10.0
M_1	88.1	91.6	99.9	100.0	69.2	76.9	97.7	98.7
M_2	99.7	99.9	100.0	100.0	97.5	98.5	99.0	100.0
M_3	100.0	100.0	100.0	100.0	99.8	99.9	100.0	100.0
M_4	90.2	93.8	99.9	100.0	77.8	84.8	99.3	99.6
M_5	99.0	99.5	100.0	100.0	98.2	96.9	100.0	100.0
M_6	98.6	99.0	100.0	100.0	96.0	98.2	100.0	100.0
M_7	99.8	99.8	100.0	100.0	99.4	99.2	100.0	100.0

Power settings: $\beta_{31} = 0.8$ and $\beta_{32} = 0.8$; $H_0 : \beta_{31} = -\beta_{11}, \quad \beta_{32} = -\beta_{22}$

Table 8: Power of LR tests on long-run coefficients - $I(2)$ and near-$I(2)$ case $\rho = 0.90$

| | $I(2)$ | | | | $near\text{-}I(2)$ | | | |
| | $T = 64$ | | $T = 128$ | | $T = 64$ | | $T = 128$ | |
$Model$	5.0	10.0	5.0	10.0	5.0	10.0	5.0	10.0
M_1	82.9	87.5	99.9	100.0	58.9	67.2	94.3	96.3
M_2	99.2	99.6	100.0	100.0	94.7	96.6	99.9	99.9
M_3	99.2	99.9	100.0	100.0	99.4	99.6	100.0	100.0
M_4	78.8	85.3	100.0	100.0	59.9	69.1	96.3	97.7
M_5	96.9	98.2	100.0	100.0	92.7	94.9	100.0	100.0
M_6	97.0	98.2	100.0	100.0	89.7	92.7	99.5	100.0
M_7	99.6	99.8	100.0	100.0	97.2	98.1	99.9	100.0

Power settings: $\beta_{31} = 0.8$ and $\beta_{32} = 1.0$; $H_0 : \beta_{31} = -\beta_{11}, \quad \beta_{32} = -\beta_{22}$

References

FANELLI, L. AND E. BACCHIOCCHI (2005): "Testing the purchasing power parity through $I(2)$ cointegration techniques", *Journal of Applied Econometrics*, 20(6), 749–770.

FRYDMAN, R. AND M. GOLDBERG (2007): *Imperfect Knowledge Economics: Exchange rates and Risk*. Princeton University Press, Princeton.

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*, ed. by A. S. Deaton. Cambridge University Press, Cambridge (UK).

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

——— (1997): "Likelihood analysis of the $I(2)$ model", *Scandinavian Journal of Statistics*, 24, 433–462.

——— (2006): "Statistical analysis of hypotheses on the cointegrating relations in the $I(2)$ model", *Journal of Econometrics*, 132(1), 81–115.

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.

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", *Econometrics Journal*, 6(1), 53–71.

KURITA, T., H. NIELSEN AND A. RAHBEK (2011): "An $I(2)$ cointegration model with piecewise linear trends", *Econometrics Journal*, 14(2), 131–155.

MACKINNON, J. G. (1996): "Numerical Distribution Functions for Unit Root and Cointegration Tests", *Journal of Applied Econometrics*, 11, 601–618.

MOSCONI, R. AND P. PARUOLO (2013): "Identification of cointegrating relations in $I(2)$ vector autoregressive models", WP, 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.

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.

PARUOLO, P. AND A. RAHBEK (1999): "Weak exogeneity in $I(2)$ VAR systems", *Journal of Econometrics*, 93(2), 281–308.