



F U N D A Ç Ã O
GETULIO VARGAS

EPGE

Escola de Pós-Graduação
em Economia

Ensaio Econômico

Escola de

Pós-Graduação

em Economia

da Fundação

Getúlio Vargas

Nº 763

ISSN 0104-8910

Forecasting Multivariate Time Series under Present-Value-Model Short- and Long-run Co-movement Restrictions

Osmani Teixeira de Carvalho Guillén, Alain Hecq, João Victor Issler,
Diogo Saraiva

Março de 2015

URL: <http://hdl.handle.net/10438/13540>

Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getúlio Vargas.

ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA

Diretor Geral: Rubens Penha Cysne

Vice-Diretor: Aloisio Araujo

Diretor de Ensino: Carlos Eugênio da Costa

Diretor de Pesquisa: Humberto Moreira

Vice-Diretores de Graduação: André Arruda Villela & Luis Henrique Bertolino Braidó

Teixeira de Carvalho Guillén, Osmani

Forecasting Multivariate Time Series under
Present-Value-Model Short- and Long-run Co-movement
Restrictions/ Osmani Teixeira de Carvalho Guillén, Alain Hecq,
João Victor Issler, Diogo Saraiva - Rio de Janeiro : FGV,EPGE,
2015

45p. - (Ensaaios Econômicos; 763)

Inclui bibliografia.

CDD-330

FORECASTING MULTIVARIATE TIME SERIES UNDER PRESENT-VALUE-MODEL SHORT- AND LONG-RUN CO-MOVEMENT RESTRICTIONS

Osmani Teixeira de Carvalho Guillén

Alain Hecq

Banco Central do Brasil and IBMEC-RJ

Maastricht University

João Victor Issler*

Diogo Saraiva

Getulio Vargas Foundation

Getulio Vargas Foundation

February 26, 2015

Abstract

Using a sequence of nested multivariate models that are VAR-based, we discuss different layers of restrictions imposed by present-value models (PVM hereafter) on the VAR in levels for series that are subject to present-value restrictions. Our focus is novel – we are interested in the short-run restrictions entailed by PVMs (Vahid and Engle, 1993, 1997) and their implications for forecasting.

Using a well-known database, kept by Robert Shiller, we implement a forecasting competition that imposes different layers of PVM restrictions. Our exhaustive investigation of several different multivariate models reveals that better forecasts can be achieved when restrictions are applied to the unrestricted

*Corresponding author: João Victor Issler, Graduate School of Economics - EPGE, Getulio Vargas Foundation, Praia de Botafogo 190 s. 1100, Rio de Janeiro, RJ 22250-900, Brazil. E-mail: Joao.Issler@fgv.br.

VAR. Moreover, imposing short-run restrictions produces forecast winners 70% of the time for the target variables of PVMs and 63.33% of the time when all variables in the system are considered.

JEL Codes: C22, C32.

Keywords: forecasting, multivariate models, vector autoregression (VAR), present-value restrictions, common cycles, cointegration, interest rates, prices and dividends.

1 Introduction

The use of multivariate time series models in economics and finance has proved fruitful, since they entail key inter-relationships between the variables being modelled. Unfortunately, most of these models have an abundance of free parameters, which poses a problem when they are used for forecasting, since their forecast accuracy measures are usually outperformed by those of more parsimonious alternatives. One way to cope with this forecasting problem is to impose restrictions, thus reducing the number of free parameters in forecasting models. This is often done for small-dimension vector autoregressive (VAR) models by testing and imposing long-run relationships among the series being modelled when they individually trend and jointly co-trend over time (see Engle and Granger, 1987; Johansen, 1988). One can also impose further similarities in their short-run dynamics, e.g., impose common cyclical feature restrictions (see Engle and Kozicki, 1993; Vahid and Engle, 1993).

The extensive work on cointegration has indeed shown that considering and imposing long-run relationships leads to forecasting gains compared to the model in first differences (see also Clements and Hendry, 1998 or Hoffman and Rasche, 1996, *inter alia*). However, less than a handful of papers (e.g., Issler and Vahid, 2001, Vahid and Issler, 2002, Anderson and Vahid, 2011) have investigated whether additional short-run co-movement restrictions generate better forecasts. Moreover, Athanasopoulos et al. (2011) only recently compared the relative importance of these two types of

restrictions using simulations and real data, and showed that existing short-run restrictions have a greater potential to improve the forecast accuracy than cointegration restrictions.

Short and long-run restrictions are implied by the present-value model (PV model or PVM, hereafter) introduced by Campbell and Shiller (1987) and studied here. However, most papers have focused on the presence of cointegration between the levels of two variables (labeled Y_t and y_t in this paper), a condition that is necessary for the validity of a present-value model linking them.¹ Hence, it is often overlooked that *another* necessary condition for the PVM to hold is that the forecast error implied by the PV model be orthogonal to the past. We refer to the studies by Hansen and Sargent (1981, 1993) and Baillie (1989) for initial work on rational expectations linked to PVMs, and Johansen and Swensen (1999, 2004, 2011) and Johansen (2000) for a recent fresh look on the subject.

Indeed, PVMs arise from a first-order stochastic difference equation, where its error term must be unforecastable regarding past information, i.e., it must have a zero conditional expectation. This is exactly what the common cyclical feature framework implies. If this fails, the PV equation will not be valid, since it will contain an additional term that captures the (non-zero) conditional expected value of all future error terms. Cointegration imposes the transversality condition allowing the limit $I(0)$ combination of Y_t and y_t to be discarded. The existence of an unforecastable linear combination of the $I(0)$ series in the difference equation guarantees that the dynamic behavior of the variables in the PVM will conform to theory.

Since we need both conditions in order to validate PVMs, we will ideally work with an integrated econometric framework that encompasses the joint existence of these two phenomena. This is the starting point of this article. We first show that PV relationships entail a weak-form common feature restriction, as per Hecq et al. (2006) and Athanasopoulos et al. (2011), for the vector error-correction model

¹Examples of Y_t and y_t include prices and dividends for a given asset, long- and short-term interest rates, and consumption and disposable income, respectively. If they are integrated processes, they will cointegrate. See also the examples from Campbell (1987) and Campbell and Deaton (1989), *inter alia*, which are reviewed by Engsted (2002), and the interesting recent contribution of Johansen and Swensen (2011).

(VECM) for Y_t and y_t . Alternatively, it also implies a polynomial serial correlation common feature relationship (Cubadda & Hecq, 2001) for the VAR representations of Δy_t and the cointegrating relationship $Y_t - \theta y_t$. These represent short-run restrictions on the dynamic system for these variables. Once we cast the PVM in these terms, it is straightforward to apply the toolkit of the *common-feature* literature for inference and testing.

Our second contribution relates to the forecasting of series that are subject to PVM restrictions. We show the relevance of the issues discussed above in an empirical exercise involving two sets of financial series. The first contains annual long- and short-maturity interest rates for the U.S. economy. The second contains real price and dividends for the S&P composite index and the real risk-free rate. Both data sets were extracted from the online library maintained and updated by Shiller (<http://www.econ.yale.edu/~shiller/data.htm>), with 142 annual observations spanning the period 1871–2012. We are careful to consider the different layers of restrictions discussed in the PVM literature: long-run restrictions (cointegration), short-run restrictions (weak-form common cycles), long- and short-run restrictions jointly, and the latter with additional specific parameter restrictions implied by economic theory. Each layer corresponds to a specific restricted representation for the reduced form VAR/VECM. Forecast accuracy measures across representations are compared in order to evaluate the benefits of imposing each set of restrictions. The final results confirm the importance of imposing short-run restrictions. Indeed, for target variables in PVMs (Y_t), forecasting models that allow for and/or impose these restrictions produce winners in 70% of cases at horizons from 1 to 5 years ahead. Overall, for Y_t and y_t , they produce winners 63.33% of the time at these same horizons.

Our last contribution is to devise a testing strategy for PV restrictions in macroeconomics and finance, incorporating more than 20 years of research on this topic. We cover several important issues. First, we discuss how to choose the lag length of the VAR consistently. Second, we discuss how to test for cointegration, common cycles, and weak-form common cycles, using a multivariate approach based on the likelihood ratio test (canonical correlation analysis) and a single-equation heteroskedasticity robust approach (GMM). Part of our suggested strategy relies

on Monte-Carlo simulation results. Finally, we also suggest integrated approaches estimating the lag length of the VAR and the long-run and short-run parameters jointly as per Athanasopoulos et al. (2011). Alternatively, we also discuss estimating long-run and short-run parameters jointly as per Centoni, Cubadda and Hecq (2007). In order to avoid using too much space in a forecasting paper with testing and estimation issues, these are discussed in the Appendix.

The rest of the paper is divided as follows. Section 2 reviews PV formulas (for both the levels and log-levels of the variables) and discusses the types of restrictions PVMs imply for the VECM, as well as for a transformed VAR. In Section 3, we present an in-sample analysis of the data used in the forecasting experiment, verifying whether the restrictions implied by economic theory hold in practice. In Section 4 we compare the forecasting gains obtained by imposing different types of PV restrictions in multivariate models. Section 5 concludes. The Appendix contains additional material on how to select the lag-length of the VAR in our context, how to implement different tests for PVMs, including their small-sample performance, and other relevant issues for examining PVM restrictions. We also present an online appendix only with self-contained material on common-cyclical features for cointegrated data.

2 Present-value models

2.1 Nesting the representation in levels with long- and short-run co-movement

Consider the present value equation $Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$, where, for simplicity of presentation, we drop the constant term. Y_t is a linear function of the present discounted value of the expected future y_t , where $\mathbb{E}_t(\cdot)$ is the conditional expectation operator, using information up to t as the information set. In most cases, Y_t and y_t are $I(1)$ variables. Examples of Y_t and y_t include, respectively: long- and short-term interest rates, real stock prices and real dividends, personal consumption and disposable income, etc. (see the survey by Engsted, 2002). In this subsection, we assume constant expected returns with a discount factor of $\delta = \frac{1}{1+r}$, where r is a time-

invariant interest rate. The coefficient θ is a factor of proportionality. For example, $\theta = \delta/(1 - \delta)$ in the price-dividend relationship in levels while $\theta = 1$ for the interest rates case, and the link with the discount factor is given by the term structure of the interest rates (see, *inter alia*, Chow, 1984; Campbell & Shiller, 1987; Johansen & Swensen, 2011). The choice of θ only impacts the value of the cointegrating vector. Hence, in what follows, we set its value equal to $\theta = \delta/(1 - \delta)$, such that:

$$Y_t = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}. \quad (1)$$

Following Campbell and Shiller (1987), the actual spread is defined as:

$$S_t := Y_t - \frac{\delta}{1 - \delta} y_t = Y_t - \theta y_t, \quad (2)$$

where S_t is $I(0)$ if Y_t and y_t are cointegrated. Subtracting $\frac{\delta}{1 - \delta} y_t$ from both sides of Eq. (1) produces the theoretical spread $S'_t := \frac{\delta}{1 - \delta} \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i}$, where:

$$S_t = Y_t - \theta y_t = \frac{\delta}{1 - \delta} \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i} = S'_t. \quad (3)$$

This shows that the series Y_t and y_t must be cointegrated, because the right-hand side is a function of $I(0)$ terms with exponentially decreasing weights. Also, S (or S') must help to predict Δy , i.e., must *Granger-cause* Δy .

If Y_t and y_t are $I(1)$ series, we discuss what types of restrictions PV relationships impose in multivariate models for $\begin{pmatrix} Y_t \\ y_t \end{pmatrix}$. Our starting point is the vector autoregressive (VAR) model, for simplicity of exposition. However, the concepts discussed here are broader, and also apply to a more general class of models. A bi-variate VAR of order p on these series takes the form:

$$\begin{pmatrix} Y_t \\ y_t \end{pmatrix} = \Gamma_1^* \begin{pmatrix} Y_{t-1} \\ y_{t-1} \end{pmatrix} + \dots + \Gamma_p^* \begin{pmatrix} Y_{t-p} \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}. \quad (4)$$

Under cointegration for Y_t and y_t as in Eq. (2), we can re-write Eq. (4) as:

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \Gamma_1 \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots + \Gamma_{p-1} \begin{pmatrix} \Delta Y_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} S_{t-1} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \quad (5)$$

where the Γ_i s are the short-run coefficient matrices, $\Gamma_j = -\sum_{i=j+1}^p \Gamma_i^*$, and α_1 and α_2

are the loadings on the error-correcting term, which entails $-\left(I - \sum_{i=1}^p \Gamma_i^*\right) \begin{pmatrix} Y_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1, -\theta) \begin{pmatrix} Y_t \\ y_t \end{pmatrix}$.

As is well known, PV relationships as in Eq. (1) imply additional restrictions on dynamic models of the data. Campbell and Shiller (1987) and Johansen and Swensen (1999) exploited the fact that VARs (Eq. (4)) have cross-equation restrictions. Here, we exploit a different aspect of these *same* restrictions – the existence of reduced-rank restrictions for the short-run coefficient matrices $\Gamma_1, \Gamma_2, \dots, \Gamma_{p-1}$ in the VECM (Eq. (5)).

To see this, move forward Eq. (1) one period, take $\mathbb{E}_t(\cdot)$, and subtract Eq. (1) from the results; we then obtain a third spread $S_t'' := \frac{\delta}{1-\delta} \mathbb{E}_t \Delta Y_{t+1}$, where $S_t = S_t' = S_t''$:

$$S_t'' \frac{1-\delta}{\delta} = \mathbb{E}_t \Delta Y_{t+1} = \delta \sum_{i=0}^{\infty} \delta^{i-1} \mathbb{E}_t \Delta y_{t+i+1} = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i} = S_t \frac{1-\delta}{\delta}, \quad (6)$$

which shows that the conditional expectation of forward changes in Y depends only on the lagged S , i.e., $\mathbb{E}_{t-1} \Delta Y_t = S_{t-1} \frac{1-\delta}{\delta}$. Of course, this constrains the dynamics of the VECM (Eq. (5)) with exclusion restrictions. We summarize this result using a proposition.

Proposition 1. (Campbell and Shiller, 1987, and Johansen and Swensen, 1999) If the elements of $(Y_t, y_t)'$ obey a PV relationship as in $S_{t-1} = \frac{\delta}{1-\delta} \Delta Y_t + u_t$, then their VECM obeys a weak-form common feature relationship (see Hecq et al., 2006, and Athanasopoulos et al., 2011): there exists a 1×2 vector γ' such that $\gamma' \Gamma_1 = \gamma' \Gamma_2 =$

.... = $\gamma' \Gamma_{p-1} = 0$, but $\gamma' \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \neq 0$. Moreover, $\gamma' = (1 : 0)$, the first row of every Γ_i , $i = 1 \dots p-1$, must be zero, and the following restriction must also be met: $\alpha_1 = \frac{1-\delta}{\delta}$.

Proposition 1 simply states that the short-run dynamics of the VECM in Eq. (5) is constrained by the existence of weak-form serial-correlation common features, making ΔY_t a linear function of S_{t-1} alone, apart from an unpredictable error term. As the proposition makes clear, the final restricted VECM is nested in Eq. (5). So, going from Eqs. (4) to (5), and then to the restricted-VECM in Proposition 1, shows a sequence of nested models, including all of the restrictions entailed by PVMs at the end.

An alternative route for nesting models is to consider the cross-equation restrictions studied by Campbell and Shiller (1987) using a transformed VAR on $\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix}$; see Johansen and Swensen (2011). To motivate this discussion, we recall that the PVM is usually generated by a first-order stochastic difference equation for the state variable S_t . Lag Eq. (6) by one period and rearrange it to obtain:

$$(1 - \delta)S_{t-1} = \delta \Delta Y_t + (1 - \delta)u_t, \quad (7)$$

where, by construction, $\mathbb{E}_{t-1}(u_t) = 0$. Recall that $\delta S_{t-1} = \delta(Y_{t-1} - \frac{\delta}{1-\delta}y_{t-1})$ and that $\delta \Delta Y_t = \delta Y_t - \delta Y_{t-1}$. Substitute this into Eq. (7), adding and subtracting $\delta \frac{\delta}{1-\delta}y_t$, and rearrange to finally obtain:

$$S_t = \frac{1}{\delta}S_{t-1} - \frac{\delta}{1-\delta}\Delta y_t + \varepsilon_t, \quad (8)$$

where $\varepsilon_t = \frac{(1-\delta)}{\delta}u_t$.

As can readily be seen, Eq. (8) is a stochastic difference equation. When solved forward in expectations, with an added transversality condition, it generates Eq. (3). Notice that, by construction, $\mathbb{E}_t(\varepsilon_{t+i}) = 0$, $i = 1, 2, \dots$, which is the reason why we obtain exactly, Eq. (3). Eq. (8) will constrain the dynamics of the transformed VAR on $\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix}$, leading to a sequence of nested models starting with Eq. (4). The first

layer of the nesting structure appears when going from the VAR in Eq. (4) to the VECM in Eq. (5). To go from Eq. (5) to the transformed VAR representation we use $C = \begin{bmatrix} 1 & -\theta \\ 0 & 1 \end{bmatrix}$, the 2×2 nonsingular matrix formed by stacking the transpose of the cointegrating vector $\begin{bmatrix} 1 & -\theta \end{bmatrix}$ and the selection vector $\begin{bmatrix} 0 & 1 \end{bmatrix}$, such that $C \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \Delta S_t \\ \Delta y_t \end{pmatrix}$.

Premultiplying both sides of Eq. (5) by C , then solving for $\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix}$, we obtain:

$$\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \underline{\Gamma}_{11}(L) & \underline{\Gamma}_{12}(L) \\ \underline{\Gamma}_{21}(L) & \underline{\Gamma}_{22}(L) \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad (9)$$

where $\underline{\Gamma}_{11}(L)$ and $\underline{\Gamma}_{21}(L)$ are polynomials of order $p-1$ and $\underline{\Gamma}_{12}(L)$ and $\underline{\Gamma}_{22}(L)$ are polynomials of order $p-2$. Indeed, one important issue to note is that the transformed VAR in Eq. (9) is a VAR of order p in both S_t and Δy_t in which the two coefficients of Δy_{t-p} are zero. Cross-equation restrictions for the system are imposed on the coefficient matrices of $\underline{\Gamma}(L) = \begin{pmatrix} \underline{\Gamma}_{11}(L) & \underline{\Gamma}_{12}(L) \\ \underline{\Gamma}_{21}(L) & \underline{\Gamma}_{22}(L) \end{pmatrix} = \underline{\Gamma}_1 + \underline{\Gamma}_2 L + \dots + \underline{\Gamma}_p L^{p-1}$ in Eq. (9), as the following proposition asserts.

Proposition 2. (Campbell and Shiller, 1987, and Johansen and Swensen, 2011) A PVM as in $S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta} \Delta y_t + \varepsilon_t$, implies a polynomial serial-correlation common feature relationship (see Cubadda and Hecq, 2001) for the transformed VAR in Eq. (9): there exists a vector $\tilde{\gamma}'_0$ such that $\tilde{\gamma}'_0 \underline{\Gamma}_2 = \dots = \tilde{\gamma}'_0 \underline{\Gamma}_p = 0$, with $\tilde{\gamma}'_0 \underline{\Gamma}_1 = \tilde{\gamma}'_1 \neq 0$. Moreover in the PVM $\tilde{\gamma}'_0 = (1 : \frac{\delta}{1-\delta})$ and $\tilde{\gamma}'_1 = (-\frac{1}{\delta} : 0)$.

As was stressed by Campbell (1987) in the context of saving, Eq. (8) plays a very important role: it is the first order stochastic difference equation that generates the PVM. There are two important conditions to go from Eq. (8) to Eq. (3): cointegration delivers the transversality condition $\lim_{k \rightarrow \infty} \delta^k \mathbb{E}_t(S_{t+k}) = 0$, whereas the unforecastability of ε_t regarding the past, i.e., $\mathbb{E}_t(\varepsilon_{t+j}) = 0$, for all $j > 0$, ensures that there is no additional term on the right-hand side of Eq. (3) to invalidate it.

The first represents a long-run restriction between Y_t and y_t . The second restricts the dynamics of the stationary representation of the system, making S_t and Δy_t specific functions of their own past alone. Thus, they are short-run restrictions on the behavior of S_t and Δy_t – exactly the types of restrictions that are studied in the common-cycle literature. Therefore, applying the *toolkit* developed there provides a fresh view of PVMs.

Remark 1. Johansen and Swensen (2011) discuss the properties of the three spreads S_t , S'_t , and S''_t . Their setup is slightly different from ours, since they define the present-value relationship to be $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$ instead of $Y_t = \theta(1-\delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$; i.e., they discount only future values of y_t , not its current value. Some authors prefer the latter to the former, using the argument that, in the discrete time setup, the cash flow accrues at the end of every period. Here, we follow Campbell and Shiller in their choice of the PV formula. The cointegrating vector is not affected by this choice, but the short-run dynamic-coefficient restrictions are – as we shall see in the next section. This is why, some of our results differ from those of Johansen and Swensen (2011).

One possible explanation for a rejection of the PVMs is the use of cross-equation restrictions that impose both reduced-rank restrictions and particular values on the parameters. Misspecifications such as proxy variables or measurement errors can affect the value of the parameters, leaving the reduced-rank restrictions unaffected. As an example, instead of the PV representation $Y_t = \theta(1-\delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$, one can find in the literature that the series Y_t is a function of the future discounted expected value of y_t , such that $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$.² This slight change is not innocuous. Apply

²Johansen and Swensen (2011), and Campbell, Lo and Mackinlay (1996), use that formulation when they consider the stock price at the end of the period.

the algebra used before to $Y_t = \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t y_{t+i}$ to obtain the following expressions:

$$\Delta Y_t = -\Delta y_t + \frac{1-\delta}{\delta} S_{t-1} + u_t, \quad (10)$$

$$\text{where } \gamma' = (1 : 1) \text{ and } \alpha_1 = \frac{1-\delta}{\delta} \text{ in Proposition 1.} \quad (11)$$

$$S_t = -\frac{1}{(1-\delta)} \Delta y_t + \frac{1}{\delta} S_{t-1} + v_t, \quad (12)$$

$$\text{where } \tilde{\gamma}'_0 = (1 : \frac{1}{1-\delta}) \text{ and } \tilde{\gamma}'_1 = (-\frac{1}{\delta} : 0) \text{ in Proposition 2.} \quad (13)$$

Now, the unpredictable linear combinations involve three variables: ΔY_t , Δy_t , and S_t , in both the VECM and the transformed VAR. Also, the weights in the linear combinations have changed. Despite the differences in parameter values in the linear combinations above, the existence of a reduced-rank model is not affected by how one writes the PV equation linking Y_t and y_t . Hence, the reduced-rank properties of the VECM and of the transformed VAR are invariant to this choice: in both cases there exist weak-form common features for the VECM and the PSCCF for the transformed VAR.

2.2 Constant versus variable expected returns: levels versus logs

Campbell and Shiller's (1987) model for the levels of prices (Y_t) and dividends (y_t) is consistent with a very restrictive assumption – that the expected return of a given stock is constant over time:

$$\mathbb{E}_t [R_{t+1}] = R, \quad (14)$$

where

$$R_{t+1} \equiv \frac{P_{t+1} + D_{t+1}}{P_t} - 1, \quad (15)$$

where P_t and D_t denote the price and the dividend, respectively, of a given stock.

In two subsequent papers, Campbell and Shiller (1988a,b) developed an alternative representation for prices and dividends, in which Eq. (14) need not hold,

which is therefore consistent with the idea of time-varying returns. This alternative representation uses the logarithms of prices and dividends:

$$h_{t+1} \equiv \log(1 + R_{t+1}) = \log(P_{t+1} + D_{t+1}) - \log(P_t), \quad (16)$$

arriving at:

$$h_{t+1} = p_{t+1} - p_t + \log(1 + \exp(d_{t+1} - p_{t+1})), \quad (17)$$

where lower-case variables represent the respective logarithmic transformation of the original variable.

They use a first-order Taylor expansion in Eq. (17), to get:

$$h_{t+1} \approx k + \rho p_{t+1} + (1 - \rho)d_{t+1} - p_t \quad (18)$$

$$= k + \rho p_{t+1} + (1 - \rho)d_{t+1} - p_t + d_t - d_t \quad (19)$$

$$= k + \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} + (d_t - p_t), \quad (20)$$

where $\rho \equiv \frac{1}{(1 + \exp(\overline{d-p}))}$, $\overline{d-p}$ is the average across time of $d_{t+1} - p_{t+1}$, and $k \equiv -\log(\rho) - (1 - \rho)\log(1/\rho - 1)$.

Notice that we can solve Eq. (20) for $(d_t - p_t)$, yielding an exact stochastic first-order difference equation for it,

$$(d_t - p_t) = -k + h_{t+1} - \Delta d_{t+1} - \rho(p_{t+1} - d_{t+1}) + \varepsilon_{t+1}, \quad (21)$$

where ε_{t+1} is an approximation error. Under the assumption that $\mathbb{E}_t(\varepsilon_{t+j}) = 0$, for all $j > 0$, Eq. (21) can be solved forward to yield a logarithmic version of Eq. (3):

$$d_t - p_t = -\frac{k}{1 - \rho} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j [h_{t+1+j} - \Delta d_{t+1+j}]. \quad (22)$$

Campbell and Shiller (1988a) argue that “there is no economic content in equation (22)”. To get economic content, they impose a restriction on the behavior and dynamics of h_t :

$$\mathbb{E}_t h_{t+1} = \mathbb{E}_t r_{t+1} + c, \quad (23)$$

i.e., that the excess-return of a stock, vis-à-vis the real risk-free rate r_t , is constant.³ If r_t is observable, Eq. (22) and (23) yield a testable econometric model:

$$d_t - p_t = \frac{c - k}{1 - \rho} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j [r_{t+1+j} - \Delta d_{t+1+j}]. \quad (24)$$

2.3 Common-cyclical feature restrictions: the logarithmic version

To test the log-linear present value model embedded in Eq. (24), we use the tridimensional system for $X_t = (p_t, d_t, r_t)'$. Notice first that Eq. (23) implies that r_t is $I(0)$, given that h_t is $I(0)$. This yields the first cointegrating vector for the system in X_t . Given that r_t is $I(0)$, from Eq. (24) $d_t - p_t$ is $I(0)$ as well, yielding the second cointegrating vector in the system.

The VECM($p - 1$) reads as:

$$\begin{aligned} \begin{bmatrix} \Delta p_t \\ \Delta d_t \\ \Delta r_t \end{bmatrix} &= \begin{bmatrix} \gamma_1 & \gamma_4 \\ \gamma_2 & \gamma_5 \\ \gamma_3 & \gamma_6 \end{bmatrix} \begin{bmatrix} (d_{t-1} - p_{t-1}) \\ r_{t-1} \end{bmatrix} + \begin{pmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 \end{pmatrix} \Delta X_{t-1} + \\ &\quad \dots + \begin{pmatrix} a_{11}^{p-1} & a_{12}^{p-1} & a_{13}^{p-1} \\ a_{21}^{p-1} & a_{22}^{p-1} & a_{23}^{p-1} \\ a_{31}^{p-1} & a_{32}^{p-1} & a_{33}^{p-1} \end{pmatrix} \Delta X_{t-p-1} + \epsilon_t. \end{aligned} \quad (25)$$

Disregarding an irrelevant constant term, the PVM implies that $\mathbb{E}_{t-1} [h_t - r_t] = 0$. From Eq. (20), we can approximate h_t as $h_t = \rho \Delta p_t + (1 - \rho)(d_{t-1} - p_{t-1})$. Consequently, in the VECM, if we pre-multiply the system in Eq. (25) by $\begin{pmatrix} \rho & 0 & -1 \end{pmatrix}$, in order to obtain an expression for $h_t - r_t$, we must have:

$$\underbrace{\begin{pmatrix} \rho & 0 & -1 \end{pmatrix} \Delta X_t}_{\rho \Delta p_t - \Delta r_t} = -(1 - \rho)(d_{t-1} - p_{t-1}) + r_{t-1} + \begin{pmatrix} \rho & 0 & -1 \end{pmatrix} \epsilon_t, \quad (26)$$

³Equation (23) implies the existence of a common cycle for h_t and r_t .

which is equivalent to

$$\underbrace{\rho\Delta p_t + (1-\rho)(d_{t-1} - p_{t-1})}_{h_t} - r_t = \begin{pmatrix} \rho & 0 & -1 \end{pmatrix} \epsilon_t, \quad (27)$$

which implies $\mathbb{E}_{t-1}[h_t - r_t] = 0$. These theoretical conditions restrict the VECM parameters as follows:

$$\begin{pmatrix} \rho & 0 & -1 \end{pmatrix} \begin{pmatrix} a_{11}^i & a_{12}^i & a_{13}^i \\ a_{21}^i & a_{22}^i & a_{23}^i \\ a_{31}^i & a_{32}^i & a_{33}^i \end{pmatrix} = 0, \quad i = 1, \dots, p-1, \text{ and} \quad (28)$$

$$\begin{pmatrix} \rho & 0 & -1 \end{pmatrix} \begin{bmatrix} \gamma_1 & \gamma_4 \\ \gamma_2 & \gamma_5 \\ \gamma_3 & \gamma_6 \end{bmatrix} \begin{bmatrix} (d_{t-1} - p_{t-1}) \\ r_{t-1} \end{bmatrix} = \underbrace{(\rho\gamma_1 - \gamma_3)}_{-(1-\rho)}(d_{t-1} - p_{t-1}) + \underbrace{(\rho\gamma_4 - \gamma_6)}_1 r_{t-1}. \quad (29)$$

The last set of restrictions in Eq. (29) are:

$$\gamma_1 = \frac{\gamma_3 - (1-\rho)}{\rho}, \text{ and} \quad (30)$$

$$\gamma_4 = \frac{\gamma_6 + 1}{\rho}. \quad (31)$$

Given the estimates of γ_3 , γ_6 , and $\rho \equiv \frac{1}{(1+\exp(d-p))}$, we can obtain the corresponding values of γ_1 and γ_4 , consistent with Eq. (24). We now summarize these results with the following proposition.

Proposition 3. (Campbell and Shiller, 1988a) If the elements in $X_t = (p_t, d_t, r_t)'$ obey a PVM as in Eqs. (24) and (23), with the latter leading to $h_t = r_t + \varepsilon_t$, where $h_t = \rho\Delta p_t + (1-\rho)(d_{t-1} - p_{t-1})$, and $\mathbb{E}_{t-1}[\varepsilon_t] = 0$, then their VECM obeys a weak-form common feature relationship (Hecq et al., 2006, and Athanasopoulos et al., 2011):

there exists a 1×3 vector $\gamma' = \begin{pmatrix} \rho & 0 & -1 \end{pmatrix}$ such that $\gamma' \begin{pmatrix} a_{11}^i & a_{12}^i & a_{13}^i \\ a_{21}^i & a_{22}^i & a_{23}^i \\ a_{31}^i & a_{32}^i & a_{33}^i \end{pmatrix} = 0$,

$i = 1, \dots, p-1$, implying that the first row of $\begin{pmatrix} a_{11}^i & a_{12}^i & a_{13}^i \\ a_{21}^i & a_{22}^i & a_{23}^i \\ a_{31}^i & a_{32}^i & a_{33}^i \end{pmatrix}$ is proportional to the

last row ($1/\rho$); that $\gamma' \begin{bmatrix} \gamma_1 & \gamma_4 \\ \gamma_2 & \gamma_5 \\ \gamma_3 & \gamma_6 \end{bmatrix} \neq 0$, with the first-row elements being restricted as follows: $\gamma_1 = \frac{\gamma_3 - (1-\rho)}{\rho}$, and $\gamma_4 = \frac{\gamma_6 + 1}{\rho}$.

Finally, as was the case for the same series in levels, i.e., P_t and D_t , slight changes on the assumptions of when dividends are accrued (at the beginning, middle or end of period t) influence the short-run parameter restrictions imposed on the multivariate system for $X_t = (p_t, d_t, r_t)'$. This is not a trivial issue, especially because we are not dealing with firm data for prices and dividends. On the contrary, we are dealing with an aggregate of several firms, where each one has its own dividend policy and some of which are varying across time, making it very difficult to know exactly what are the appropriate short-run parameter restrictions that should be imposed in PVMs. Although short-run parameter restrictions are not invariant to changes in the dividend policy, the rank restrictions discussed in Propositions 1 and 3 are invariant to them. We take this into account explicitly when devising the forecast models considered in the forecast experiment.

3 In-sample analysis of the data used in the forecast experiment

Here, we analyze three different annual data sets which are used later in the forecast experiment. The first includes the levels of interest rates, with long and short maturities, labelled i_{lr} and i_{sr} , respectively. The second includes the levels of real price and the dividend for the S&P composite index, labelled P_t and D_t , and the last involves

a logarithmic transformation of prices and dividends, $p_t = \ln P_t$ and $d_t = \ln D_t$, with the inclusion of the real risk-free rate, labelled r_t . Data are collected for the period 1871–2012 by Robert Shiller at <http://www.econ.yale.edu/~shiller/data.htm>. The results are presented in Table 1, and the econometric tools used in this section are discussed at great length in the Appendix. Note that we investigate the robustness of the results to the presence of data for the recent financial crisis by considering two samples: 1871–2007 and 1871–2012. The existence of a cointegrating relationship (resp. a common cyclical feature relationship) is evaluated when the null of no cointegration is rejected (resp. of common features is not rejected).

For each data set, we use the Hannan-Quinn information criterion to determine the lag length of the VAR in levels.⁴ This is reported in the third column of Table 1. Columns (4), (5) and (6) refer to the cointegration analysis. We do not reject the null of no long-run relationships when levels are used, but the sensible value of the discount factor may lead to the conclusion that we have a power issue. There is a clear indication of cointegration for the interest rates as well as for prices/dividends in logs. Note that we should have found two cointegrating vectors for the system with $(p_t, d_t, r_t)'$. Strictly speaking, we only found one cointegrating vector at the 5% significance level, albeit the test statistic is very close to the critical value at that level. So, imposing two vectors could be a possibility. As far as common-cyclical features are concerned, the results in columns (7) to (9) show a clear indication of weak-form common features for interest rates; we also conclude that there is a common feature vector for the system $(P_t, D_t)'$ if we use the robust GMM test J_2 . For the system with $(p_t, d_t, r_t)'$, our optimal choice of model is the VECM(0), for which there always exist SCCFs.

Therefore, we were able to find empirically that data that are subject to the theory of PVMs conform to some of the restrictions implied by theory. In our forecast experiment below, we ask a different question: whether or not imposing these restrictions in unrestricted multivariate models for the same data leads to an improvement in standard forecast accuracy measures. This experiment has bearings not only on

⁴In a common-cycle context, the Monte-Carlo exercise by Vahid and Issler (2002) reached this conclusion, and it is also supported by evidence from Athanasopoulos et al. (2011).

Table 1: Long- and short-run properties of the series in the forecast experiment.

(1) System	(2) Ends	(3) VAR(p)	Cointegration		(6) $\hat{\theta}$	Common cycle		
			(4) H_0	(5) cv5%		(7) ξ_{LR}	(8) J_2	(9) cv5%
(P_t, D_t)	2012	3	$r = 0$: 10.47	15.49	60.75	9.76	5.65	7.81
	2007	2	$r = 0$: 12.67	15.49	69.76	10.46	1.69	3.84
(p_t, d_t, r_t)	2012	1	$r = 0$: 78.74	29.79	1.764	\exists SCCF	\exists SCCF	–
			$r = 1$: 14.58	15.49				
	2007	1	$r = 0$: 75.18	29.79	1.802	\exists SCCF	\exists SCCF	–
			$r = 1$: 14.37	15.49				
(i_{lr}, i_{sr})	2012	2	$r = 0$: 31.52	15.49	1.071	0.12	0.10	3.84
	2007	2	$r = 0$: 31.66	15.49	1.096	0.02	0.02	3.84

theory, but also on practical issues.

4 Out-of-sample forecasting

4.1 Forecasting strategies imposing different co-movement restrictions

Next describe the different forecasting strategies used in this paper. Each of them imposes a different set of co-movement restrictions on the unrestricted VAR, our benchmark forecasting model. We use the Hannan-Quinn (HQ) criterion to choose the lag length of estimated models; see the results of Vahid and Issler (2002) and Athanasopoulos et al. (2011) in a common-cycle context. All models are estimated by conditional maximum likelihood (ML),⁵ then we use the estimated results to forecast the variables in the system up to h periods ahead:

1. **VAR(p) in levels (benchmark):** Select p using the Hannan-Quinn (HQ) criterion.
2. **VECM(HQ-PIC):** This is the VECM possibly (but not necessarily) restricted

⁵Model (2) is estimated by two-stage conditional ML.

by cointegration and/or by weak-form serial-correlation common features: jointly select p , the rank of the short-run matrices, and the cointegrating rank, by a combination of the use of the posterior information criterion (PIC) and the HQ criterion, as suggested by Athanasopoulos et al. (2011).

3. **VECM(HQ-J):** This is the VECM using solely the PVM-cointegration restriction. Select p according to the HQ criterion, and impose the cointegrating-rank restriction consistent with the PVM. Conditional on that restriction, estimate the cointegrating vector consistently (Johansen, 1991). No short-run restrictions are imposed here.
4. **VECM(HQ)Rank:** This is the VECM using solely the PVM cointegration and weak-form serial-correlation-common-feature rank restrictions. No coefficient restrictions are imposed. Select p according to the HQ criterion, where we impose the rank restrictions for cointegration and weak-form serial-correlation-common-feature that are consistent with the PVM.
5. **PV model:** This is the VECM using the PVM cointegration and weak-form serial-correlation-common-feature rank restrictions, in addition to the theoretical restrictions discussed in Proposition 1. Select p according to the HQ criterion, imposing the restrictions outlined in Proposition 1, where we use the cointegrating vector estimate of θ , $\hat{\theta}$ (T -consistent), to constrain α_1 , as $\hat{\alpha}_1 = \frac{1-\hat{\delta}}{\hat{\delta}} = 1/\hat{\theta}$. We recover the estimated reduced-form parameters from the quasi-structural form, and forecast the variables in the system.
6. **PV model (logs):** This is the VECM for $X_t = (p_t, d_t, r_t)'$, using the PVM cointegration and weak-form serial-correlation-common-feature rank restrictions, in addition to the theoretical restrictions discussed in Proposition 3. Select p according to the HQ criterion. Estimate the last two reduced-form equations under the previous restrictions and set $\hat{\rho} \equiv \frac{1}{(1+\exp(\overline{d-p}))}$, where $\overline{d-p}$ is the time-average of $d_t - p_t$. With the last two equations of reduced-form estimates and $\hat{\rho}$, we assemble the quasi-structural form, recover the reduced-form,

and forecast the variables in the trivariate system for $X_t = (p_t, d_t, r_t)'$.⁶

In the list above, we compare models with different layers of restrictions, starting with the unrestricted model in (1), all the way through to the PVM-restricted models in (5) and (6), in levels and in log-levels, respectively. There is a rationale for choosing each of these different models. Model (2) is the preferred choice of Athanasopoulos et al. (2011), when the series being modelled are subject to long- and short-run restrictions. This choice is data-driven, since we let information criteria (PIC and HQ) choose the cointegrating and short-run-matrix ranks, rather than their being imposed *a priori*. Models (3) and (4) impose (i.e., without testing for the existence of those restrictions) cointegration and weak-form serial-correlation-common-feature rank restrictions, respectively. These restrictions are motivated by the theory of PV models for trending data. First, notice that model (3) refrains from imposing short-run restrictions on the VECM coefficients, consistent with Propositions 1 and 3. Second, notice that model (4) refrains from imposing the parameter restrictions listed in Propositions 1 and 3, but only imposes the rank restrictions. This is due to the fact that parameter restrictions are sensitive to small changes in the assumptions underlying PV models. As was discussed above, one of these assumptions is related to the timing in which dividends are accrued in the price-dividend model; see Remark 1 and the discussion at the end of Section 2.2.

Finally, models (5) and (6) are completely theory-based and impose all restrictions listed in either Proposition 1 or Proposition 3. Comparisons of models (5) and (6) with the unrestricted VAR and with model (2) can determine whether imposing *structural* restrictions helps in forecasting, settling – at least from the point of view of forecasting using PVM restrictions – the dispute between theory-based econometrics (structural form) and *atheoretical* econometric models (reduced form).⁷ It is interesting to note that, for the dividend-price series, there is a long history of literature in finance and forecasting about the issue of return predictability; see for example Goyal

⁶Loss functions here are computed vis-à-vis the logged variables, i.e., vis-à-vis the variables in $X_t = (p_t, d_t, r_t)'$.

⁷Regarding the dispute between reduced- and structural-form models in econometrics, see the recent paper by Keane (2010) and further comments on it.

and Welch (2003, 2008), Campbell and Thomson (2008), and Cochrane (2008). For PVMs, the cointegrating relationship is a natural predictor of the variables in the dynamic system. Our propositions stress the role of S_{t-1} as a predictor in the dynamic systems – restricted VECMs and transformed VARs, which holds for $d_{t-1} - p_{t-1}$, as well as in the log version. Models (5) and (6) respectively have S_{t-1} and $d_{t-1} - p_{t-1}$ exclusively as predictor, being similar in spirit to the exercises implemented in this body of literature.⁸

The estimation details for the forecasting models are as follows. First, we divide our full sample into a “estimation sample” and a “forecasting sample.” Since the *great recession* (2008–2009) has had a huge influence on both asset prices and the prices of bonds (interest rates), we consider two separate forecasting samples: the first ending in 2007, just prior to the great recession, and the second ending in 2012, with all available information to date. We set $h = 12$ years in the forecasting exercise. This enables the measurement of the short- and medium-term forecast accuracy (h between 1 and 5 years), as well as the long-term ($h > 10$ years). When forecasting until 2007, we have 70 observations for the estimation sample (1871–1940) and 67 observations for the forecasting sample (1941–2007).⁹ When forecasting until 2012, the estimation sample has 75 observations (from 1871–1945) and the forecasting sample has 67 (1946–2012). Estimation is performed with a rolling window, kept constant throughout the out-of-sample exercise.

The forecast accuracies of all restricted models are compared to that of the VAR in levels. We use the ratio of the root mean squared forecast error for each model (or variable in them) to that of the VAR in levels – our benchmark:

$$RRMSFE_h^M = \frac{RMSFE_h^M}{RMSFE_h^{VAR}}, \quad (32)$$

where $RRMSFE_h^M$ is the root mean squared forecast error (RMSFE) statistic of model M , relative to that of the unrestricted VAR, for h -step-ahead forecasting. All

⁸We thank an anonymous referee for pointing this out.

⁹Notice that, the number of out-of-sample observations differ for $h = 1, 2, \dots, 12$: it is 67 for the first but 56 for the last.

comparisons are made using the embedded first-difference forecast errors.

We want to be able to distinguish the forecast accuracies of models (1) through (6), asking whether their accuracy measures are statistically equal or not. We do this using the unconditional predictive ability test of Giacomini and White (2006), i.e. the rolling window version of the Diebold and Mariano (1995) test, comparing each model M with the unrestricted VAR, and with all other models as well though these results are not shown here to save space, but are available upon request.

4.2 Forecasting results

We have forecast results for the three different data sets described earlier. We computed the relative measure of the forecast accuracy (RMSFE) of model M vis-à-vis that of the VAR, described in Eq. (32). Since all of the restricted representations (models (2) (6)) forecast the first differences of the data, but the VAR (model (1)) forecast their level, we transform the VAR forecasts errors into first-difference errors in order to compute the ratio in Eq. (32).

Following the empirical financial literature, which relies much more on results for individual data, we focus on forecast measures for the individual target variables in the PV formulas, i.e., for the Y s, instead of forecast measures for the system as a whole (e.g., the determinant of the RMSFE matrix). However, we must stress that the forecasting exercise performed is multivariate, since we use the system as a whole to produce forecasts for individual series. With multivariate forecasts in hand, we then focus only on forecasts of asset prices (P), long-run interest rates (i_{lr}), and logged asset prices (p). This limitation is justified by the finance literature, since, for example, most forecast studies are interested in asset prices and there is very little interest in forecasting dividends, even if forecasting dividends is an intermediate step for forecasting prices.

In Tables 2 and 3, we present forecast results for i_{lr} . When we exclude the great recession period – forecasts up to 2007, strategy (4), VECM(HQ)Rank dominates at all horizons, although strategy (3), VECM(HQ-J) and the PV model (5) perform well on occasion. If we extend the forecast period to 2012, strategy (4) dominates again,

Table 2: Relative RMSFEs of restricted models vs VAR for i_{lr} . Forecast period up to 2007.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.944	1.06	1.064	1.174	0.993	1.111	0.988	0.914	0.927	0.905	1.022	0.872
	[0.176]	[0.254]	[0.224]	[1.549]	[0.011]	[1.564]	[0.012]	[0.564]	[0.509]	[0.668]	[0.195]	[1.831]
VECM(HQ-J)	0.943	0.939	0.996	0.981	0.905	0.913	0.874	0.944	0.889	0.845	0.874	0.864
	[0.699]	[1.221]	[0.015]	[0.226]	[1.693]	[2.021]	[1.422]	[1.242]	[1.411]	[1.495]	[1.03]	[1.856]
VECM(HQ)Rank	0.915	0.806	0.996	0.924	0.923	0.938	0.859	0.892	0.899	0.831	0.881	0.862
	[0.492]	[1.632]	[0.000]	[1.292]	[2.373]	[1.667]	[1.464]	[1.108]	[1.333]	[1.536]	[1.015]	[1.837]
PV	0.979	0.938	0.963	0.933	0.979	0.933	0.986	0.967	0.881	0.855	0.979	0.809
	[0.15]	[1.051]	[0.587]	[0.715]	[0.161]	[1.807]	[0.039]	[0.237]	[1.045]	[0.936]	[0.067]	[1.733]

Ratio of the root mean squared forecast error (RMSFE) of the model in each row to that of the VAR in levels, transformed to first differences.

Best models in blue.

*Denotes rejection of the null of equal forecast accuracy at the 10% level, according to the unconditional predictive ability test of Giacomini and White (2006). The number in square brackets is the test statistic and the critical value is equal to 2.705.

**Denotes rejection of the null of equal forecast accuracy at the 5% level, according to the unconditional predictive ability test of Giacomini and White (2006). The number in square brackets is the test statistic and the critical value is equal to 3.842.

***Denotes rejection of the null of equal forecast accuracy at the 1% level, according to the unconditional predictive ability test of Giacomini and White (2006). The number in square brackets is the test statistic and the critical value is equal to 6.349.

Table 3: Relative RMSFEs of restricted models vs VAR for i_{lr} . Forecast period up to 2012.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	1.105	1.021	1.068	1.118	0.961	1.098	1.234	0.948	0.978	1.05	0.959	0.886
	[0.913]	[0.024]	[0.326]	[1.713]	[0.427]	[1.398]	[1.536]	[0.316]	[0.098]	[0.377]	[0.328]	[2.175]
VECM(HQ-J)	0.971	0.958	1.026**	0.985	0.973	0.958	0.972*	0.984	0.958	0.944	0.928	0.909
	[2.144]	[1.472]	[4.44]	[0.238]	[1.74]	[2.646]	[3.089]	[0.943]	[1.117]	[1.351]	[0.706]	[1.999]
VECM(HQ)Rank	0.927	0.823	1.011	0.95	0.957	0.948	0.94	0.937	0.922	0.918	0.916	0.888
	[0.216]	[1.468]	[0.292]	[1.03]	[2.219]	[2.006]	[2.561]	[0.944]	[1.233]	[1.445]	[0.765]	[1.876]
PV	0.937	0.881	0.965	0.939	0.982	0.97	0.998	0.959	0.899	0.882	0.871	0.849
	[0.512]	[1.616]	[0.611]	[0.716]	[0.003]	[0.004]	[0.315]	[0.255]	[0.845]	[0.892]	[0.752]	[1.853]

See the notes to Table 2.

although the PV model (5) performs well in the medium- to long-term. Thus, for i_{lr} forecasts, we conclude that strategy (4) performs really well, followed by the PV model (5). Despite that, no strategy has a forecast performance that is statistically superior to that of the VAR when the unconditional Giacomini and White (2006) test is employed.¹⁰

In Tables 4 and 5, we present forecast results for prices of the S&P 500 portfolio, P_t . The PV model (5) performed really well regardless of the forecast sample (2007

¹⁰The rest is well known, in that it is developed in the typical Diebold-Mariano-West framework. One could also easily adopt the more standard Diebold-Mariano (DM) test, which is less powerful and undersized. As Buseti and Marcucci (2013) showed, if one can reject the null of equal forecast accuracy with the simple DM test, there is no need to try more complicated tests based on bootstrapped or simulated critical values.

Table 4: Relative RMSFEs of restricted models vs VAR for P_t . Forecast period up to 2007.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.951 [0.281]	0.951 [2.461]	0.954* [3.031]	0.959 [2.144]	0.958 [2.005]	0.971 [1.386]	0.99 [1.323]	0.995 [1.363]	0.987 [1.737]	0.994 [0.515]	0.998 [0.223]	0.993 [1.828]
VECM(HQ-J)	0.945 [0.185]	0.954* [3.743]	0.962* [3.483]	0.965 [2.621]	0.96 [2.35]	0.971 [1.924]	0.987 [1.368]	0.991 [1.467]	0.984** [6.389]	0.992 [0.706]	0.997 [0.302]	0.993 [1.799]
VECM(HQ)Rank	0.95 [0.252]	0.955 [1.487]	0.958 [1.394]	0.962* [2.778]	0.958 [2.369]	0.97 [0.025]	0.989 [0.091]	0.994 [0.771]	0.985** [4.737]	0.992 [0.791]	0.997 [0.353]	0.992 [1.847]
PV	0.965 [0.189]	0.968 [1.371]	0.983 [0.904]	0.985 [0.829]	0.973 [1.356]	0.982 [1.192]	0.991 [1.134]	0.98 [1.066]	0.974* [2.766]	0.983 [0.822]	0.989 [0.737]	0.993 [0.302]

See the notes to Table 2.

Table 5: Relative RMSFEs of restricted models vs VAR for P_t . Forecast period up to 2012.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.998 [0.941]	0.967** [4.334]	0.938 [2.145]	0.926 [1.713]	0.898 [1.629]	0.949 [1.769]	0.926 [1.216]	0.918 [1.31]	0.924 [1.251]	0.921 [1.217]	0.888 [1.188]	0.905 [1.106]
VECM(HQ-J)	0.991 [1.343]	0.97** [4.962]	0.948 [2.321]	0.935 [1.883]	0.907 [1.737]	0.944 [1.295]	0.926 [1.288]	0.917 [1.326]	0.923 [1.268]	0.921 [1.223]	0.89 [1.186]	0.904 [1.109]
VECM(HQ)Rank	0.991 [0.544]	0.965 [2.667]	0.936 [1.962]	0.925 [1.91]	0.893 [1.665]	0.941 [0.822]	0.919 [1.354]	0.909 [1.322]	0.915 [1.242]	0.914 [1.217]	0.879 [1.185]	0.895 [1.107]
PV	0.98 [0.605]	0.908 [1.333]	0.828 [0.894]	0.777 [0.713]	0.727 [1.277]	0.652 [1.519]	0.564 [1.17]	0.478 [1.197]	0.443 [1.215]	0.374 [1.175]	0.315 [1.155]	0.264 [1.106]

See the notes to Table 2.

or 2012), although its performance in the short term for the sample up to 2007 was beaten by that of strategy (2), VECM(HQ-PIC) and strategy (4). Thus, for P_t , we conclude that the PV model (5) is the best, followed by strategies (2) and (4).

In Tables 6 and 7, we present forecast results for p_t . Regardless of the sample period, we find that the best model for p_t is strategy (3), whereas, on occasion (middle to long horizons), strategy (4), performed well. Most of the time, strategy (3), forecasts are statistically different from the VAR forecasts using the unconditional Giacomini-White test.

Wrapping up the results across all target assets and horizons (up to 12 years), we find that the PV model (5) performs well. This is in line with well-known results on return predictability for the price-dividend spread and (logged) price-dividend ratio; see Goyal and Welch (2003, 2008), Campbell and Thomson (2008) and Cochrane (2008). The same can be said about strategy (4), VECM(HQ)Rank, and, to a lesser extent, strategy (3), VECM(HQ-J), which only performed really well in the

Table 6: Relative RMSFEs of restricted models vs VAR for p_t . Forecast period up to 2007.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.929** [3.985]	0.938** [4.604]	0.934** [6.117]	0.947* [3.034]	0.958 [1.007]	0.965 [0.803]	0.974 [0.719]	0.995 [0.211]	0.993 [0.149]	0.995 [0.118]	0.999 [0.018]	0.997 [0.019]
VECM(HQ-J)	0.897** [6.36]	0.921*** [6.689]	0.923*** [8.232]	0.936** [5.103]	0.935* [3.233]	0.958 [1.271]	0.97 [0.81]	0.99 [0.366]	0.981 [0.641]	0.987 [0.397]	0.994 [0.126]	0.994 [0.101]
VECM(HQ)Rank	0.901** [4.958]	0.931** [4.593]	0.929** [5.979]	0.942* [3.393]	0.935* [3.102]	0.959 [1.214]	0.972 [0.812]	0.991 [0.357]	0.979 [0.668]	0.984 [0.487]	0.992 [0.171]	0.995 [0.051]
PV (logs)	2.025*** [42.408]	2.917*** [79.156]	3.119*** [67.565]	3.178*** [58.132]	3.176*** [48.619]	3.286*** [43.036]	3.407*** [37.053]	3.352*** [32.576]	3.325*** [28.27]	3.328*** [25.879]	3.335*** [23.389]	3.345*** [21.084]

See the notes to Table 2.

Table 7: Relative RMSFEs of restricted models vs VAR for p_t . Forecast period up to 2012.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.868* [3.641]	0.909 [2.255]	0.945 [0.932]	0.962 [1.489]	0.969 [0.803]	0.984 [0.1]	0.981 [0.091]	0.988 [0.103]	1.002 [0.054]	1.001 [0.003]	1.000 [0.002]	1.003 [0.025]
VECM(HQ-J)	0.876** [5.729]	0.901* [3.358]	0.924** [4.425]	0.943* [2.949]	0.958 [1.599]	0.975 [0.42]	0.976 [0.322]	0.984 [0.203]	0.989 [0.553]	0.995 [0.09]	0.999 [0.001]	1.001 [0.001]
VECM(HQ)Rank	0.879** [6.201]	0.908* [2.751]	0.936 [2.425]	0.949 [1.906]	0.957 [1.369]	0.976 [0.364]	0.98 [0.145]	0.986 [0.108]	0.991 [0.449]	0.998 [0.032]	1.001 [0.004]	1.000 [0.000]
PV (logs)	2.001*** [32.221]	2.834*** [81.764]	3.044*** [56.782]	3.104*** [48.517]	3.14*** [38.35]	3.204*** [36.047]	3.212*** [34.488]	3.19*** [33.279]	2.948*** [32.957]	3.027*** [33.91]	3.037*** [32.3]	3.051*** [37.36]

See the notes to Table 2.

logarithmic case for p_t . For financial series, this highlights the importance of building forecasting models which consider strategy (2), include strategy (4), or impose PV model (5) short-run restrictions. We stress this because, of all strategies, the only strategy that rules out an important role for short-run restrictions is strategy (3). Although it did really well for the logarithmic case (p_t), strategies that allow a role for short-run restrictions do much better for P_t and i_{lr} , which is in line with the findings of the common-feature literature.

As is well known, econometric models are built mainly for forecasting at shorter horizons – as the horizon increases, conditional forecasts converge to their unconditional counterparts. This highlights the importance of focusing only on results at business-cycle horizons (1 to 5 years ahead). In this case, there are two winners – strategy (3), and strategy (4), tied as the best strategy 30% of the time each. These are followed closely by PV model (5), which is the best 26.67% of the time, and strategy (2), which is the best 13.33% of the time. Again, we must stress that strategies that allow a role for short-run restrictions produced winners in 70% of cases at

business-cycle horizons.

All in all, if we had to recommend a single forecast strategy for multivariate models containing series that are subject to PVM restrictions, we would recommend either strategy (3), VECM(HQ-J) or strategy (4), VECM(HQ)Rank.

5 Conclusion

This paper makes two original contributions. The first is to show that PV relationships entail a weak-form SCCF restriction, as per Hecq et al. (2006) and Athanasopoulos et al. (2011). It also implies a polynomial serial correlation common feature relationship (Cubadda & Hecq, 2001) for the VAR representation for Δy_t and the cointegrating relationship $Y_t - \theta y_t$. These represent short-run restrictions on the dynamic system for these variables, something that has not been discussed before in this literature.

Our second contribution relates to the forecasting of multivariate time series that are subject to PVM restrictions, which has a wide application in macroeconomics and finance. We benefit from previous work showing the benefits for forecasting when the short-run dynamics of the system are constrained for stationary data (Vahid & Issler, 2002), and when they are constrained for data that are subject to long- and short-run restrictions (Issler & Vahid, 2001, Anderson & Vahid, 2011, and Athanasopoulos et al., 2011). The reason why appropriate common-cycle restrictions improve forecasting is because they find linear combinations of the first differences of the data that cannot be forecast by past information. This embeds natural exclusion restrictions, preventing the estimation of useless parameters, which would otherwise contribute to increase the forecast variance with no expected reduction in squared bias.

We applied the techniques discussed in this paper to data that are known to be subject to PV restrictions: the online series maintained and updated by Robert Shiller at <http://www.econ.yale.edu/~shiller/data.htm>. We focus on three different data sets. The first includes the levels of interest rates with long and short maturities, the second includes the levels of real prices and dividends for the S&P composite index, and the third includes the logarithmic transformation of prices and dividends,

for which PV analysis requires the inclusion of the real risk-free rate. Our exhaustive investigation of six different multivariate models reveals that better forecasts can be achieved when restrictions are applied to them. The strategies that performed best were, strategy (3), VECM(HQ-J) and strategy (4), VECM(HQ)Rank, followed closely by the PV model (5).

All in all, if we had to recommend one forecast strategy for multivariate models containing series subject to PVM restrictions, we would recommend either strategy (3), or strategy (4).

Finally, when taken as a whole, our results show the usefulness of imposing short-run restrictions for forecasting economic series. It should be stressed that strategies that allow a role for short-run restrictions performed best 70% of the time at business cycle horizons – (4), VECM(HQ)Rank, PV model (5), and (2), VECM(HQ-PIC). This is in line with the findings of more than 20 years of research in multivariate analysis using short-run restrictions.

6 Acknowledgements

This paper supersedes that of Hecq and Issler (2012), with two additional coauthors. We gratefully acknowledge the comments and suggestions given by Ricardo Brito, Rob Hyndman, Søren Johansen, Farshid Vahid, three anonymous referees, participants in the 2011 SBE conference in Foz do Iguaçu, and participants in the 2012 IIF Conference in Boston. All remaining errors are ours. Issler thanks CNPq, Faperj and INCT for financial support. Guillén and Hecq thank INCT for financial support. We thank Rafael Burjack, Marcia Waleria Machado, Andréa Machado and Marcia Marcos for excellent research assistance.

References

- [1] ANDERSON, H.M., & VAHID, F. (2011). VARs, cointegration and common cycle restrictions. in *The Oxford Handbook of Economic Forecasting*. Edited by Michael P. Clements & David F. Hendry.

- [2] ATHANASOPOULOS, G., GUILLEN, O.T.C., ISSLER, J.V., & VAHID, F. (2011). Model selection, estimation and forecasting in VAR models with short-run and long-run restrictions. *Journal of Econometrics*, 164(1), 116-129.
- [3] BAILLIE, R. (1989). Econometric tests of rationality and market efficiency. *Econometric Reviews*, 8, 151-186.
- [4] BASISTHA, A., & STARTZ, R. (2008). Measuring the NAIRU with reduced uncertainty: a multiple-indicator common-cycle approach. *The Review of Economics and Statistics*, 90(4), 805-811.
- [30] BUSETTI, F., & MARCUCCI, J. (2013). Comparing forecast accuracy: A monte carlo investigation. *International Journal of Forecasting*, 29(1), 13-27.
- [6] CAMPBELL, J.Y. (1987). Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis. *Econometrica*, 55(6), 1249-73.
- [7] CAMPBELL, J.Y., & SHILLER, R. (1987). Cointegration tests and present value models. *Journal of Political Economy*, 95, 1062-1088.
- [8] CAMPBELL, J.Y., & SHILLER, R. (1988A). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 58, 495-514.
- [9] CAMPBELL, J.Y., & SHILLER, R. (1988B). Stock prices, earnings and expected dividends. *Journal of Finance*, 43, 661-676.
- [10] CAMPBELL, J.Y., & DEATON, A. (1989). Why is consumption so smooth?. *The Review of Economic Studies*, 56, 357-374.
- [11] CAMPBELL, J.Y., LO, A.W., & MACKINLAY, A.C. (1996). *The Econometrics of Financial Markets*. Princeton University Press.

- [30] CAMPBELL, J. Y., & THOMPSON S.B. (2008). Predicting excess stock returns out of sample: can anything beat the historical average?. *The Review of Financial Studies*, 21(4), 1509-1531.
- [13] CANDELON, B., & HECQ, A. (2000). Stability of Okun's law in a codependent system. *Applied Economic Letters*, 7, 687-693.
- [14] CENTONI, M., CUBADDA, G., & HECQ, A. (2007). Common shocks, common dynamics and the international business cycle. *Economic Modelling*, 24, 1, 149-166.
- [15] CHOW, G., (1984). *Econometrics*. McGraw-Hill, New York.
- [16] Cochrane, John H. (2008). The dog that did not bark: a defense of return predictability. *The Review of Financial Studies*, 21(4), 1533-1575.
- [17] CUBADDA, G. AN, & HECQ, A. (2001). On non-contemporaneous short-run comovements. *Economics Letters*, 73, 389-397.
- [18] ENGLE, R.F., & GRANGER, C.W.J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica*, 55, 251-276.
- [19] ENGLE, R.F., & KOZICKI, S. (1993). Testing for common features (with comments). *Journal of Business and Economic Statistics*, 11, 369-395.
- [20] ENGLE, R.F., & ISSLER, J.V. (1995). Estimating common sectoral cycles. *Journal of Monetary Economics*, 35, 83-113.
- [21] ENGLE, R.F., & YOO, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35(1), 143-159.
- [22] ENGSTED, T. (2002). Measures of fit for rational expectations models. *Journal of Economic Surveys* 16, 301-355.
- [23] GIACOMINI, R., & WHITE, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545-1578.

- [24] GOYAL, AMIT, & WELCH, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5), 639–654.
- [25] GOYAL, AMIT, & WELCH, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.
- [26] HAMILTON, J.D. (1994). *Time Series Analysis*. (Princeton University Press: Princeton).
- [27] HANSEN, L.P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica*, 50, 1029-1054.
- [28] HANSEN, L.P., & SARGENT, T.J. (1981). A note on Wiener-Kolmogorov prediction formulas for rational expectations models. *Economics Letters*, 8(3), 255-260.
- [29] HANSEN, L.P., & SARGENT, T.J. (1993). Recursive linear models of dynamic economies. *Proceedings, Federal Reserve Bank of San Francisco*, March issue.
- [30] HECQ, A., & ISSLER, J.V., (2012). A common-feature approach for testing present-value restrictions with financial data. economics working papers (Ensaaios Economicos da EPGE) 728, FGV/EPGE.
- [31] HECQ, A., PALM, F., & URBAIN, J.P. (2000). Permanent-transitory decomposition in VAR models with cointegration and common cycles. *Oxford Bulletin of Economics & Statistics*, 62, 511-532.
- [32] HECQ, A., PALM, F.C., & URBAIN, J.P. (2006). Testing for common cyclical features in VAR models with cointegration. *Journal of Econometrics*, 132(1), 117-141.
- [33] HECQ, A., LAURENT, S., & PALM, F.C. (2011). Pure portfolios models. RM Maastricht University.

- [34] HOFFMAN D.L, & RASCHE, R.H. (1996). Assessing forecast performance in a cointegrated system. *Journal of Applied Econometrics*, 11, 495–517.
- [35] ISSLER, J.V., & VAHID, F. (2001). Common cycles and the importance of transitory shocks to macroeconomic aggregates. *Journal of Monetary Economics*, 47(3), 449–475.
- [36] JOHANSEN, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.
- [37] JOHANSEN, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press: Oxford.
- [38] JOHANSEN, S. (2000). Modelling of cointegration in the vector autoregressive model. *Economic Modelling*, 17, 359–373.
- [39] JOHANSEN, S., & SWENSEN, A. (1999). Testing rational expectations in vector autoregressive models. *Journal of Econometrics*, 93, 73–91.
- [40] JOHANSEN, S., & SWENSEN, A. (2004). More on testing exact rational expectations in cointegrated vector autoregressive models: restricted drift terms. *Econometrics Journal*, 7, 389–397.
- [41] JOHANSEN, S., & SWENSEN, A. (2011). On a numerical and graphical technique for evaluating some models involving rational expectations. *Journal of Time Series Econometrics*, 3, (1), 1–29.
- [42] KEANE, MICHAEL P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, 156 (1), 3–20.
- [43] PATTON, A., RAMADORAI, T. & STREATFIELD, M. (2013). Change you can believe in? Hedge fund data revisions. Working paper: Duke University.
- [44] REINSEL, G.C., & AHN, S.K. (1992). Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio tests, and forecasting. *Journal of Time Series Analysis*, 13, 353–375.

- [45] VAHID, F., & ENGLE, R.F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 8, 341-360.
- [46] VAHID, F., & ENGLE, R.F. (1997). Codependent cycles. *Journal of Econometrics*, 80, 199-221.
- [47] VAHID, F., & ISSLER, J.V. (2002). The importance of common cyclical features in VAR analysis: A monte carlo study. *Journal of Econometrics*, 109, 341-363.

A Appendix: Testing present-value models with a common-cycle approach

The discussion in this paper suggests that, for integrated Y_t and y_t , there are three different instances in which we can investigate the validity of PVMs. First, the cointegration test for Y_t and y_t , if both are $I(1)$. Second, the (invariant) rank restrictions in the VECM or the transformed VAR. Third, the coefficient restrictions and unpredictability properties for linear combinations. In order to test for PVMs, we propose the following steps:

1. Choose the order of the $VAR(p)$ for the joint $I(1)$ process $(Y_t, y_t)'$ consistently using different information criteria.
2. Given our choice of p , test for the existence of cointegration between Y_t and y_t . If such is the case (that is, there exists one cointegrating vector), estimate the long-run coefficient θ in $S_t = Y_t - \theta y_t$ super-consistently, using the likelihood-based trace test proposed by Johansen (1995). Alternatively, the Engle and Granger (1987) regression test can be carried out. In either case, form $\hat{S}_t = Y_t - \hat{\theta} y_t$. If there is no cointegration, the PVM is rejected.
3. Given p and \hat{S}_t , test for the weak form common feature using a reduced rank test for $(\Delta Y_t, \Delta y_t)'$. We can use both a likelihood ratio multivariate approach (e.g., a canonical correlation analysis) and a single-equation approach (e.g.,

GMM). Because most present-value relationships apply to heteroskedastic financial data, one may prefer a GMM framework on the basis that it embeds robust variance-covariance matrices for parameters estimates easily. Indeed, the canonical correlation approach assumes *i.i.d.* disturbances.

Note that we can improve on steps 1 to 3 using steps 4 and/or 5 below. Given that we only work with bivariate systems for relatively large numbers of observations in this paper we do not introduce such small sample improvements into our analysis. However, these are:

- 4 Integrate steps 2 and 3, estimating long-run and short-run parameters jointly as per Centoni, Cubadda and Hecq (2007).
- 5 Integrate steps 1, 2 and 3, estimating the lag length of the VAR and long-run and short-run parameters jointly as in Athanasopoulos *et al.* (2011).

A.1 LR tests for i.i.d. disturbances

The canonical-correlation approach entails the use of a likelihood ratio (reduced-rank regression) test for the weak-form common features in the $VECM(p-1)$ for $(\Delta Y_t, \Delta y_t)'$. It can be undertaken using the canonical-correlation test on zero eigenvalues, which are computed from:

$$CanCor \left\{ \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix}, \begin{pmatrix} \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix} \mid (D_t, \hat{S}_{t-1}) \right\}, \quad (33)$$

where $CanCor\{X_t, W_t|G_t\}$ denotes the computation of canonical correlations between the two sets of variables X_t and W_t , concentrating out the effect of G_t (deterministic terms and a disequilibrium error-correction term) by multivariate least

squares. The previous program (Eq. (33)) is numerically equivalent to

$$CanCor \left\{ \begin{pmatrix} \Delta Y_t \\ \Delta y_t \\ \hat{S}_{t-1} \end{pmatrix}, \begin{pmatrix} \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \hat{S}_{t-1} \end{pmatrix} \mid D_t \right\}, \quad (34)$$

which is more convenient for obtaining the coefficient of \hat{S}_{t-1} in Eq. (8) directly. The likelihood ratio test, denoted by ξ_{LR} , considers the null hypothesis that there exist at least s common feature vectors. It is obtained as $\xi_{LR} = -T \sum_{i=1}^s \ln(1 - \hat{\lambda}_i)$, $s = 1, 2$, where $\hat{\lambda}_i$ are the i th smallest squared canonical correlations computed from Eqs. (33) or (34) above, namely from $\hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XW} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{WX}$, or similarly from the symmetric matrix $\hat{\Sigma}_{XX}^{-1/2} \hat{\Sigma}_{XW} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{WX} \hat{\Sigma}_{XX}^{-1/2}$, where $\hat{\Sigma}_{ij}$ are the empirical covariance matrices, $i, j = X, W$.

In the bivariate case, the unrestricted VECM has $4(p-1)+2$ parameters, whereas the restricted model has $2(p-1)+2+1$. The number of restrictions when testing the hypothesis that there exists one WF common feature is then $2(p-1)-1 = 2p-3$ for $p > 1$.¹¹ As was proposed by Issler and Vahid (2001), we can obtain the same statistics by computing the difference between the log-likelihoods in the unrestricted $VECM(p-1)$ for $(\Delta Y_t, \Delta y_t)'$ and in the pseudo-structural form estimated by FIML:

$$\begin{pmatrix} 1 & -\gamma_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{\Gamma}_1 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots \\ + \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix} \begin{pmatrix} \Delta Y_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} (\alpha_1 - \gamma_0 \alpha_2) \\ \tilde{\alpha}_2 \end{pmatrix} S_{t-1} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

¹¹In the VECM, the general formula for n series that can be annihilated by s combinations is $sn(p-1) - s(n-s)$.

For the transformed VAR, the restriction underlying the restricted PSCCF can be tested using:

$$CanCor \left\{ \begin{pmatrix} \hat{S}_t \\ \Delta y_t \\ \hat{S}_{t-1} \end{pmatrix}, \begin{pmatrix} \hat{S}_{t-1} \\ \hat{S}_{t-2} \\ \vdots \\ \hat{S}_{t-p} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix} \mid D_t \right\},$$

where the number of parameters in the unrestricted model is $4(p-1)+2$; the restricted model has $6 + 2(p-2)$, with the number of restrictions being $2p-4$ in the case of an unrestricted $\tilde{\gamma}_1$:

$$\begin{pmatrix} 1 & -\tilde{\gamma}_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \tilde{\Gamma}_{1a} \\ \tilde{\Gamma}_{1b} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots \\ + \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix} \begin{pmatrix} S_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \tilde{\Gamma}_{p,p} & 0 \end{pmatrix} \begin{pmatrix} S_{t-p} \\ \Delta y_{t-p} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

If $\tilde{\gamma}_1$ is restricted, we have $2p-3$ restrictions, and the pseudo structural form is

$$\begin{pmatrix} 1 & -\tilde{\gamma}_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \vartheta_1 & 0 \\ \tilde{\Gamma}_{2,1} & \tilde{\Gamma}_{2,2} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \\ \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix} \begin{pmatrix} S_{t-2} \\ \Delta y_{t-2} \end{pmatrix} + \dots + \begin{pmatrix} 0 & 0 \\ \tilde{\Gamma}_{p,p} & 0 \end{pmatrix} \begin{pmatrix} S_{t-p} \\ \Delta y_{t-p} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Notice that this set of rank restrictions is identical to that of Campbell and Shiller (1987) if one imposes zero restrictions in the last matrix coefficient in their setup.¹² The approach for testing PVMs proposed here is to first test the rank condition (necessary), without imposing any further parameter restrictions yet. As was argued

¹²This is probably implicit in their analysis, but is not discussed in the paper itself.

above, the rank condition is invariant to the way in which we write the PV equation linking Y_t and y_t . If it is not rejected, then we can test the additional restrictions on matrix coefficients, which are not invariant to the way we write the PV equation. Putting more weight on invariant restrictions satisfies robustness, since, a different definition of the timing of Y_t and/or y_t , along with the presence of measurement error and data revisions, will all lead to the correct rank condition to be met, but will imply different parameter values in the difference equation generating PVMs. An additional reason to follow this path is that we will be able to split the two effects, shedding light on the exact reason for a rejection of the theory, if that is the case. It is important to understand why we reject a given PVM, since various authors have complained that cross-equation restriction tests reject PVMs too often, even in cases where theory is firmly believed to hold and a graphical analysis seems to support that view.

A.2 Regression-Based GMM Tests

Testing using a GMM approach entails testing the common feature null hypothesis using an orthogonality condition between a combination of variables in the model $(\Delta Y_t, \Delta y_t, \hat{S}_{t-1})'$ and the conditioning set W_t' . For example, in the context of Eq. (7), we would have the following moment restrictions:

$$\mathbb{E}([\Delta Y_t - \gamma_1 \Delta y_t - \gamma_2 \hat{S}_{t-1}] \otimes W_t') = 0, \quad (35)$$

where we would also have to test $H_0: \gamma_1 = 0$ and $\gamma_2 = \frac{1-\delta}{\delta}$ using a Wald test. Prior to that, we want to estimate γ_1 and γ_2 and test the validity of the over-identifying restrictions in Eq. (35). The use of IV- type estimators and the associated orthogonality tests is straightforward in this context. Let us consider W_t , the vector of instruments, defined as before (an intercept is added). The GIVE estimator is simply the 2SLS or the IV estimator when the instruments are the past of the series, namely

$$\hat{\theta}_{GIVE} = (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{Y}), \quad (36)$$

with $\Delta \mathbf{X}_t = (\Delta y_t, \hat{S}_{t-1}, 1)'$. The validity of the orthogonality condition, and consequently, the presence of a common feature vector, is obtained via an overidentification J-test (Hansen, 1982) $J_1(\theta) = T g_T(\theta; \cdot)' P_T^{-1} g_T(\theta; \cdot)$, the empirical counterpart of which is:

$$J_1(\theta_{IV}) = (\mathbf{u}' \tilde{\mathbf{W}})(\hat{\sigma}_u^2 \tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1}(\tilde{\mathbf{W}}' \mathbf{u}).$$

Under usual regularity properties, the variance-covariance matrix of the orthogonality condition has the sample counterpart $\hat{P}_T = (1/T) \hat{\sigma}_u^2 (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})$, with $u_t = \Delta Y_t - \hat{\gamma}_1 \Delta y_t - \hat{\gamma}_2 \hat{S}_{t-1}$. $\tilde{\mathbf{W}}$ is the demeaned \mathbf{W} , namely $\tilde{\mathbf{W}} = \mathbf{W} - \mathbf{i}(\mathbf{i}' \mathbf{i})^{-1} \mathbf{i}' \mathbf{W}$ (with $\mathbf{i} = (1 \dots 1)'$), because we do not want to impose the restriction that the common feature vector also annihilates the constant vector.

All of the estimates and tests presented above embed the assumption of homoskedasticity. This may be fine for macroeconomic data, such as consumption and income, but is clearly not appropriate for financial data. In a Monte Carlo exercise, Candelon *et al.* (2005) have illustrated that ξ_{LR} has large size distortions in the presence of GARCH disturbances. We implement the GIVE estimator by using White's HCSE estimator such that (see Hamilton, 1994):

$$\hat{\theta}_{GMM} = (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{B} \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{B} \mathbf{W})^{-1} \mathbf{W}' \Delta \mathbf{Y}), \quad (37)$$

where the only difference between $\hat{\theta}_{GMM}$ and the usual $\hat{\theta}_{GIVE}$ is the presence of an additional diagonal matrix $\mathbf{B} = \mathbf{diag}(u_1^2, u_2^2, \dots, u_T^2)$, where $u_t = \Delta Y_t - \hat{\gamma}_{1IV} \Delta y_t - \hat{\gamma}_{2IV} \hat{S}_{t-1}$ are the residuals obtained under homoskedasticity using the GIVE estimation in a first step. For testing, we form the following new sequence of residuals $u_t^* = \Delta Y_t - \hat{\gamma}_{1GMM} \Delta y_t - \hat{\gamma}_{2GMM} \hat{S}_{t-1}$, and use these to compute a new J-test that is robust to heteroskedasticity:

$$J_2(\theta_{GMM}) = (\mathbf{u}^{*'} \tilde{\mathbf{W}})(\tilde{\mathbf{W}}' \mathbf{B} \tilde{\mathbf{W}})^{-1}(\tilde{\mathbf{W}}' \mathbf{u}^*). \quad (38)$$

Note that alternative approaches have been evaluated by Hecq and Issler (2012).

A.3 Small sample properties of PVM tests

A small Monte Carlo simulation might help to determine which of the tests considered to use. We use $T = 100$ observations with 10,000 replications. The lag length of the VAR in levels in the data generating process is chosen to be $p = 3$. However, we estimate the model for $p = 2, 3$ and 5. The DGP that ensures $\gamma' = (1 : 0)$ is:

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0.5 & 0.2 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots \begin{pmatrix} 0 & 0 \\ -0.4 & 0.2 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-2} \\ \Delta y_{t-2} \end{pmatrix} \\ + \begin{pmatrix} 1 \\ 0.75 \end{pmatrix} \begin{pmatrix} 1 & -\frac{\delta}{1-\delta} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

We considered two types of error terms for the VECM above: in the first DGP, labelled DGP # 1 in Table 8, the disturbance term is bivariate normal with a unit variance and a correlation of 0.5; in the second process, labelled DGP #2, the disturbance terms are governed by a bivariate GARCH process (CCC) with a yesterday's news coefficient of 0.25, a coefficient of persistence of 0.74, and a long run variance of 0.01. Note that the coefficient on yesterday's news is larger than what is usually found empirically (between 0.10 and 0.15). The theoretical coefficients in the relationship $\Delta Y_t = -\gamma_1 \Delta y_t + \gamma_2 S_{t-1} + u_t$ are $\gamma_1 = 0$ and $\gamma_2 = \frac{\delta}{1-\delta}$. Here, for simplicity, we set $\frac{\delta}{1-\delta} = 1$ in the DGP, but the cointegrating vector $Y_t - \hat{\theta} y_t$ is estimated using Johansen's approach.

Table 8 reports the empirical rejection frequency at the 5% significance level (nominal size). In the *iid* case, the behavior of the tests are rather similar. The results get much worse in the presence of time-varying conditional variances. With heteroskedastic data, the robust White GMM test has the proper size.

In the last three columns of Table 8 we report the rejection frequencies of the same three tests applied to the logarithms of the variables Y_t and y_t , but for the same DGP in levels. However, it emerges that taking logs does not affect the rejection frequencies.

Table 8: Empirical size (nominal by set to the 5% level)

		$T = 100$					
		Levels			Log levels		
	$VAR(p)$	ζ_{LR}	J_1	J_2	ζ_{LR}	J_1	J_2
DGP #1: iid	$p = 2$	6.27	5.96	5.95	6.25	5.99	5.98
	$p = 3$	7.91	6.93	6.17	7.89	6.93	6.18
	$p = 5$	8.62	6.47	5.00	8.65	6.45	4.98
DGP #2: GARCH	$p = 2$	13.2	12.9	6.07	13.3	13.0	6.05
	$p = 3$	17.6	16.5	5.76	17.6	16.5	5.74
	$p = 5$	21.3	17.8	4.53	21.3	17.8	4.54

B Online Appendix: Common Cyclical Features with Cointegrated data

Parts of this material were extracted from the work of Issler and Vahid (2001), Athanasopoulos et al. (2011), and Issler et al. (2014). For a full discussion, see Engle and Granger (1987), Johansen (1988, 1991), Vahid and Engle (1993, 1997), Hecq et al. (2000, 2006), Cubadda and Hecq (2001), Vahid and Issler (2002), and Athanasopoulos et al. (2011).

B.1 Co-Movement Restrictions in Dynamic Models

Following the previous notation (especially regarding the logarithmic case discussed in this paper), first assume that X_t is a n -vector of $I(1)$ variables, with the stationary ($MA(\infty)$) Wold representation given by:

$$\Delta X_t = C(L) \varepsilon_t, \quad (39)$$

where $C(L)$ is a matrix polynomial in the lag operator, L , with $C(0) = I_n$, $\sum_{j=1}^{\infty} \|C_j\| < \infty$. The vector ε_t is a $n \times 1$ vector of stationary one-step-ahead linear forecast errors

in X_t , given information on the lagged values of X_t . We can rewrite Eq. (39) as:

$$\Delta X_t = C(1) \varepsilon_t + \Delta C^*(L) \varepsilon_t, \quad (40)$$

where $C_i^* = \sum_{j>i} -C_j$ for all i . In particular, $C_0^* = I_n - C(1)$.

If we integrate both sides of Eq. (40), we get:

$$\begin{aligned} X_t &= C(1) \sum_{s=0}^{\infty} \varepsilon_{t-s} + C^*(L) \varepsilon_t \\ &= T_t + C_t. \end{aligned} \quad (41)$$

Eq. (41) is the multivariate version of the Beveridge-Nelson trend-cycle representation (Beveridge and Nelson, 1981). The series X_t are represented as being the sum of a random walk part T_t , which is called the “trend”, and a stationary part C_t , which is called the “cycle”.

Definition 1. The variables in y_t are said to have common trends (or to be cointegrated) if there are r linearly independent vectors, $r < n$, stacked in an $r \times n$ matrix α' , with the following property¹³:

$$\alpha'_{r \times n} C(1) = 0.$$

Definition 2. The variables in y_t are said to have common cycles if there are s linearly independent vectors, $s \leq n - r$, stacked in an $s \times n$ matrix $\tilde{\alpha}'$, with the

¹³This definition could alternatively be expressed in terms of an $n \times r$ matrix γ , such that:

$$C(1)\gamma = 0.$$

The Granger-Representation Theorem (Engle and Granger (1987) shows that if the series in y_t are cointegrated, α and γ in Eq. (43) satisfy:

$$\begin{aligned} C(1)\gamma &= 0, \text{ and,} \\ \alpha' C(1) &= 0. \end{aligned}$$

property that:

$$\tilde{\alpha}'_{s \times n} C^*(L) = 0.$$

Thus, cointegration and common cycles represent restrictions on the elements of $C(1)$ and $C^*(L)$ respectively.

We now discuss restrictions on the dynamic autoregressive representation of economic time series that arise from cointegration (common trends) and common cycles. First, we assume that X_t is generated by a vector autoregression (VAR):

$$X_t = \Gamma_1^* X_{t-1} + \dots + \Gamma_p^* X_{t-p} + \varepsilon_t \quad (42)$$

If elements of X_t are cointegrated, then the matrix $I - \sum_{i=1}^p \Gamma_i^*$ must have less than full rank, which imposes cross-equation restrictions on the VAR. In this case, Engle and Granger (1987) show that the system in Eq. (42) can be written as a vector error-correction model (VECM):

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_{p-1} \Delta X_{t-p+1} + \gamma \alpha' X_{t-1} + \varepsilon_t \quad (43)$$

where γ and α are full rank matrices of order $n \times r$, r is the rank of the cointegrating space, $-\left(I - \sum_{i=1}^p \Gamma_i^*\right) = \gamma \alpha'$, and $\Gamma_j = -\sum_{i=j+1}^p \Gamma_i^*$, $j = 1, \dots, p-1$. Given the cointegrating vectors stacked in α' , it can be seen that Eq. (43) parsimoniously encompasses Eq. (42). Conditional on a knowledge of the cointegrating vectors, the VECM has $n^2(p-1) + n \cdot r$ parameters in the conditional mean, while the VAR has $n^2 \cdot p$ parameters. Thus, the former has $n \cdot (n-r)$ fewer parameters, since $r < n$. If we take into account the free parameters in the cointegrating vector, the VECM has $n^2(p-1) + 2n \cdot r - r^2$ mean parameters, $(n-r)^2$ fewer than the VAR.

Vahid and Engle (1993) show that a dynamic representation of the data X_t may have additional cross-equation restrictions if there are common cycles. To see this, recall that the cofeature vectors $\tilde{\alpha}'_i$, stacked in an $s \times n$ matrix $\tilde{\alpha}'$, eliminate all serial correlation in ΔX_t , i.e. $\tilde{\alpha}' \Delta X_t = \tilde{\alpha}' \varepsilon_t$. Since the cofeature vectors are identified only up to an invertible transformation, we can, without loss of generality, rotate $\tilde{\alpha}$

to have an s dimensional identity sub-matrix:

$$\tilde{\alpha} = \begin{bmatrix} I_s \\ \tilde{\alpha}_{(n-s) \times s}^* \end{bmatrix}.$$

Considering $\tilde{\alpha}' \Delta X_t = \tilde{\alpha}' \varepsilon_t$ as s equations in a system, and completing the system by adding the unconstrained VECM equations for the remaining $n - s$ elements of ΔX_t , we obtain

$$\begin{bmatrix} I_s & \tilde{\alpha}^{*'} \\ \mathbf{0}_{(n-s) \times s} & I_{n-s} \end{bmatrix} \Delta X_t = \begin{bmatrix} \mathbf{0}_{s \times (np+r)} \\ \Gamma_1^{**} & \dots & \Gamma_{p-1}^{**} & \gamma^* \end{bmatrix} \begin{bmatrix} \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-p+1} \\ \alpha' X_{t-1} \end{bmatrix} + v_t, \quad (44)$$

where Γ_i^{**} and γ^* represent the partitions of Γ_i and γ respectively, corresponding to the bottom $n - s$ reduced form VECM equations, and $v_t = \begin{bmatrix} I_s & \tilde{\alpha}^{*'} \\ \mathbf{0}_{(n-s) \times s} & I_{n-s} \end{bmatrix} \varepsilon_t$.

It is easy to show that Eq. (44) parsimoniously encompasses Eq. (43). Since $\begin{bmatrix} I_s & \tilde{\alpha}^{*'} \\ \mathbf{0}_{(n-s) \times s} & I_{n-s} \end{bmatrix}$ is invertible, it is possible to recover Eq. (43) from Eq. (44). Notice however that the latter has $s \cdot (np + r) - s \cdot (n - s)$ fewer parameters. Estimation of the parameters in the system Eq. (44), imposing short- and long-run rank conditions, can be performed by full-information maximum likelihood (FIML).

Hecq, Palm and Urbain (2000, 2006) made a distinction between what they labelled *strong-form serial-correlation common features* and *weak-form serial-correlation common features*. The former entails the usual restrictions considered by Vahid and Engle (1993):

$$\tilde{\alpha}' \Gamma_1 = \tilde{\alpha}' \Gamma_2 = \dots = \tilde{\alpha}' \Gamma_{p-1} = 0, \text{ and} \quad (45)$$

$$\tilde{\alpha}' \gamma = 0. \quad (46)$$

They argue that, in some instances, imposing restrictions in short-run dynamics (Eq.

(45)), coupled with those regarding the cointegrating vector (Eq. (46)), may be too restrictive. As we saw in Section 2.1, this is the case with PVMs, where a necessary condition to have a first-order difference equation generating the restrictions in PVM is that Eq. (46) does not hold.

Whenever we only impose the restrictions in Eq. (45), not those in Eq. (46), we obtain what they labelled *weak-form serial-correlation common features*. In this case:

$$\tilde{\alpha}' [\Delta X_t - \gamma \alpha' X_{t-1}] = \tilde{\alpha}' \varepsilon_t, \quad (47)$$

i.e., we only inherit an unpredictable linear combination of ΔX_t once we *control* for the long-run deviations $\alpha' X_{t-1}$ that stem from cointegration. From Eq. (45) and (46), it is immediately seen that *weak-form SCCF* is less restrictive, although it still imposes rank restrictions on the short-term coefficient matrices.

Athanasopoulos et al. (2011) show that *weak-form serial-correlation common features* can be represented using the triangular representation of a cointegrated system that is used extensively in the cointegration literature, where the n -dimensional time series, partitioned as:

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix},$$

where X_{1t} is $r \times 1$ (implying that X_{2t} is $(n - r) \times 1$) is generated from:

$$\begin{aligned} X_{1t} &= \beta X_{2t} + u_{1t} \\ \Delta X_{2t} &= u_{2t} \end{aligned} \quad (48)$$

where β is a $r \times (n - r)$ matrix of parameters, and

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

is a strictly stationary process with mean zero and a positive definite covariance matrix. In other words, Eq. (48) is a data generating process (DGP) of a system of n cointegrated $I(1)$ variables with r cointegrating vectors. The extra feature that

they add to this fairly general DGP is that u_t is generated from a VAR of finite order $p - 1$ and rank q ($< K$):

$$u_t = B_1 u_{t-1} + B_2 u_{t-2} + \cdots + B_{p-1} u_{t-p+1} + \eta_t, \quad (49)$$

where B_1, B_2, \dots, B_p are $n \times n$ matrices with $\text{rank} \begin{bmatrix} B_1 & B_2 & \dots & B_{p-1} \end{bmatrix} = q$, and η_t is an i.i.d. sequence with mean zero, a positive definite variance-covariance matrix, and finite fourth moments. Estimation imposing these rank restrictions can be performed using the two-step maximum likelihood iterative procedure described at length in the text.

B.2 Remaining empirical results for forecasting models

In this section we present the tables containing the remaining empirical results for the forecasting models. As was explained in Section 4.2, we focused there on results for the target variables in the PV formula – Y_t . Here, however, we also present results for y_t , the series on the right-hand side of Eqs. (1) or (3).

Table 9: Relative RMSFEs of restricted models vs VAR for i_{sr} . Forecast period up to 2007.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	1.058 [0.27]	1.056 [0.216]	1.014 [0.013]	1.057 [0.124]	1.051 [0.246]	1.206 [1.409]	1.005 [0.001]	0.886 [0.795]	0.907 [0.63]	0.955 [0.122]	0.908 [0.59]	0.842 [1.208]
VECM(HQ-J)	0.952 [0.906]	0.955 [0.839]	0.997 [0.011]	0.919 [1.452]	0.949 [0.9]	0.96 [0.58]	0.927 [2.119]	0.892 [1.56]	0.845 [1.517]	0.88 [1.3]	0.86 [1.345]	0.84 [1.317]
VECM(HQ)Rank	0.952 [0.948]	0.871 [1.122]	0.901 [1.317]	0.987 [0.471]	0.931 [0.931]	0.958 [0.341]	0.955 [1.971]	0.914 [1.593]	0.894 [1.597]	0.886 [1.406]	0.865 [1.367]	0.878 [1.273]
PV	0.946 [1.372]	0.923 [0.843]	0.932 [0.965]	0.956 [1.131]	0.923 [0.836]	0.896 [1.191]	0.874 [1.514]	0.895 [1.196]	0.804 [1.549]	0.801 [1.328]	0.81 [1.235]	0.801 [1.248]

See the notes to Table 2.

Table 10: Relative RMSFEs of restricted models vs VAR for i_{sr} . Forecast period up to 2012.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	1.13 [1.345]	1.043 [0.212]	0.946 [0.119]	1.166 [1.372]	0.955 [0.163]	1.133 [0.877]	1.191 [0.996]	0.919 [0.635]	0.94 [0.393]	1.067 [0.359]	0.862 [0.855]	0.775 [1.331]
VECM(HQ-J)	1.008 [0.112]	0.944 [1.594]	0.986 [0.601]	0.973 [1.142]	0.939 [1.21]	0.928 [1.172]	0.935 [2.348]	0.94 [1.587]	0.895 [1.475]	0.913 [1.262]	0.866 [1.257]	0.873 [1.207]
VECM(HQ)Rank	1.038 [0.518]	0.912 [0.916]	0.896 [1.289]	1.003 [0.000]	0.912 [1.41]	0.887 [1.41]	0.885 [1.941]	0.926 [1.62]	0.843 [1.469]	0.895 [1.342]	0.832 [1.3]	0.853 [1.246]
PV	0.984 [0.303]	0.927 [1.119]	0.922 [1.209]	0.967 [1.243]	0.882 [1.64]	0.876 [1.394]	0.886 [1.626]	0.883 [1.292]	0.805 [1.504]	0.829 [1.349]	0.76 [1.32]	0.758 [1.292]

See the notes to Table 2.

Table 11: Relative RMSFEs of restricted models vs VAR for D_t . Forecast period up to 2007.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.977 [0.927]	0.989 [0.623]	0.901** [4.317]	0.867*** [9.852]	0.875* [3.5]	0.902 [0.085]	0.928 [0.034]	0.946 [1.51]	0.949 [0.000]	0.973 [0.448]	0.995 [0.034]	0.973 [0.285]
VECM(HQ-J)	0.931*** [6.922]	0.887*** [7.557]	0.841*** [9.489]	0.838*** [13.597]	0.895 [0.766]	0.924 [0.036]	0.944 [0.125]	0.941 [1.598]	0.931 [0.055]	0.958 [1.425]	0.982 [0.204]	0.963 [0.585]
VECM(HQ)Rank	0.958* [3.092]	0.983 [2.676]	0.869*** [7.495]	0.854*** [9.022]	0.898 [0.824]	0.917 [0.016]	0.935 [0.237]	0.933 [1.76]	0.923 [0.028]	0.955 [1.885]	0.982 [0.206]	0.965 [0.584]
PV	0.958** [4.794]	1.107 [0.059]	1.022 [0.472]	1.032 [0.042]	1.267 [2.458]	1.397** [5.715]	1.467*** [6.753]	1.42 [1.279]	1.375 [2.271]	1.401 [1.486]	1.403 [2.189]	1.337* [2.89]

See the notes to Table 2.

Table 12: Relative RMSFEs of restricted models vs VAR for D_t . Forecast period up to 2012.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.931 [0.039]	0.9 [2.609]	0.871* [2.742]	0.907** [4.489]	0.991 [0.501]	1.093 [0.445]	1.108 [0.506]	0.948 [0.000]	0.619 [1.363]	0.632 [1.15]	0.739 [1.056]	0.868 [0.722]
VECM(HQ-J)	0.872*** [11.218]	0.809*** [9.108]	0.819*** [9.954]	0.891* [3.088]	1.023 [0.084]	1.114 [0.502]	1.123 [0.58]	0.954 [1.017]	0.618 [1.378]	0.63 [1.156]	0.723 [1.06]	0.868 [0.74]
VECM(HQ)Rank	0.889* [3.562]	0.878* [3.281]	0.863** [5.57]	0.937 [0.279]	1.06 [0.212]	1.151 [0.59]	1.164 [0.649]	0.989 [0.362]	0.584 [1.368]	0.626 [1.148]	0.76 [0.993]	0.923 [0.383]
PV	0.888** [4.878]	0.982 [1.198]	0.971 [1.406]	1.075 [0.567]	1.359** [4.698]	1.446** [6.42]	1.422*** [6.763]	1.161 [0.249]	0.843 [0.492]	0.624 [0.795]	0.691 [0.674]	0.833 [0.445]

See the notes to Table 2.

Table 13: Relative RMSFEs of restricted models vs VAR for d_t . Forecast period up to 2007.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.98 [2.433]	0.964 [1.977]	0.911 [1.995]	0.891 [0.998]	0.911 [0.761]	0.937 [0.571]	0.952 [0.718]	0.966 [0.992]	0.971 [0.908]	0.977 [0.475]	0.968 [0.338]	0.949 [0.474]
VECM(HQ-J)	0.952 [2.304]	0.873** [5.904]	0.806*** [7.568]	0.826** [5.988]	0.928 [0.99]	0.958 [0.477]	0.965 [0.604]	0.941 [2.013]	0.917 [2.646]	0.937 [1.434]	0.939 [1.022]	0.93 [1.026]
VECM(HQ)Rank	0.964 [0.234]	0.94 [0.002]	0.812 [2.574]	0.825** [4.742]	0.919 [1.618]	0.945 [0.755]	0.955 [0.706]	0.934 [2.188]	0.912* [2.952]	0.929 [1.762]	0.938 [1.161]	0.934 [1.022]
PV	0.907 [1.742]	3.324 [2.356]	5.371** [5.903]	5.929** [4.93]	5.909*** [9.502]	6.016*** [9.275]	6.221*** [10.018]	6.355*** [8.809]	7.327*** [9.065]	8.269*** [10.561]	8.408*** [10.227]	8.308*** [10.967]

See the notes to Table 2.

Table 14: Relative RMSFEs of restricted models vs VAR for d_t . Forecast period up to 2012.

Horizon:	1	2	3	4	5	6	7	8	9	10	11	12
VECM(HQ-PIC)	0.961 [0.069]	0.966 [0.363]	0.929 [0.387]	0.893 [1.373]	0.86 [0.212]	0.873 [0.61]	0.88 [1.491]	0.885 [2.175]	0.942 [2.097]	0.969 [0.963]	0.991 [0.157]	0.991 [0.071]
VECM(HQ-J)	0.921** [4.918]	0.889* [2.804]	0.826** [4.763]	0.809** [4.612]	0.869 [1.375]	0.86 [1.124]	0.851 [1.438]	0.841 [2.494]	0.928* [2.875]	0.96 [1.666]	0.977 [0.774]	0.976 [0.641]
VECM(HQ)Rank	0.933** [4.472]	0.976 [0.679]	0.865 [2.077]	0.807* [3.544]	0.844 [1.975]	0.848 [1.899]	0.859 [1.671]	0.852 [2.515]	0.924* [3.122]	0.952 [1.942]	0.974 [0.816]	0.973 [0.823]
PV	0.819*** [7.448]	2.777*** [18.576]	4.943*** [24.418]	6.345*** [20.669]	6.77*** [14.578]	6.801*** [9.772]	6.864*** [8.519]	7.005*** [8.137]	5.678*** [7.423]	5.863** [6.5]	5.779** [5.757]	5.538** [5.273]

See the notes to Table 2.