FUNDAÇÃO GETULIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Diogo Vinícius Menezes Saraiva

Essays in Macroeconometrics

Rio de Janeiro
2015

Diogo Vinícius Menezes Saraiva

Essays in Macroeconometrics

Tese submetida à Escola de Pós-Graduação em Economia como requisito parcial para a obtenção do grau de Doutor em Economia

Área de concentração: Macroeconometria

Orientador: João Victor Issler

Rio de Janeiro
2015

Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Saraiva, Diogo Vinícius Menezes
Essays in macroeconometrics/ Diogo Vinícius Menezes
Saraiva. – 2015.
78 f.

Tese (doutorado) - Fundação Getulio Vargas, Escola de Pós-
Graduação em Economia.
Orientador: João Victor Issler.
Inclui bibliografia.

1. Macroeconomia. 2. Econometria. 3. Produto interno bruto. 4. Bolsa de mercadorias. 5. Macroeconomia - Rússia. I. Issler, João Victor. II. Fundação GetulioVargas. Escola de Pós-Graduaçãoem Economia. III. Título.

CDD – 339

114203002

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA
CURSO DE DOUTORADO EM ECONOMIA

ATA DE DEFESA DE TESE DO CURSO PARA CONCESSÃO DO GRAU DE
DOUTOR EM ECONOMIA

ALUNO (A): DIOGO VINICIUS MENEZES SARAIVA

ORIENTADOR (A): JOÃO VICTOR ISSLER

BANCA EXAMINADORA :

JOÃO VICTOR ISSLER
CAIO IBSEN RODRIGUES DE ALMEIDA
FELIPE SARAIVA IACHAN
OSMANI TEIXEIRA DE CARVALHO GUILLEN
TIAGO COUTO BERRIEL

TÍTULO DA TESE: ESSAYS IN MACROECONOMETRICS.

LOCAL: FUNDAÇÃO GETULIO VARGAS **SALA:** CONGREG../11º **HORÁRIO DE INÍCIO:** 15H00MIN

( ) APROVADO;
(X) APROVADO COM RESTRIÇÃO, DEVENDO O(A) ALUNO(A) FAZER AS ALTERAÇÕES PROPOSTAS PELA BANCA ATÉ ___/___/___ ;
( ) REPROVADO.

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA:

JOÃO VICTOR ISSLER

CAIO IBSEN RODRIGUES DE ALMEIDA

Felipe S. Iachan

FELIPE SARAIVA IACHAN

OSMANI TEIXEIRA DE CARVALHO GUILLEN

TIAGO COUTO BERRIEL

Rio de janeiro, 27 de novembro de 2015.

ASSINATURA DO ALUNO: Diogo Saraiva

FGV

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA
CURSO DE DOUTORADO EM ECONOMIA

FOLHA DE ALTERAÇÕES PROPOSTAS PELA BANCA EXAMINADORA DA DEFESA DE TESE

ALUNO (A): DIOGO VINICIUS MENEZES SARAIVA

ORIENTADOR (A): JOÃO VICTOR ISSLER

BANCA EXAMINADORA :

JOÃO VICTOR ISSLER
CAIO IBSEN RODRIGUES DE ALMEIDA
FELIPE SARAIVA IACHAN
OSMANI TEIXEIRA DE CARVALHO GUILLEN
TIAGO COUTO BERRIEL

ALTERAÇÕES PROPOSTAS PELA BANCA

implantar mudanças de forma sugeridas pela banca.

RESULTADO FINAL:

ALTERAÇÕES ATENDIDAS (✓)
ALTERAÇÕES NÃO ATENDIDAS ( )

Rio de Janeiro, 6 de junho de 2016

ASSINATURA DO ORIENTADOR: ____________

Nowcasting Russia GDP: Macroeconomic News and Commodity Prices Predictability

Diogo Saraiva*

London Business School and FGV

October 1, 2015

Abstract

The knowledge of the current state of the economy is crucial for policy makers, economists and analysts. However, a key economic variable, the gross domestic product (GDP), are typically colected on a quartely basis and released with substancial delays by the national statistical agencies. The first aim of this paper is to use a dynamic factor model to forecast the current russian GDP, using a set of timely monthly information. This approach can cope with the typical data flow problems of non-synchronous releases, mixed frequency and the curse of dimensionality. Given that Russian economy is largely dependent on the commodity market, our second motivation relates to study the effects of innovations in the russian macroeconomic fundamentals on commodity price predictability. We identify these innovations through a news index which summarizes deviations of offical data releases from the expectations generated by the DFM and perform a forecasting exercise comparing the performance of different models.

Keywords: GDP Nowcasting, dynamic factor models, news index, commodity price predictability.

*Under the supervision of Prof. Lucrezia Reichlin (LBS)

1 Introduction

Policy makers, economists and analysts have imperfect knowledge of the present state of the economy and are continually concerned of how to assess it. Policy measures taken by Governments and central banks are driven by the ongoing economic situation. Therefore an accurate and timely assesment of gross domestic product (GDP) growth rates for the current quarter have a vital importance for policy makers. However the first official estimates for GDP are published with a considerable delay, around 2 or 3 months after the end of the reference quarter (which are often revised). The first aim of the present work is, therefore, to exploit a framework which aims to cope with this issue. We nowcast russian GDP using a large set of monthly hard and soft data.

For this purpose, we use a dynamic factor model (DFM) approach studying how it compares with other approaches and your forecasting performance. We employ the DFM methodology advocated by Bańbura et al. (2010) and Giannone et al. (2008). Following Doz et al. (2012) and Bańbura and Modugno (2014) the parameters in the model are estimated by maximum likelihood implemented by the Expected Maxisimation (EM) algorithm.

The results of our empirical study generally show that the DFM model performs well in nowcasting Russian GDP, generally outperforming the statistical benchmarking autoregressive model (AR) and professional forecasters. The model's RMSFE overall has a declining pattern over the prediction period, meaning that is useful to update the nowcast of GDP with each data release. Also, we notice that surveys are particularly important in the early part of the calendar, while in the last months hard variable releases become more important.

A second objective of this paper is to study the relationship between innovations in macroeconomic fundamentals of Russia economy and commodity price. Russia is the world's largest producer of oil and natural gas, on a par with Saudia Arabia and USA, respectively. Russia alone accounts for around 20% of the world's natural gas production and 14% of oil production. Russia exports around 70% of its crude oil and 30% of its natural gas production, which account for 70% of the value of Russia's exports. At the same time Russia is the world's second largest natural gas consumer after the U.S.. Therefore, it is not surprising that the energy sector comprises a huge part of the domestic economy. Following that we investigate if movements of Liquified Natural Gas (LNG) price are linked to russian macroeconomic fundamentals.

Following Bańbura et al. (2010), the innovations in macroeconomic fundamentals are based on macroeconomic *news* defined as the difference between the official data releases from the expectations generated by the DFM. From these surprises an index is constructed, the news are weighted where the weights depend on the contribution of the associated economic indicator to the forecast of GDP in the nowcasting model. Some

recent academic paper has given attention to these kind of instrument, see Altavilla et al. (2014).

The in-sample results indicates that there is some relationship among the indexes and the LNG price. Investigating further through an out-of-sample analysis, we found out that autoregressive models added with the index had the overall best performance. We should highlight that the results are stronger for the short-term, which makes it even more significant.

The remainder of the paper is organized as follows. In section 2 we describe the problem of nowcasting in general and explain the details of our DFM approach. In section 3 we discuss the related literature, while in section 4 we describe the data used in the forecast exercise. Section 5 shows the forecasting results of our DFM. Section 6 is devoted to explain our news index and demonstrates it effects on the LNG price. Section 7 concludes this study.

2 Methodology

In Russia, the GDP is released usually between two and three months after the close of the reference quarter. The nowcasting framework arises from the fact we can estimate the GDP in the meantime using an inflow of new information which is available in a higher-frequency(monthly). For this purpose, in this paper we use the methodology described in Bańbura et al. (2010), to produce "nowcasts" for Russia GDP, where its defined as prediction of the present, the very near future and the very recent past.

Let us denote by Ω_v, a vintage of data at v where v is the date-time of a particular economic variable release. Further consider the GDP growth at time t as y_t^Q, the nowcasting of y_t^Q is defined as the orthogonal projection of y_t^Q on the available information set Ω_v:

$$\mathbb{E}\left[y_t^Q|\Omega_v\right] \tag{1}$$

where $\Omega_v = \left\{x_{i,t_i}, t_i = 1, 2, ...T_{i,v}, i = 1, ..., n; y_{3k}^Q, 3k = 3, 6, ..., T_{Q,v}\right\}$, where $T_{i,v}$ corresponds to the last period for which in vintage v the series i has been observed. The structure of information set Ω_v is tipically referred as "jagged edge", it means that data are released in a non-synchronous manner, with different degrees of delay and also it contains mixed frequency series, monthly and quartely in our exercise. Then $T_{i,v}$ is not the same across variables.

The first nowcasts are usually made with very little or no information on the reference quarter, as long as subsequent data are released, a sequence of projections are performed: $\mathbb{E}\left[y_t^Q|\Omega_v\right]$, $\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right]$,.. where v, $v+1$,... refer to dates of consecutive data releases.

Analysing the difference between Ω_v and Ω_{v+1}, at time $v+1$, a certain group of

variables are released, $x_{j,T_{j,v+1}|j\in\mathbb{J}_{v+1}}$, and consequently the information set is expanded by the introduction of these new information. More formally, $\Omega_v \subseteq \Omega_{v+1}$ and $\Omega_{v+1}\backslash\Omega_v = x_{j,T_{j,v+1}|j\in\mathbb{J}_{v+1}}$. Therefore, the new forecast can be decomposed as:

$$\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right] = \mathbb{E}\left[y_t^Q|\Omega_v\right] + \mathbb{E}\left[y_t^Q|I_{v+1}\right] \tag{2}$$

where I_{v+1} is the information in Ω_{v+1} not in Ω_v, i.e.:

$$I_{v+1,j} = x_{j,T_{j,v+1}} - \mathbb{E}\left[x_{j,T_{j,v+1}}|\Omega_v\right] \tag{3}$$

and $I_{v+1} = \left(I_{v+1,1}...I_{v+1,J_{v+1}}\right)'$, where J_{v+1} denotes the number of elements in $\mathbb{J}_{v+1}$.

From the equation above we can read that the change in the nowcast is due only to the "unexpected" part of the data release, i.e., the surprising part of the release with respect to the model forecast. We label this part of the release, I_{v+1}, as the *news*. For instance, a negative *news* in the industrial production index should revise the GDP forecast downwards, on the other hand, if the release number is exactly as predicted by the model, the nowcast for GDP will not be revised. To make this point clear, assuming that the data are Gaussian, we can further develop the following equation:

$$\underbrace{\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right]}_{\text{new forecast}} - \underbrace{\mathbb{E}\left[y_t^Q|\Omega_v\right]}_{\text{old forecast}} = \sum_{j\in\mathbb{J}_{v+1}} b_{j,t,v+1} \underbrace{\left(x_{j,T_j,v+1} - \mathbb{E}x_{j,T_{j,v+1}|\Omega_v}\right)}_{\text{news}}. \tag{4}$$

Where the weight $b_{j,t,v+1}$ quantifies the relevance of the *news* for the target variable. From (4), it is possible to decompose the contribution from the *news* of each individual variable into the forecast revision. Therefore, it is possible to comment the revision of the target variable in relation to unexpected developments of a particular input.

In this paper we follow the approach proposed by Giannone et al. (2008) to deal with the problems of mixed frequency, jagged edge, possibly missing data and the curse of dimensionality arisen from the nowcasting framework. Their solution to these problems consists of modelling the monthly data as a parametric dynamic factor model cast in a state space representation, using the Kalman filter techiniques to perform the projections. The estimation is made by maximum likelihood adopted in Bańbura and Modugno (2014).

2.1 Monthly factor model

Let $x_t = (x_{1,t}, x_{2,t}, ..., x_{n,t})'$ denote the monthly series, which can be in level, difference or month-on-month growth rates so as to satisfy the assumption of stationarity. We assume that x_t obey the following factor model representation:

$$\begin{aligned} x_t &= \mu + \Lambda f_t + \epsilon_t & (5) \\ \epsilon_{i,t} &= \beta_i \epsilon_{i,t-1} + e_t, \quad e_t \sim N(0, \sigma_i^2) & (6) \end{aligned}$$

where f_t is a 3×1 vector of (unbsorved) common factor and ϵ_t is a vector of idionsyncratic components both assumed to have mean zero. Λ denotes the factor loadings for the monthly variables and $\mu = (\mu_1, \mu_2, ..., \mu_n)'$ are the unconditional means. We assume that the idionsyncratic component of the monthly variables follow an AR(1) process, with $\mathbb{E}[e_{i,t}e_{j,s} = 0]$ for $i \neq j$. Factors are modelled as a VAR(2) process:

$$f_t = A_1 f_{t-1} + A_p f_{t-2} + u_t, \quad u_t \sim i.i.d.N(0, Q) \tag{7}$$

where $A_1, ..., A_p$ are 3×3 matrices of autoregressive coefficients.

The factors f_t are partioned into mutually independent global, hard and soft. It is assumed that global factor is loaded by all the variables while hard and soft factors are specific to hard and soft variables, respectively. This framework is used to account for the local cross-sectional correlation within the hard and soft blocks, which helps for a more efficient extraction of the global factor. Maximum likelihood estimation accomodates this type of restrictions. Therefore, we have:

$$\begin{pmatrix} Q = & \Lambda_{S,G} & \Lambda_{S,S} & 0 \\ & \Lambda_{H,G} & \Lambda_{H,H} & 0 \end{pmatrix}$$

$$\begin{pmatrix} & f_t^G \\ f_t = & f_t^S \\ & f_t^H \end{pmatrix} \text{ and } \begin{pmatrix} & A_{i,G} & 0 & 0 \\ A_i = & 0 & A_{i,S} & 0 \\ & 0 & 0 & A_{i,H} \end{pmatrix},$$

$$\begin{pmatrix} & Q_G & 0 & 0 \\ Q = & 0 & Q_S & 0 \\ & 0 & 0 & A_H \end{pmatrix}$$

2.2 Quartely variables

In our model, we incorporate two quartely variables: GDP and consumer confidence index. Quartely variables are introduced by constructing for each of them a partially observed monthly counterpart. The value of the quartely variable is "assigned" to the third month of the respective quarter and the quartely value is defined as the sum of the unobserved monthly contributions:

$$Y_t^Q = Y_t^M + Y_{t-1}^M + Y_{t-2}^M \quad t = 3, 6, 9... \tag{8}$$

where for instance, for the GDP variable, $Y_t^Q = 100 \times \log(GDP_t^Q)$ and $Y_t^M = 100 \times \log(GDP_t^M)$. We assume that the unobserved monthly growth rate of GDP, $y_t = \Delta Y_t^M$,

admits the same factor representation as the monthly hard variables:

$$
\begin{aligned}
y_t &= \mu_Q + \Lambda_Q f_t + \epsilon_t^Q && (9)\\
\epsilon_t^Q &= \alpha_Q \epsilon_{t-1}^Q + e_t^Q, \quad e_t^Q \sim N(0, \sigma_Q^2) && (10)
\end{aligned}
$$

with $\Lambda_Q = (\Lambda_{Q,G}, 0, \Lambda_{Q,H})$

The link between the onbserved GDP data and the monthly y_t is made through the following relation:

and use the approximation of Mariano and Murasawa (2003):

$$
\begin{aligned}
y_t^Q &= Y_t^Q - Y_{t-3}^Q && (11)\\
&\approx (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - (Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) && (12)\\
&= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}, \quad t = 3, 6, 9... && (13)\\
& && (14)
\end{aligned}
$$

2.3 Estimation and forecasting

Let us define $\bar{x}_t = (x_t', y_t^Q)'$ and $\bar{\mu} = (\mu', \mu_Q)'$. The model described by equations (6)-(14) can be cast in a state space representation:

$$
\begin{aligned}
\bar{x}_t &= \bar{\mu} + Z(\theta)\alpha_t && (15)\\
\alpha_t &= T(\theta)\alpha_{t-1} + \eta_t, \quad \eta \sim N(0, \Sigma_\eta(\theta)) && (16)
\end{aligned}
$$

where the vector of states $\alpha_t = (f_t', f_{t-1}', f_{t-2}', f_{t-3}', f_{t-4}', \epsilon_{1,t}, ..., \epsilon_{n,t}, \epsilon_t^Q, \epsilon_{t-1}^Q, \epsilon_{t-2}^Q, \epsilon_{t-3}^Q, \epsilon_{t-4}^Q)'$ and the parameters are collected in $\theta = (\bar{\mu}, \Lambda, \Lambda_Q, A_1, A_2, \beta_1, ...\beta_n, \beta^Q, \gamma_1, ..., \gamma_n, \gamma^Q)$. The details of the state space representation and the structure of the matrices, $Z(\theta)$ $T(\theta)$ and $\Sigma_\eta(\theta)$ can be found in Banbura, Giannone and Reichlin (2010).

The estimation of θ is made by maximum likelihood implemented by the the Expectation Maximisation (EM) algorithm, approach proposed by Doz et al. (2012) and Bańbura and Modugno (2014), which can be seen for more details. The method consists of a iterative procedure, which involves a sequence of two alternating steps:

1. E-step - Conditional on the data, the expectation of the log-likelihood is calculated using the estimates from the previous iteration, $\theta(j)$:

$$
L(\theta, \theta(j)) = \mathbb{E}_{\theta(j)}\left[l(\bar{x}, \alpha; \theta)\right]; \tag{17}
$$

2. M-step - the new parameters, $\theta(j+1)$, are estimated through the maximisation of the expected log-likelihood (from the previous iteration) with respect to θ.

Given an estimate of θ, the nowcasts as well as the estimates of the factors or any missing observation in $\bar{x}_t$, can be obtained from the Kalman filter or smoother. This procedure allows to deal easily with the missing data issue, restriction on the parameters, serial correlation of the idionsycratic component and also should be more efficient than the popular non-parametric method based on principal components.

3 Related Literature

The nowcasting methodology described in the previous section , relies on the assumption that the data are driven by few unobservable factors. The econometric framework and the estimation procedure we have used in this paper relies on the work of Giannone et al. (2008), Bańbura et al. (2010) and Bańbura and Modugno (2014). Some other applications of the factor model approach can be found in Angelini et al. (2008), Marcellino and Schumacher (2008), Bańbura and Rünstler (2011), Antipa et al. (2012), amongst others.

A traditional approach to nowcasting, largely used at various central banks, its the called bridge models. It is composed of a single equation of the quartely target variable which is regressed on its lags and on some monthly predictors that are converted to the target variable frequency. Some applications are on Trehan (1989), Parigi and Schlitzer (1995), Golinelli and Parigi (2007), amongst others.

Another solution to forecasting low frequency variables with high frequency predictors was proposed by Ghysels et al. (2004), called MIDAS (Mixed Data Sampling) regressions. As the bridge models, the MIDAS strategy is a single equation approach, however it does not require the frequency conversion as it involves a parsimoniously parameterized distributed lag polynomial for the high frequency regressors. Clements and Galvão (2008) and Marcellino and Schumacher (2010) used the MIDAS strategy in the context of nowcasting. However the single equation approaches are quite simple and can deal with the forecasting problem, they hinder an understanding of nowcast revisions in terms of consecutive data releases.

Rautava (2004) discusses the impact of international oil prices and the real exchange rate on the russian economy using a vector autoregressive (VAR) model and cointegration techniques.

Regarding the literature on commodity prices, Chen (2014) shows that oil-sensitive stock price indices in the energy sector have predictive power over nominal and real crude oil prices at short horizons (one month ahead predictions). Borensztein and Reinhart (1994) study the main economic fundamentals behind the behavior of commodity prices

Release name	**F**	**S**	**SD**	**Units/Transf**
PMI: Manufacturing	M	HSBC/MKT	1997	Index/levels
PMI: Services Composite	M	HSBC/MKT	2001	Index/levels
Output: Industrial Production Index: Total	M	GKS/H	1995	MoM %/MoM
Exports of Goods	M	CBRF/H	1995	MoM %/MoM
Imports of Goods	M	CBRF/H	1995	MoM %/MoM
Total Real Retail Trade [Sales]	M	GKS/H	1995	MoM %/MoM
Labor Market: Employment	M	GKS/H	1996	MoM %/MoM
Business Confidence: Manufacturing	M	GKS	2005	Index/levels
Consumer Confidence Index	Q	GKS	2006	Index/diff
Gross Domestic Product	Q	GKS	1995	QoQ %/QoQ

Table 1: Dataset

and try to quantify the importance of these factors. Chen et al. (2010) show that currencies of countries which depend heavily on the export of commodities have power in predicting global commodity prices, both in-sample and out-of-sample. In a more recent work, Gargano and Timmermann (2014) combine the commodity currencies with macroeconomic variables to forecast commodity price indexes, finding that both have predictive power.

4 Data

Let's comment on the data used in the nowcasting exercise. It contains ten major indicators on the Russia economy. The series are presented in table (1). The dataset can be divided between "hard" data, composed by real activity indicators, and "soft" data, formed by surveys. The data was downloaded from Haver database, starting from January 1995 until December 2013. It cointains monthly and quarterly variables.

There are substancial differences in terms of timeliness among the variables. Survey series often are already available at the end of the respective reference period, on the other hand, hard data on real activity are released with longer delay, usually with 2-3 months after the end of the reference period. While surveys are important for their timeliness, being the only information available at that time, hard data tipically carry a more precise signal for GDP developments. In this paper the data set is composed of soft and hard variables so as to benefit from timeliness as well as precision.

4.1 The calendar

The 'preliminary' estimate of GDP in Russia is published by the Office for National Statistics (ONS) towards the end of the month following the reference quarter (i.e. around 130 days after the beginning of the reference quarter), followed by a second estimate a month later (160 days) and a final estimate released at the end of the next quarter (around 180 days). Our aim is to predict GDP before these official numbers are published by taking advantage of the information in the flow of economic data releases that precede them, updating our prediction with each successive data release. The calendar of data releases that we use for the Russia model is shown in table (2).

As in other countries, surveys are particularly important in the early part of the calendar. The timeliest indicators is the Business Confidence Manufacturing survey which is published around the end of the third week of the reference month; following these, Manufacturing and Services Composite PMIs are published within the first few working days of the month following the reference month. Consumer Confidence Index is published around the first week of the following quarter.

Hard data are published typically starting from the second half of the month following the reference month. Retail trade, Employment and Output Indicators (Construction, Agriculture and Transportation) are released with a delay of two/three weeks from the end of the reference month, while Exports and Imports of goods are published around the second week of the second following month. Table (2) below give a calendar overview.

4.2 Surveys

The surveys that we use within the Russia model are:

1. Russia: Business Confidence: Manufacturing
2. Russia:PMI: Manufacturing
3. Russia:PMI: Services Composite
4. Russia: Consumer Confidence Index

In Russia, the business confidence index is based on the management reports of around 4000 companies representing three basic industries - subsurface resource extraction, processing and electricity, gas and water production and distribution. The index shows the difference between the percentage share of executives that are optimistic and the percentage of that is pessimistic. The index takes a value between -100 (all responding entities asses their situation as poor and expect it to become worse) up to 100 (all participants are satisfied with the current situation and expect it to improve); 0 indicates neutrality.

Data Series	Days	Reference Month
Russia: Business Confidence: Manufacturing	-6	Reference Month/Quarter
	...	
Russia:PMI: Manufacturing	+1	First Following Month
	+2	
	+3	
Russia:PMI: Services Composite	+4	
	+5	
	+6	
Russia: Consumer Confidence Index	+7	
	...	
Russia: Output: Industrial Production Index: Total	+16	
	+17	
	+18	
Russia: Total Real Retail Trade	+19	
Russia: Labor Market: Employment	+19	
	...	Second Following Month
Russia: Exports of Goods	+41	
Russia: Imports of Goods	+41	
	...	
Russia: Gross Domestic Product	+45	

Table 2: Calendar

The HSBC/Markit Purchasing Managers' Index (PMI) indices are derived from continuous monthly surveys of business conditions and track what is actually happening at individual company level. The PMI indices are based on carefully selected panels of executives in companies who report each month on real events. The PMIs used in the Russia model cover manufacturing and service sector activity.

The Consumer Confidence Overall Index is based on a survey of 5,000 people aged 16+ leaving in all regions of the country. The index is an arithmetical average of 5 indices: the change in the respondent's personal financial situation over the last 12 months and next 12 months, the change in the country's economic situation over the last year and in the next 12 months, and the current climate for durable goods purchase. The index shows the difference between the percentage share of persons that are optimistic and the percentage of persons that are pessimistic. It takes a value between -100 (all respondents asses their situation as poor and expect it to become worse) and 100 (all participants are satisfied with the current situation and expect it to improve); 0 indicates neutrality.

4.3 Hard data

As for other countries, we include in the model key series in the five main categories: production, domestic demand, trade, labour and housing.

For production, we track the monthly Index of Production (IoP) which is released by the Federal State Statistics Service (GKS) and which has a high relevance index for Bloomberg (95). This index measures the volume of production at base year prices for the manufacturing, mining and quarrying, electricity, gas and water supply. The IoP provides a timely indicator of growth in the output of production industries, at constant prices. We also include three basic indicators of GDP by industry for the YoY model: Construction, Agriculture and Transportation, which are indicators of the output regarded to each sector.

For domestic demand we consider Total Real Retail Trade. All large and medium outlets (22 000 units) are covered through monthly questionnaires. Small outlets are covered through a quarterly sample survey of about 150 000 outlets, being about 20% of the total number of small outlets. A quarterly sample survey is also carried out on markets and bazaars, which account for about 34% of total retail trade.

We also include Imports and Exports of Goods FOB, which are recorded by the Federal Customs Service of the Russian Federation (FCS) in accordance with the general foreign trade recording system, and goods not registered by the FCS.

For the labour market we use the Employment serie, which data are disseminated on the estimated number of the persons employed in the national economy. The data describe the size of the employed permanent population of the Russian Federation. A sample labor force survey is conducted in all of the regions of Russia once per quarter. The planned sample size is approximately 65,000 people between 15 and 72 years of age, or 0.06% of the population.

5 Results

In this section we evaluate the methodology described before in a recursive out-of-sample exercise to forecast quarter-on-quarter Russia GDP growth. We perform a pseudo-real time exercise, replicating the pattern of data availability at each point in time. We call it "pseudo" because the reconstruction of the vintages use the last revision for all the data.

Consecutive releases of different data revise the forecast and affect the associted forecast uncertainty. We consider updates of next, current and previous quarter forecasts at every new data release, also labelled as "forecasts", "nowcasts" and "backcasts", respectively. The forecasts are produced from 2008-Q1 until 2013-Q4.

As explained before, the sequence of forecasts for the reference quarter is based on "expanding" information sets. The difference between two consecutives forecasts is the sum over all the released variables of the product of the *news* related to a particular variable and the associated weight in the GDP estimate. The parameters are estimated at the beginning of each year, before the first forecast in the sequence is made, and kept

constant for all the subsequent forecast updates throughout the year.

Our main outcome is the QoQ growth rate result for the Russia economy, however there is no other professional forecast institute with whom we could compare ourselves. For this purpose from the QoQ results we can reconstruct annual forecasts so as to compare our performance with other professional forecasters. We use forecasts from the International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), European Comission (EC) and World Bank (WB). As we can see in the figure (8), our model has a good performance, tracking the GDP close from the others forecasts during the whole sample analysed and recurent performing better than the others forecasting professionals.

Also, we compute the results for the YoY growth rate. Different from the annual specification, the YoY results are obtained from not only a forecasting reconstruction, but the DFM underlining the results is different. The variable of interest in the DFM its the yearly growth rate instead of the quartely growth rate, which implies in different restrictions in the model representatition. The YoY model specification and results are found in the appendix.

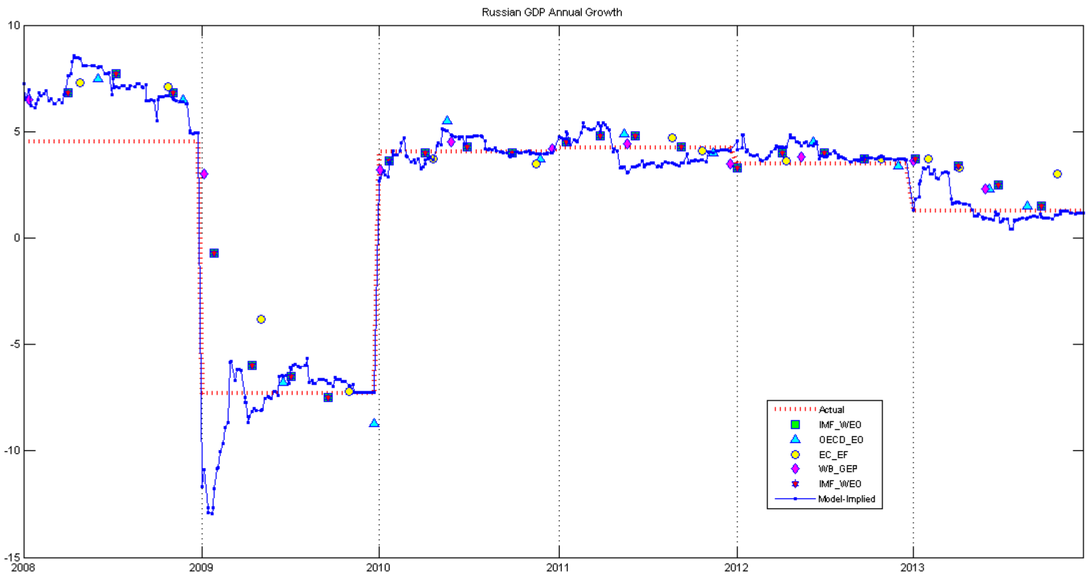


Figure 1: QoQ GDP growth

In figure (2) we compare the performance of the model, on average for all of the calendar quarters in the historic reconstruction period (January 2008 to December 2013), with an autoregressive model forecast, which changes only when GDP is released. From the chart, we can see that our model had the best performance mainly during the 2008/2009 crisis where it could track the deepening GDP drop much better than the autoregressive model.

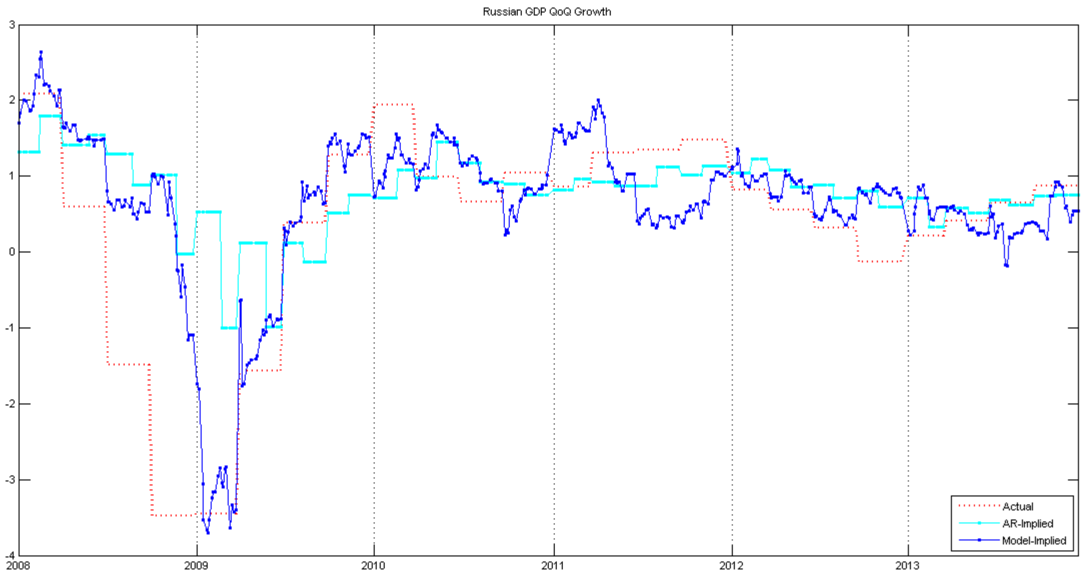


Figure 2: Annual GDP growth

To examine the average accuracy of the model we choose the Root Mean Squared Forecast Error (RMSFE) evaluated over the period 2008-2013. In Figure (3), the X axis reports the days of the prediction period for each calendar quarter. The model's quarterly GDP growth prediction is first made 90 days before the start of a given quarter, and is then updated with each successive data release until the release of the preliminary GDP estimate which for Russia takes place around 135 days after the beginning of the reference quarter. The prediction period is conventionally split into three subsamples which correspond to the 'forecast' (from -90 to -1), the 'nowcast' (from 0 to 90) and the 'backcast' periods (from 91 until GDP release). The Y axis measures the evolution of the Root Mean Squared Forecast Error (RMSFE) in forecasting GDP as new data are released over the prediction period. For comparision we plot the same average uncertainty measure for forecasts produced by the autoregressive univariate model. We can see from Figure 2 that the model's RMSFE overall has a declining pattern over the prediction period which means that new information has a general monotonic and negative effect on uncertainty: it is useful to update the nowcast of GDP with each data release, as the accuracy of the predictions made by the model increases. We can see that the RMSFE is reduced by 50% as we move from the first 'forecast' to the last 'nowcast', illustrating the ability of the moodel to incorporate increasingly richer information as time progress is key for improving now-cast accuracy. Also, we note the difference between the model forecast from the auto regressive forecast becomes more sizeable as more information related to the target period accrues.

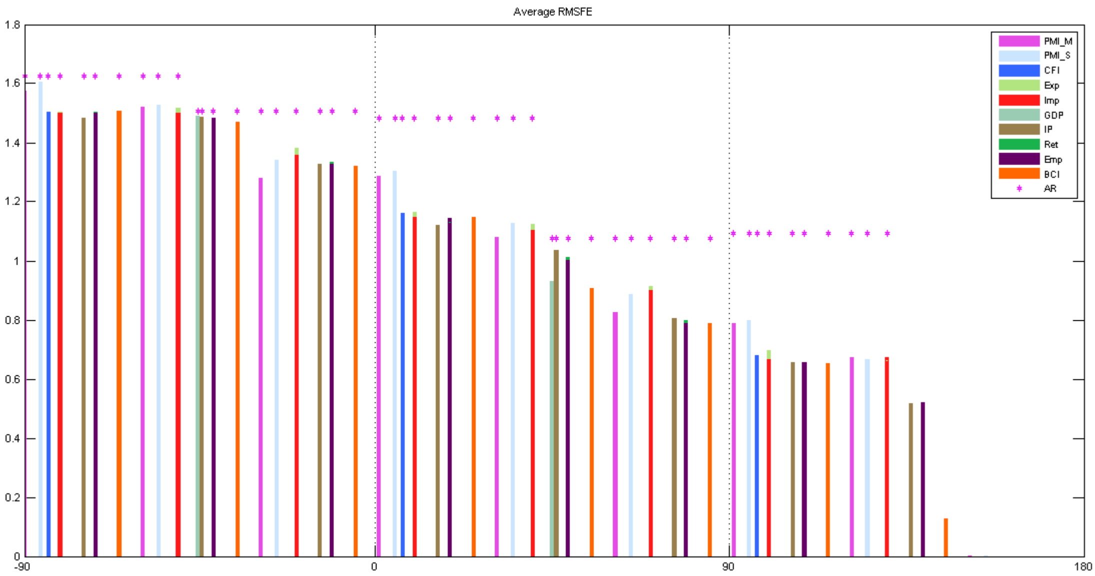


Figure 3: Average Root Mean Square Forecast Error

In table (3), we can identify the average RMSFE reduction caused by the release of each variable. As we could presume, surveys are particularly important in the early part of the calendar. For the first two months of the quarter, the PMI manufacturing is the responsible for the biggest drop in the forecast error, followed by the consumer confidence index, which is released only in the first month of each quarter. Although, in the last month by far the biggest RMSFE reduction comes from the GDP releases, which refers to the previous quarter. In second place comes another "hard" variable, the Industrial Production Index.

Table (4) and figure (4) report the contribution of the news component of the various variables to the nowcast revision. In the first and second month of the quarter PMI manufacturing has the largest impact while in the third month GDP has it. However, PMI manufacturing news maintain its important contribution in the last month, as it has the second largest impact. Also, consumer confidence index has a considerable impact for the nowcast revisions.

	Frequency	Average MSFE reduction		
		m1	**m2**	**m3**
Russia PMI: Manufacturing	M	**-0.17**	**-0.20**	-0.08
Russia PMI: Services Composite	M	-0.03	0.15	0.12
Russia: Consumer Expectations: Consumer Confidence Index (%)	Q	-0.15	0	0
Russia: Output: Industrial Production Index: Total	M	-0.06	-0.13	-0.15
Russia: Total Real Retail Trade	M	0.02	-0.04	-0.02
Russia: Labor Market: Employment	M	0.03	-0.02	-0.02
Russia: Business Confidence: Manufacturing (%)	M	0.00031	-0.00281	0.00070
Russia: Exports of Goods, fob	M	-0.01	-0.005	-0.04
Russia: Imports of Goods, fob	M	-0.04	-0.05	-0.03
Russia: Gross Domestic Product	Q	-0.0017	-0.04	**-1.00**

Table 3: Average MSFE reduction by variable

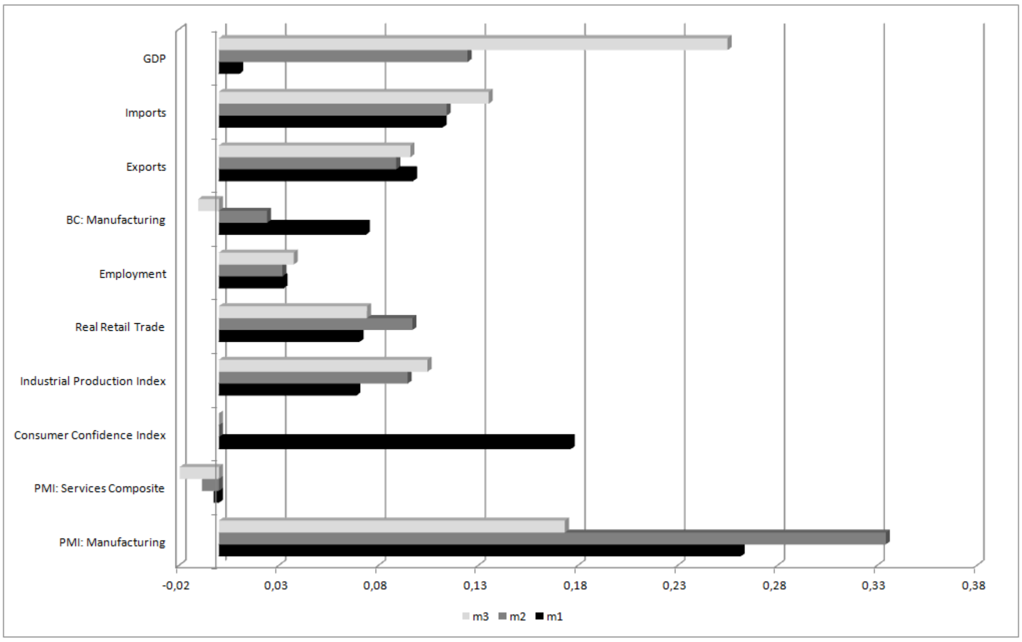


Figure 4: Variables relevance

		A=Average Weight			B=News StD			Average Impact=A*B		
	F	**m1**	**m2**	**m3**	**m1**	**m2**	**m3**	**m1**	**m2**	**m3**
PMI Manufacturing	M	0.25	0.19	0.09	1.03	1.74	1.94	**0.26**	**0.33**	0.17
PMI Services	M	-0.0014	-0.0030	-0.00631	1.83	2.86	3.13	-0.0026	-0.0085	-0.02
Consumer Confidence Index (%)	Q	0.018	0	0	9.91			0.18		
Industrial Production Index	M	0.05	0.06	0.05	1.27	1.70	2.30	0.07	0.10	0.10
Real Retail Trade	M	0.11	0.10	0.09	0.64	0.97	0.87	0.07	0.10	0.07
Employment	M	0.06	0.05	0.05	0.55	0.59	0.72	0.03	0.03	0.04
BC Manufacturing	M	0.04	0.0093	-0.0042	2.08	2.59	2.43	0.07	0.02	-0.01
Exports of Goods	M	0.01	0.02	0.01	6.84	5.80	7.00	0.10	0.09	0.10
Imports of Goods	M	0.02	0.03	0.03	4.53	4.11	5.20	0.11	0.11	0.14
Gross Domestic Product	Q	0.01	0.26	0.24	0.71	0.48	1.08	0.01	0.12	**0.26**

Table 4: Impact of the Releases on the Nowcast

6 News Index

In this section is proposed a methodology to construct pseudo real time indexes and then test its predictive power over commodity prices, in our example gas price. Indexes are useful to give a reading about the actual state of the economy, aggregating a large amount of economic indicators released at different times and at different frequency. Our news indexes summarize recent economic data surprises measuring the deviation of official data releases from the expectations generated by the Dynamic Factor Model described before. The indexes are a weighted average of the surprises where the weights depend on the contribution of the associated economic indicator to the forecast of GDP in the Nowcasting model. The difference among the indexes are in the size of the rolling window used: one month, two and three months.

Altavilla, Giannone and Modugno (2014) also identify innovations in macroeconomic fundamentals based on macroeconomic news, although their definition of news is the

difference between the actual macroeconomic releases and the median market predictions. They analyze the reaction of the U.S. Treasury bond market to those innovations in macroeconomic fundamentals.

The predictive power of the indexes are tested through in-sample and out-of-sample analysis. For the in-sample exercise we estimate the relationship and the predictive power of the indexes for one, two and three steps ahead. For the out-of-sample exercise we compare the forecasting results among different models, computing the mean squared forecast error (MSFE) so as to measure the forecast accuracy of each model.

6.1 Methodology

The rolling window indexes are defined as the sum of impacts of the news on the nowcast of GDP. We define impacts as the weights of the Nowcasting model $b_{j,t}$ multiplied by the Nowcasting news $\eta_{j,t}$. The indexes have windows of one, two and three months. The monthly index is constructed as follows:

$$\sum_{\mathbb{J}_{v+1}\in M_t} \sum_{j\in\mathbb{J}_{v+1}} b_{j,v+1} \left(x_{j,T_j,v+1} - \mathbb{E}x_{j,T_{j,v+1}|\Omega_v}\right) \tag{18}$$

the two months index is:

$$\sum_{\mathbb{J}_{v+1}\in M_{t-1:t}} \sum_{j\in\mathbb{J}_{v+1}} b_{j,v+1} \left(x_{j,T_j,v+1} - \mathbb{E}x_{j,T_{j,v+1}|\Omega_v}\right) \tag{19}$$

the quarterly index is:

$$\sum_{\mathbb{J}_{v+1}\in M_{t-2:t}} \sum_{j\in\mathbb{J}_{v+1},} b_{j,v+1} \left(x_{j,T_j,v+1} - \mathbb{E}x_{j,T_{j,v+1}|\Omega_v}\right) \tag{20}$$

We can see the monthly index is the sum of impacts within the currently month, whereas the two months index and quarterly index are the sum over the last two and three months, respectively. Index starts from zero and evolve moving the window on month by month. We register the macroeconomic news and subsequentlty the impacts from July 2002 to December 2013.

6.2 Data

In our exercise we use a monthly price data for the Liquified Natural Gas (LNG) in U.S. downloaded from the Global Financial Data database. The price series is in US dollars and was deflated using the consumer price index for the U.S. Then log prices were computed in order to obtain the growth rate for the price serie. The indexes are computed with the 10 macroeconomic variables showed previously.

Our choice for this LNG price series its related to the longest data serie available for the U.S. market and for some LNG features. The development of an international market for liquefied natural gas (LNG) and the resulting opportunities for intercontinental arbitrage are seen as creating a world in which movements in natural gas prices are linked between continents. Increased flows of LNG into the United States and the potential sensitivity of these shipments to price differentials between Europe and North America suggests the possibility of a strengthening relationship between natural gas prices on these two continents. At the same time , there is considerable evidence linking natural gas price movements in Europe and North America to those for crude oil. (See Brown and Yücel (2009)).

6.3 In-sample results

We start analysing the relationship between the indexes and the LNG serie graphically as showed in the figure (5). We plot the three indexes against the LNG price growth rate to see if there is any correlation between our indexes and the performance of the LNG price. We normalize the series in order to have comparable scales in the chart.

Figures (5) shows a comovement between the indexes and the LNG variations. For instance, both series share a drop during the financial crisis which took place at the end of 2008 and beginning of 2009, followed by a recovery pattern afterwards. We still notice a similar pattern out of this period and even more, we can appreciate an antecipating component of the indexes series regarding the LNG behavior, which can lead us to believe that there is a possible explicatory power of the indexes over the LNG serie.

Besides the graphically analysis, we run some regressions using the monthly, two months and quarterly indexes and the log price difference for the LNG. For robustness regarding autocorrelation and heteroskedasticity we use the Newey West estimator of the covariance matrix. For our in-sample exercise we use the following regression:

$$\Delta p_{t+h} = \beta_1 + \beta_2 \Delta p_t + \beta_3 Index_t + \epsilon_t \tag{21}$$

where Δp_{t+h} is the log difference price of the LNG h steps ahead, h being equal to 1,2 and 3 months. The sample period is from July 2002 to December 2013. The regression is run at monthly frequency and we compare the standardized coefficients ($t_\beta = \hat{\beta}/s.e.(\hat{\beta})$) with an infinite number of degrees of freedom and we test the significance of the coefficients in the regression at 10%, 5% and 1% confidence level.

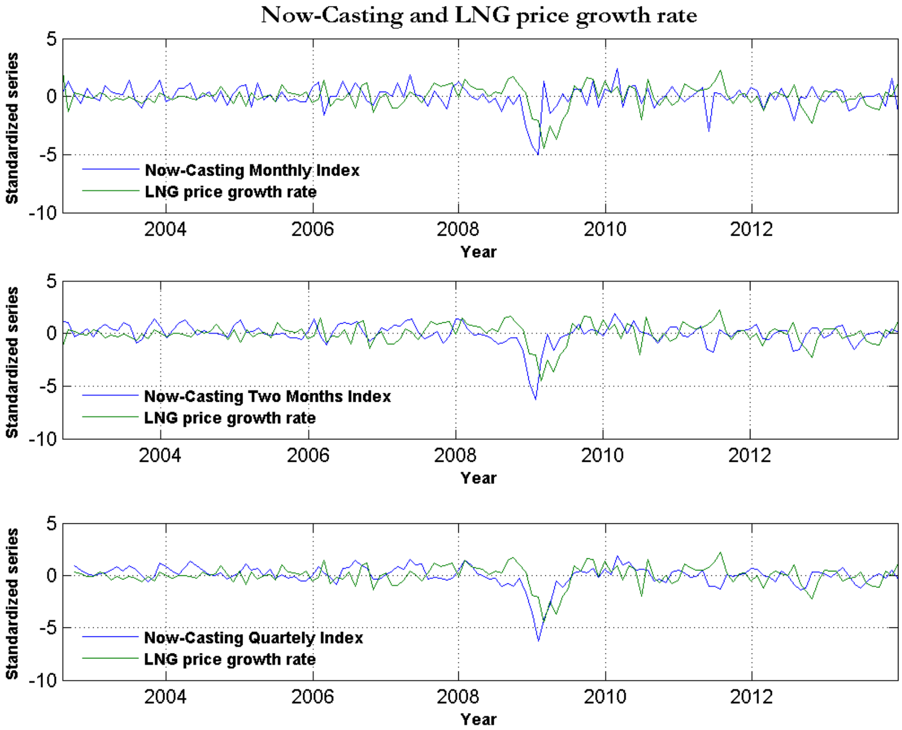


Figure 5: Nowcasting index and LNG price

From the table (5) the coefficient β_3 for the monthly index is 10% significant for 1 and 2 steps ahead, while 1% significant for 3 steps ahead. Results are even stronger for the two months and quarterly index, which coefficients are 1% significant for all the cases, except for 1 step ahead when we use the two months index. The signs of the coefficients are positive, what means that a positive surprise in the economy for the last period (one month, two months, quarter) relative to the value forecasted of the Nowcasting model, weighted by the nowcasting weights, has a positive correlation with the prices of the LNG. Regarging the current LNG price coefficient β_2, its significant at a 1% confidence level for 1 and 2 steps ahead, although its not significant when used for forecasting 3 months ahead. The R-squared has values between 17,28% and 33,42%, with the quarterly index model showing the highest values. It is worthwhile to underline these numbers are not low when compared with the typical results in the literature.

The presented in-sample exercise is useful to investigate if there is some relationship among the indexes and the LNG price, indicating that the indexes can have some explanatory power over the LNG price. However, an out-of-sample exercise provides best information about causation and forecast ability.

	1 step ahead			2 steps ahead		
	Monthly Index	Two Months Index	Quarterly Index	Monthly Index	Two Months Index	Quarterly Index
β_1	0.0036	0.0039	0.0034	0,0046	0,0042	0,0041
t-statistic	(0,84)	(1,08)	(1,03)	(0,96)	(1,02)	(0,98)
β_2	0,3919***	0,3930***	0,3387***	0,3244***	0,2682***	0,2301***
t-statistic	(2,89)	(3,73)	(3,35)	(2,61)	((2,97)	(2,94)
β_3	0,0300*	0,0253**	0,0270***	0,034*	0,0356***	0,0288***
t-statistic	(1,43)	(2,05)	(2,99)	(1,63)	(3,15)	(3,41)
R^2	0,2249	0,2788	0,3342	0,1855	0,2640	0,2665

	3 steps ahead		
	Monthly Index	Two Months Index	Quarterly Index
β_1	0,0053	0,0052	0,0052
t-statistic	(1,06)	(1,10)	(1,08)
β_2	0,1525	0,1055	0,0527
t-statistic	(1,68)	(1,11)	(0,46)
β_3	0,0520***	0,0371***	0,0317***
t-statistic	(2,57)	(2,85)	(2,94)
R^2	0,1728	0,1845	0,2008

Table 5: In-sample results

6.4 Out-of-sample analysis

In this section we perform an out-of-sample exercise using different windows in order to cope with the trade-off between the length of the estimation and the out-of-sample period : out-of-sample windows of 3,4,5 and 6 years. Also, since the great recession (2008-2009) has had a huge influence on asset prices and on the prices of commodities, we considered forecast samples which included the crisis period and not. The following models are used to forecast 1,2 and 3 steps ahead:

- **Random Walk:** $\Delta p_{t+h} = \mu + \epsilon_t$
- **AR(1):** $\Delta p_{t+h} = \beta_1 + \beta_2 \Delta p_t + \epsilon_t$
- **AR(1) + Index:** $\Delta p_{t+h} = \beta_1 + \beta_2 \Delta p_t + \beta_3 Index_t + \epsilon_t$
- **AR(2):** $\Delta p_{t+h} = \beta_1 + \beta_2 \Delta p_t + \beta_3 \Delta p_{t-1} + \epsilon_t$
- **AR(2) + Index:** $\Delta p_{t+h} = \beta_1 + \beta_2 \Delta p_t + \beta_3 \Delta p_{t-1} + \beta_4 Index_t + \epsilon_t$

So we compare the most common models used as benchmark in the literature, autoregressive and random walk models, with autoregressive models added with the index, so as we can investigate if the index has some additional explanatory power to forecasting the gas price. In order to be clear how the exercise develops, for instance, let´s consider the situation which the out-of-sample window is from January 2008 until December 2013. So in the first loop the estimation sample starts on July 2002 until December 2007, using the parameters estimated of that period to forecast one step ahead (January 2008), two steps ahead (February 2008) and three steps ahead (March 2008). Then we move our date one month on and we start again the process of estimation and forecasting.

The forecast accuracy of all models are compared to that of the random walk with drift. We use the ratio of the root-mean-squared-forecast error for each model vis-a-vis that of the randon walk with drift– our benchmark:

$$RRMSFE_h^M = \frac{RMSFE_h^M}{RMSFE_h^{RW}}, \tag{22}$$

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

We want to be able to distinguish the forecast accuracy of the competitor models,asking whether their accuracy measures are statistically equal or not. We do this using the Clark and West (2007) test, comparing each model M with the random walk with drift. We have forecast results for four different out-of-sample windows, using the

same dataset, from July 2002 until December 2013. The results are displayed in the tables below.

The monthly index models had the best performance 50% of the time, 33,33% of the time AR(1)+Index and 16,67% AR(2)+Index. AR(1)+Index on average performed 1,65% better than the random walk model, whether we keep our analysis in the short-term (1 step ahead) this number increases to 11,39%. Many of the results we found out were statistically significant when the Clark and West(2007) test was employed.

Regarding the two months index models, they performed really well regardless of the forecasting sample. It has beaten the models without index (random walk and pure autoregressive) 83,33% of the time, 50% of the time AR(1)+Index was the best model and 33,33% AR(2)+Index was the best one. Averagin the RMSFE results, AR(1)+Index performed 9,01% better than the random walk model, whether we keep our analysis in the short-term (1 step ahead) this number increases to 13,19%.

Finally, the quarterly index had the best performance when compared with the other two indexes mentioned above. The quarterly index models had the lowest RMSFE 100% of time, 58,33% for the AR(1)+Index and 41,67% for the AR(2)+Index. The AR(1)+Index on average performed 14.29% better than the random walk while keeping the analysis for one step ahead the number goes to 23,88%. Again, most of the results were statistically significant. We can find these results in the tables below. The other indexes tables are found in the appendix.

Out-of-Sample length: Jan 2008-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,781***	0,942*	1,065*
AR(1) + Q Index	0,667*	**0,864***	**0,950***
AR(2)	0,770*	0,984*	1,144
AR(2) + Q Index	**0,651***	0,861*	0,979*

Out-of-Sample length: Jan 2009-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,753**	0,9117*	1,036
AR(1) + Q Index	0,631*	0,837*	0,881*
AR(2)	0,736*	0,934*	1,09
AR(2) + Q Index	**0,609***	**0,824***	**0,879***

Out-of-Sample length: Jan 2010-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	1,069	1,14	1,07
AR(1) + Q Index	0,973	**0,951***	**0,977****
AR(2)	1,072	1,150	1,094
AR(2) + Q Index	**0,970**	0,960*	1,009***

Out-of-Sample length: Jan 2011-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,801**	1,076	1,133
AR(1) + Q Index	**0,774***	**0,859***	**0,922****
AR(2)	0,915**	1,115	1,161
AR(2) + Q Index	0,8443*	0,875*	0,979**

*,** and *** mean 10%,5% and 1% of significance respectively (Clark-West test (2007))

Wrapping up results across all horizons and forecasting samples the AR(1)+Index, mainly the quarterly index, had the overall best performance. Perhaps the best result for

the quarterly index vis-a-vis the others relies on the low frequency nature of macroeconomic variables performance. More, given that econometric models are mainly buit for short term forecasts, the results are more striking as considering the one month ahead forecast the accuracy difference from the models without index is even bigger.

From the demand side, one possible reason for the results above mentioned comes from the fact that as the second largest consumer, when russian weaken growth read from the news index, that causes a drop in the gas price. From the supply side, given the heavily importance of the natural gas for the russian economy, some economic indicators are forward-looking and embodies information about future movements in the commodity markets.

7 Conclusion

This paper has two main motivations. The first is to use the dynamic factor model approach for nowcasting russian GDP, studying its methodologial advantages and how it is your forecasting performance compared to professional forecasters and an autoregressive model(AR). Our findings suggests that overall the DFM model are successful in competing wuith all the competitors analysed.

A second finding from our results, regards to the importance of update the nowcast for each new release. We can see that the RMSFE is reduced by 50% as we move from the first estimation of the GDP to the last one, illustrating the ability of the model to incorporate increasingly richer information available. Also, we can identify the importance of including surveys data in the model, as they are particurlarly important in reducing the RMSFE in the early part of the calenadar.

Given that Russia is the world's largest producer of oil and natural gas (on a par with Saudia Arabia and USA), our second motivation relates to study the effects of innovations in the russian macroeconomic fundamentals on commodity price predictability. We identify these innovations through a news index which summarizes deviations of official data releases from the expectations generated by the DFM.

We perform an exercise comparing common models used as benchmark in the literature, autoregressive and random walk models, with autoregressive models added with the index to forecast liquified natural gas (LNG) price. Our results suggest that macroeconomic news can have sizeable low frequency effects on commodity prices. Although Russia is a big country and its economic dependency of the commodity market is well known, possibly the importance of macroeconomic news for commodity price in underestimated in our work. Innovations in the fundamentals of other countries could also be important for LNG and other commodities prices.

References

Altavilla, C., Giannone, D. and Modugno, M. (2014): "Low Frequency Effects of Macroeconomic News on Government Bond Yields," Working Papers ECARES ECARES 2014-34, ULB – Universite Libre de Bruxelles.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. and Rünstler, G. (2008): "Short-Term Forecasts of Euro Area GDP Growth," Working Papers ECARES ECARES 2008-035, ULB – Universite Libre de Bruxelles.

Antipa, P., Barhoumi, K., Brunhes-Lesage, V. and Darné, O. (2012): "Nowcasting German GDP: A comparison of bridge and factor models," *Journal of Policy Modeling*, 34(6), 864–878.

Bańbura, M., Giannone, D. and Reichlin, L. (2010): "Nowcasting," Discussion paper.

Bańbura, M. and Modugno, M. (2014): "Maximum Likelihood Estimation Of Factor Models On Datasets With Arbitrary Pattern Of Missing Data," *Journal of Applied Econometrics*, 29(1), 133–160.

Bańbura, M. and Rünstler, G. (2011): "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting {GDP}," *International Journal of Forecasting*, 27(2), 333 – 346.

Borensztein, E. and Reinhart, C. M. (1994): "The macroeconomic determinants of commodity prices," *Staff Papers-International Monetary Fund*, pp. 236–261.

Chen, S.-S. (2014): "Forecasting Crude Oil Price Movements with Oil-Sensitive Stocks," *Economic Inquiry*, 52(2), 830–844.

Chen, Y.-C., Rogoff, K. S. and Rossi, B. (2010): "Can Exchange Rates Forecast Commodity Prices?," *The Quarterly Journal of Economics*, 125(3), 1145–1194.

Clark, T. E. and West, K. D. (2007): "Approximately normal tests for equal predictive accuracy in nested models," *Journal of econometrics*, 138(1), 291–311.

Clements, M. P. and Galvão, A. B. (2008): "Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States," *Journal of Business & Economic Statistics*, 26(4), 546–554.

Doz, C., Giannone, D. and Reichlin, L. (2012): "A Quasi Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94(4), 1014–1024.

GARGANO, A. and TIMMERMANN, A. (2014): "Forecasting commodity price indexes using macroeconomic and financial predictors," *International Journal of Forecasting*, 30(3), 825–843.

GHYSELS, E., SANTA-CLARA, P. and VALKANOV, R. (2004): "The MIDAS touch: Mixed data sampling regression models," *Finance*.

GIANNONE, D., REICHLIN, L. and SMALL, D. (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55(4), 665–676.

GOLINELLI, R. and PARIGI, G. (2007): "The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries," *Journal of Forecasting*, 26(2), 77–94.

MARCELLINO, M. and SCHUMACHER, C. (2008): "Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP," CEPR Discussion Papers 6708, C.E.P.R. Discussion Papers.

——— (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP*," *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550.

MARIANO, R. S. and MURASAWA, Y. (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18(4), 427–443.

PARIGI, G. and SCHLITZER, G. (1995): "Quarterly forecasts of the Italian business cycle by means of monthly economic indicators," *Journal of Forecasting*, 14(2), 117–141.

RAUTAVA, J. (2004): "The role of oil prices and the real exchange rate in Russia's economy-a cointegration approach," *Journal of comparative economics*, 32(2), 315–327.

TREHAN, B. (1989): "Forecasting growth in current quarter real GNP," *Economic Review*, (Win), 39–52.

A Appendix

A.1 The YoY model results

In the same way we did for the QoQ model, quartely variables are introduced by constructing for each of them a partially observed monthly counterpart. Let y_t^{MY} denote the unobserved monthly YoY rate which admits the same factor structure as the monthly variables:

$$\begin{aligned} y_t^{MY} &= \mu_Q + \Lambda_Q f_t + \epsilon_t^Q & (23) \\ \epsilon_t^Q &= \alpha_Q \epsilon_{t-1}^Q + e_t^Q, \quad e_t^Q \sim N(0, \sigma_Q^2) & (24) \end{aligned}$$

However, the link between monthly unobserved and partially observed quartely GDP is different:

$$\begin{aligned} y_t^{QY} &= Y_t^Q - Y_{t-12}^Q = (1 - L^{12})Y_t^Q \\ &\approx (1 - L^{12})(1 + L + L^2)Y_t^M \\ &= (1 + L + L^2)y_t^{MY} = y_t^{MY} + y_{t-1}^{MY} + y_{t-2}^{MY} \end{aligned}$$

For the yearly growth model we use a larger dataset as we can include some non-seasonally adjusted data. We include three output indexes: Construction, Agriculture and Transportation. The results are showed below.

Release name	**F**	**S**	**SD**	**Units/Transf**
PMI: Manufacturing	M	HSBC/MKT	1997	Index/levels
PMI: Services Composite	M	HSBC/MKT	2001	Index/levels
Output: Industrial Production Index: Total	M	GKS/H	1995	MoM %/MoM
Exports of Goods	M	CBRF/H	1995	MoM %/MoM
Imports of Goods	M	CBRF/H	1995	MoM %/MoM
Total Real Retail Trade [Sales]	M	GKS/H	1995	MoM %/MoM
Labor Market: Employment	M	GKS/H	1996	MoM %/MoM
Business Confidence: Manufacturing	M	GKS	2005	Index/levels
Output: Agriculture	M	GKR	2000	YoY %/YoY
Output: Construction	M	GKS	1999	YoY %/YoY
Output: Transportation	M	GKR	2000	YoY %/YoY
Consumer Confidence Index	Q	GKS	2006	Index/diff
Gross Domestic Product	Q	GKS	1995	QoQ %/QoQ

Table 6: YoY model dataset

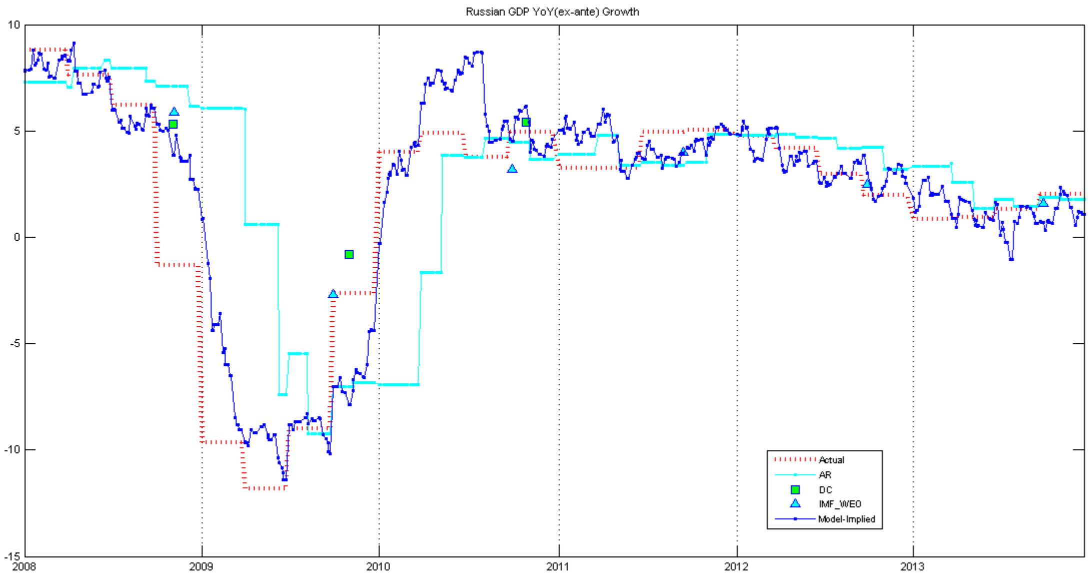


Figure 6: YoY GDP growth

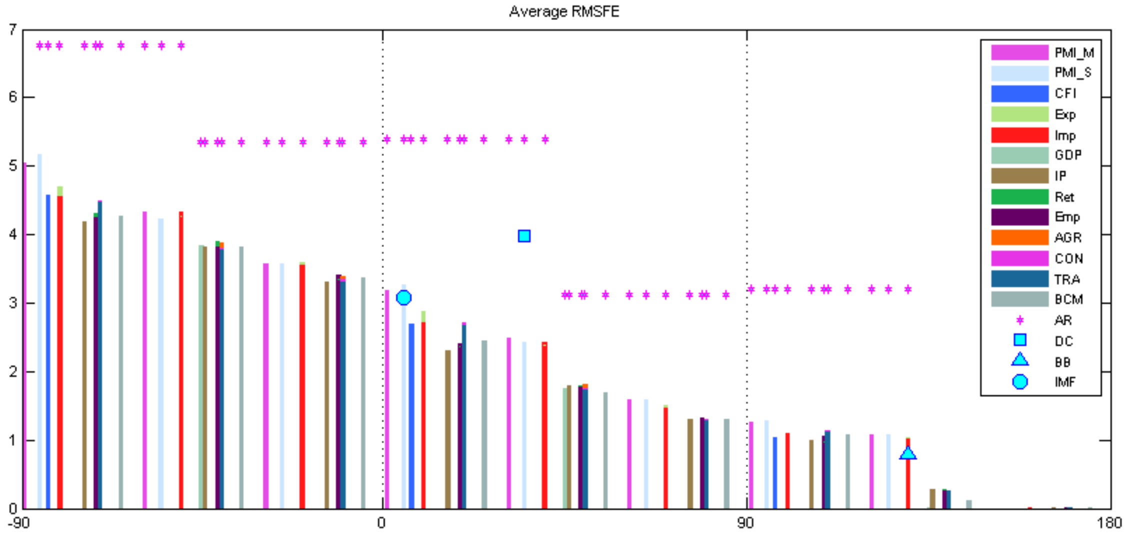


Figure 7: Average Root Mean Square Forecast Error for YoY model

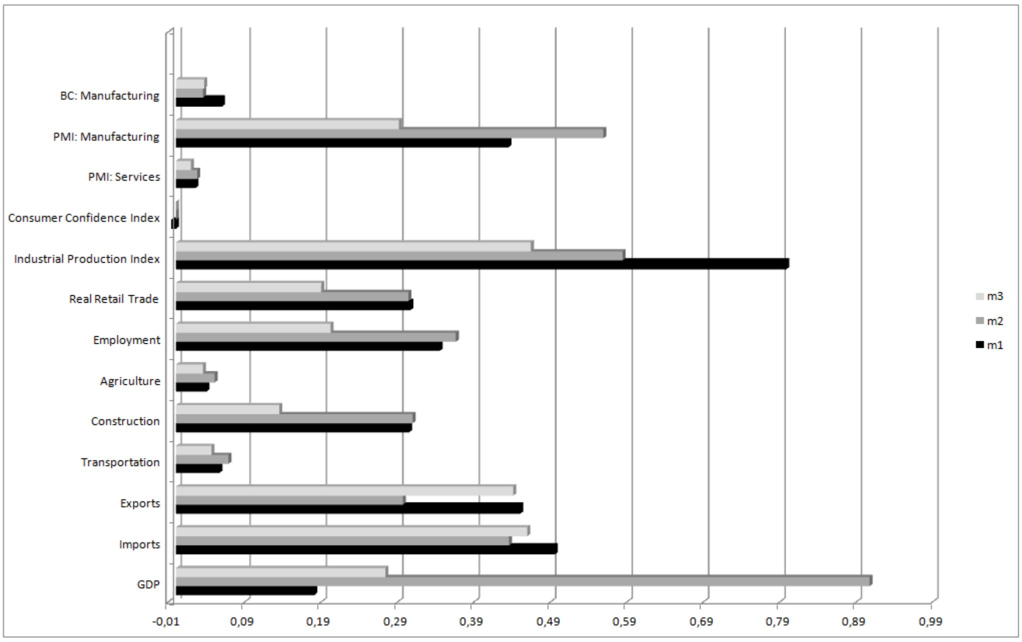


Figure 8: Variables relevance for YoY model

B Remaining News Index Results

In this section we present the out-of-sample results for the monthly and two months index.

B.0.1 Monthly Index

Out-of-Sample length: Jan 2008-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,778**	**0,936***	1,067
AR(1) + M Index	0,762*	1,001*	**1,059**
AR(2)	0,768*	0,984*	1,142
AR(2) + M Index	**0,728***	1,017*	1,120

Out-of-Sample length: Jan 2009-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,767*	**0,9124***	1,039
AR(1) + M Index	0,7465*	0,984*	**0,997****
AR(2)	0,739*	0,932*	1,085
AR(2) + M Index	0,706*	0,986*	1,038*

Out-of-Sample length: Jan 2010-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	**1,064**	**1,139**	1,067
AR(1) + M Index	1,199	1,242	**0,866*****
AR(2)	1,069	1,148	1,085
AR(2) + M Index	1,208	1,253	0,891***

Out-of-Sample length: Jan 2011-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	**0,806****	**1,075**	1,132
AR(1) + M Index	0,837**	1,180	**0,928****
AR(2)	0,916**	1,120	1,164
AR(2) + M Index	0,973**	1,242	0,950**

*,** and *** mean 10%,5% and 1% of significance respectively (Clark-West test (2007))

B.0.2 Two months Index

Out-of-Sample length: Jan 2008-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,781**	0,938*	1,065
AR(1) + TM Index	0,731*	**0,916***	**1,008**
AR(2)	0,773*	0,985*	1,144
AR(2) + TM Index	**0,710***	0,926*	1,069

Out-of-Sample length: Jan 2009-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	0,756**	0,913*	1,036
AR(1) + TM Index	0,709*	0,887*	**0,970****
AR(2)	0,734*	0,933*	1,086
AR(2) + TM Index	**0,676***	**0,857***	1,016**

Out-of-Sample length: Jan 2010-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	1,067	1,14	1,07
AR(1) + TM Index	1,139	**0,978***	**0,885*****
AR(2)	**1,075**	1,152	1,092
AR(2) + TM Index	1,175	0,983*	0,887

Out-of-Sample length: Jan 2011-Dec 2013			
RMSFE relative to RW	Steps ahead		
	1	2	3
AR(1)	**0,806****	1,076	1,132**
AR(1) + TM Index	0,894*	**0,981**	**0,822****
AR(2)	0,916**	1,117	1,163
AR(2) + TM Index	0,998	1,018	0,859

*,** and *** mean 10%,5% and 1% of significance respectively (Clark-West test (2007))

International Journal of Forecasting 31 (2015) 862–875

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting multivariate time series under present-value model short- and long-run co-movement restrictions

Osmani Teixeira Guillén [a], Alain Hecq [b], João Victor Issler [c,*], Diogo Saraiva [c]

[a] Banco Central do Brazil and IBMEC-RJ, Brazil
[b] Maastricht University, Netherlands
[c] Getulio Vargas Foundation, Brazil

ARTICLE INFO

Keywords:
Forecasting
Multivariate models
Vector autoregression (VAR)
Present-value restrictions
Common cycles
Cointegration
Interest rates
Prices and dividends

ABSTRACT

Using a sequence of VAR-based nested multivariate models, we discuss the different layers of restrictions that are imposed on the VAR in levels by present-value models (PVM hereafter) 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 & Engle, 1993, 1997) and their implications for forecasting.

Using a well-known database, maintained 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 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.

1. Introduction

The use of multivariate time series models in economics and finance has proved fruitful, since they contain 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 the 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 & Granger, 1987; Johansen, 1988). One can also impose further commonalities in their short-run dynamics, e.g., impose common cyclical feature restrictions (see Engle & Kozicki, 1993; Vahid & 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 & Hendry, 1998, or Hoffman & Rasche, 1996, *inter alia*). However, only a handful of papers (e.g., Anderson & Vahid, 2011; Issler & Vahid, 2001; Vahid & Issler, 2002) have investigated whether including additional short-run co-movement restrictions generates better forecasts. Moreover, Athanasopoulos, Guillen, Issler, and Vahid (2011) only recently compared the relative im-

* Correspondence to: Graduate School of Economics - EPGE, Getulio Vargas Foundation, Praia de Botafogo 190 s. 1100, Rio de Janeiro, RJ 22250-900, Brazil.
E-mail address: Joao.Issler@fgv.br (J.V. Issler).

http://dx.doi.org/10.1016/j.ijforecast.2015.02.002
0169-2070/

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

Both 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 Baillie (1989) and Hansen and Sargent (1981, 1993) for initial work on rational expectations linked to PVMs, and those by Johansen (2000) and Johansen and Swensen (1999, 2004, 2011) for a recent fresh look at the subject.

Indeed, PVMs arise from a first-order stochastic difference equation, where its error term must be unforecastable with regard to 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 Athanasopoulos et al. (2011) and Hecq, Palm, and Urbain (2006), for the vector error-correction model (VECM) for Y_t and y_t. Alternatively, they also imply 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 to 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 US economy. The second contains real prices 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 last 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 are compared across representations 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 one to five years ahead. Overall, for Y_t and y_t, they produce winners 63.33% of the time at these 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 the joint estimation of the long-run and short-run parameters, as per Centoni, Cubadda, and Hecq (2007). In order to avoid taking up too much space in a forecasting paper with testing and estimation issues, these are discussed in Appendix A.

The rest of the paper is arranged as follows. Section 2 reviews PV formulas (for both the levels and log-levels of the variables) and discusses the types of restrictions that 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, to verify whether the restrictions implied by economic theory hold in practice. In Section 4, we compare the forecasting gains obtained in multivariate models by imposing different types of PV restrictions. Section 5 concludes. Appendix A 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 performances, and other issues that are relevant when examining PVM restrictions. We also present an online appendix (see Appendix B) including only 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 we drop the constant term for

[1] Examples 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 be cointegrated. See also the examples from Campbell (1987) and Campbell and Deaton (1989), *inter alia*, which are reviewed by Engsted (2002); together with the interesting recent contribution by Johansen and Swensen (2011).

simplicity of presentation. 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 time 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*, Campbell & Shiller, 1987; Chow, 1984; 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:

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

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

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

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:

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

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 for $\binom{Y_t}{y_t}$ in multivariate models. 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 bivariate VAR of order p on these series takes the form:

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

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

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

where the Γ_js 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 implies $-\left(I - \sum_{i=1}^{p} \Gamma_i^*\right)\binom{Y_t}{y_t} = \binom{\alpha_1}{\alpha_2}(1 : -\theta)\binom{Y_t}{y_t}$.

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 $\Gamma_1, \Gamma_2, \ldots, \Gamma_{p-1}$ in the VECM (Eq. (5)).

To see this, move Eq. (1) forward 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 $S_t = S'_t = S''_t$:

$$\begin{aligned} S''_t \frac{1-\delta}{\delta} &= \mathbb{E}_t \Delta Y_{t+1} = \delta \sum_{i=0}^{\infty} \delta^{i-1} \mathbb{E}_t \Delta y_{t+i+1} \\ &= \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i} = S_t \frac{1-\delta}{\delta}, \end{aligned} \tag{6}$$

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 & Shiller, 1987*, *and Johansen & 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 Athanasopoulos et al., 2011, and Hecq et al., 2006), and there exists a* 1×2 *vector* γ' *such that* $\gamma'\Gamma_1 = \gamma'\Gamma_2 = \cdots = \gamma'\Gamma_{p-1} = 0$, *but* $\gamma'\binom{\alpha_1}{\alpha_2} \neq 0$. *Moreover,* $\gamma' = (1 : 0)$, *the first row of every* Γ_i, $i = 1, \ldots, p-1$, *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) are 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 Eq. (4) to Eq. (5), and then to the restricted-VECM in Proposition 1, shows a sequence of nested models, eventually including all of the restrictions entailed by PVMs.

An alternative route for nesting models is to consider the cross-equation restrictions studied by Campbell and Shiller (1987) using a transformed VAR on $\binom{S_t}{\Delta y_t}$; see Johansen and Swensen (2011). As a motivation for 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:

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

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

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

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

As can be seen readily, Eq. (8) is a stochastic difference equation. When solved forward in expectations, with an added transversality condition, it generates Eq. (3). Note that, by construction, $\mathbb{E}_t(\varepsilon_{t+i}) = 0,\ i = 1, 2, \ldots,$ 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 we go 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:

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

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 $\underline{\Gamma}(L) = \begin{pmatrix} \underline{\Gamma_{11}}(L) & \underline{\Gamma_{12}}(L) \\ \underline{\Gamma_{21}}(L) & \underline{\Gamma_{22}}(L) \end{pmatrix} = \underline{\Gamma}_1 + \underline{\Gamma}_2 L + \cdots + \underline{\Gamma}_p L^{p-1}$in Eq. (9), as the following proposition asserts.

Proposition 2 (*Campbell & Shiller, 1987, and Johansen & 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 & Hecq, 2001*) for the transformed VAR in Eq.* (9)*: there exists a vector* $\tilde{\gamma}_0'$ *such that* $\tilde{\gamma}_0'\underline{\Gamma}_2 = \cdots = \tilde{\gamma}_0'\underline{\Gamma}_p = 0$, *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 in going from Eq. (8) to Eq. (3): cointegration delivers the transversality condition $\lim_{k\to\infty}\delta^k\mathbb{E}_t(S_{t+k}) = 0$, whereas the unforecastability of ε_t regarding the past, i.e., $\mathbb{E}_t\left(\varepsilon_{t+j}\right) = 0$, 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 to 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 next. 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 the algebra used before to $Y_t = \sum_{i=1}^{\infty}\delta^i\mathbb{E}_t y_{t+i}$ to obtain the following expressions:

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

where $\gamma' = (1:1)$ and

$$\alpha_1 = \frac{1-\delta}{\delta} \quad \text{in Proposition 1.} \tag{11}$$

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

where $\tilde{\gamma}_0' = \left(1 : \dfrac{1}{1-\delta}\right)$ and

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

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 the way in which 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

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

restrictive assumption: that the expected return of a given stock is constant over time,

$$\mathbb{E}_t\left[R_{t+1}\right] = R, \tag{14}$$

where

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

with P_t and D_t denoting the price and the dividend of a given stock, respectively.

In two subsequent papers, Campbell and Shiller (1988a,b) developed an alternative representation for prices and dividends in which Eq. (14) need not hold, and which is therefore consistent with the idea of time-varying returns. This alternative representation uses the logarithms of prices and dividends:

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

arriving at:

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

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:

$$\begin{aligned} h_{t+1} &\approx k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t && (18)\\ &= k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t + d_t - d_t && (19)\\ &= k + \rho\left(p_{t+1} - d_{t+1}\right) + \Delta d_{t+1} + \left(d_t - p_t\right), && (20) \end{aligned}$$

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 $k \equiv -\log(\rho) - (1-\rho)\log(1/\rho - 1)$.

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

$$\begin{aligned} (d_t - p_t) &= -k + h_{t+1} - \Delta d_{t+1} \\ &\quad - \rho\left(p_{t+1} - d_{t+1}\right) + \varepsilon_{t+1}, \end{aligned} \tag{21}$$

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

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

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

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

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

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

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

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 $X_t = (p_t, d_t, r_t)'$. Note 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 $I(0)$, Eq. (24) indicates that $d_t - p_t$ is $I(0)$ as well, yielding the second cointegrating vector in the system.

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

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

Disregarding an irrelevant constant term, the PVM implies that $\mathbb{E}_{t-1}\left[h_t - r_t\right] = 0$. 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 $\left(\rho \quad 0 \quad -1\right)$, in order to obtain an expression for $h_t - r_t$, we must have:

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

which is equivalent to

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

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

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

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

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

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

$$\gamma_4 = \frac{\gamma_6 + 1}{\rho}. \tag{31}$$

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

			Cointegration			Common cycle		
(1) System	(2) Ends	(3) VAR(p)	(4) H_0	(5) $cv_{5\%}$	(6) $\hat{\theta}$	(7) ξ_{LR}	(8) J_2	(9) $cv_{5\%}$
(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	$r = 0: 78.74$ $r = 1: 14.58$	29.79 15.49	1.764	∃ SCCF	∃ SCCF	–
	2007	1	$r = 0: 75.18$ $r = 1: 14.37$	29.79 15.49	1.802	∃ SCCF	∃ SCCF	–
(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

Given the estimates of γ_3, γ_6, and $\rho \equiv \frac{1}{(1+\exp(\overline{d-p}))}$, 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 & 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 (Athanasopoulos et al., 2011; Hecq et al., 2006): there exists a 1×3 vector $\gamma' = \begin{pmatrix} \rho & 0 & -1 \end{pmatrix}$ such that $\gamma' \begin{pmatrix} a^i_{11} & a^i_{12} & a^i_{13} \\ a^i_{21} & a^i_{22} & a^i_{23} \\ a^i_{31} & a^i_{32} & a^i_{33} \end{pmatrix} = 0$, $i = 1, \ldots, p-1$, implying that the first row of $\begin{pmatrix} a^i_{11} & a^i_{12} & a^i_{13} \\ a^i_{21} & a^i_{22} & a^i_{23} \\ a^i_{31} & a^i_{32} & a^i_{33} \end{pmatrix}$ is proportional to the last row $(1/\rho)$; and 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 in the assumptions as to 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 these policies vary across time, making it very difficult to know exactly what short-run parameter restrictions it would be appropriate to impose in the 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 prices 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 by Robert Shiller for the period 1871–2012 and stored 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 Appendix A. 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. Thus, imposing two vectors could be a possibility. As far as common-cyclical features are concerned, the results in columns (7)–(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 imposing these restrictions in unrestricted multivariate

4 In 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).

models for the same data leads to an improvement in standard forecast accuracy measures or not. This experiment has bearings not only on theory, but also on practical issues.

4. Out-of-sample forecasting

4.1. Forecasting strategies imposing different co-movement restrictions

Next, we describe the different forecasting strategies used in this paper. Each of the strategies 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 the estimated models; see the results of Athanasopoulos et al. (2011) and Vahid and Issler (2002) in a common-cycle context. All models are estimated by conditional maximum likelihood (ML),[5] after which 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 that is possibly (but not necessarily) restricted by cointegration and/or weak-form serial-correlation common features. Jointly select p, the rank of the short-run matrices, and the cointegrating rank, by a use of the combination of the posterior information criterion (PIC) and the HQ criterion, as was suggested by Athanasopoulos et al. (2011).
3. **VECM(HQ-J)**: This is the VECM that uses only the PVM-cointegration restriction. Select p according to the HQ criterion, and impose the cointegrating-rank restriction that is 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 that uses only 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, $\widehat{\theta}$ (T-consistent), to constrain α_1, as $\widehat{\alpha}_1 = \frac{1-\widehat{\delta}}{\widehat{\delta}} = 1/\widehat{\theta}$. 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 $\widehat{\rho} \equiv \frac{1}{(1+\exp(\overline{d-p}))}$, where $\overline{d-p}$ is the time-average of $d_t - p_t$. With the last two equations of reduced-form estimates and $\widehat{\rho}$, 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 (1), all the way through to the PVM-restricted Models (5) and (6), in levels and log-levels, respectively. There is a rationale for choosing each one 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 imposing them *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, which is consistent with Propositions 1 and 3. Second, notice that Model (4) refrains from imposing the parameter restrictions listed in Propositions 1 and 3, imposing only 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 of the restrictions listed in either Proposition 1 or Proposition 3. Comparisons of Models (5) and (6) with the unrestricted VAR and with Model (2) can help to determine whether imposing *structural* restrictions helps in forecasting, thus 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 studies in finance and forecasting about the issue of return predictability; see for example Campbell and Thompson (2008), Cochrane (2008), and Goyal and Welch (2003, 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—the 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 predictors, being similar in spirit to the exercises implemented in this body of literature.[8]

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

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

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

8 We thank an anonymous referee for pointing this out.

The estimation details for the forecasting models are as follows. First, we divide our full sample into an "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, including all of the information available to date. We set $h = 12$ years in the forecasting exercise. This enables us to measure the forecast accuracy in the short- and medium-term (h between 1 and 5 years), as well as in 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 to 1945) and the forecasting sample has 67 (1946–2012). Estimation is performed using a rolling window, which is 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:

$$RRMSFE_h^M = \frac{RMSFE_h^M}{RMSFE_h^{VAR}}, \tag{32}$$

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

We want to be able to distinguish the forecast accuracies of Models (1) to (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 also with all of the other models, although the latter 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 each model M vis-à-vis that of the VAR, described in Eq. (32). Since all of the restricted representations (Models (2) to (6)) forecast the first differences of the data, but the VAR (Model (1)) forecasts their level, we transform the VAR forecasts errors into first-difference errors in order to compute the ratio in Eq. (32).

Following the empirical finance 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 Ys, instead of 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 in 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), Model (4), VECM(HQ)Rank, dominates at all horizons, although Models (3) and (5), VECM(HQ-J) and the PV model, perform well on occasion. If we extend the forecast period to 2012, Model (4) dominates again, although the PV model (Model (5)) performs well in the medium- to long-term. Thus, for i_{lr} forecasts, we conclude that Model (4) performs really well, followed by the PV model (Model (5)). Nevertheless, 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 (Model (5)) performed really well regardless of the forecast sample (2007 or 2012), although its performance in the short term for the sample up to 2007 was beaten by those of Model (2) (VECM(HQ-PIC)) and Model (4). Thus, for P_t, we conclude that the PV model (Model (5)) is the best, followed by Models (2) and (4).

In Tables 6 and 7, we present forecasting results for p_t. Regardless of the sample period, we find that the best model for p_t is Model (3), while Model (4) performed well on occasion (middle to long horizons). Most of the time, the forecasts from Model (3) 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 (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 Campbell and Thompson (2008), Cochrane (2008) and Goyal and Welch (2003, 2008). The same can be said about Model (4) (VECM(HQ)Rank), and to a lesser extent Model (3) (VECM(HQ-J)), which only performed really well in the logarithmic case for p_t. For financial series, this highlights the importance of building forecasting models which consider Model (2), include Model (4), or impose Model (5) short-run restrictions. We stress this because, of all of the strategies, the only strategy that rules out an important

[9] Note that the numbers of out-of-sample observations for $h = 1, 2, \ldots, 12$ vary: the number is 67 for the first but 56 for the last.

[10] The 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 Busetti 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 2
Relative RMSFEs of restricted models vs. the 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]

The table shows the 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. The best models are given in bold.
* 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 numbers in square brackets are test statistics 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 numbers in square brackets are test statistics 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 numbers in square brackets are test statistics and the critical value is equal to 6.349.

Table 3
Relative RMSFEs of restricted models vs. the 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.

Table 4
Relative RMSFEs of restricted models vs. the 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	**0.951**	**0.954***	**0.959**	0.958	0.971	0.99	0.995	0.987	0.994	0.998	0.993
	[0.281]	[2.461]	[3.031]	[2.144]	[2.005]	[1.386]	[1.323]	[1.363]	[1.737]	[0.515]	[0.223]	[1.828]
VECM(HQ-J)	**0.945**	0.954*	0.962*	0.965	0.96	0.971	**0.987**	0.991	0.984**	0.992	0.997	0.993
	[0.185]	[3.743]	[3.483]	[2.621]	[2.35]	[1.924]	[1.368]	[1.467]	[6.389]	[0.706]	[0.302]	[1.799]
VECM(HQ)Rank	0.95	0.955	0.958	0.962*	**0.958**	**0.97**	0.989	0.994	0.985**	0.992	0.997	**0.992**
	[0.252]	[1.487]	[1.394]	[2.778]	[2.369]	[0.025]	[0.091]	[0.771]	[4.737]	[0.791]	[0.353]	[1.847]
PV	0.965	0.968	0.983	0.985	0.973	0.982	0.991	**0.98**	**0.974***	**0.983**	**0.989**	0.993
	[0.189]	[1.371]	[0.904]	[0.829]	[1.356]	[1.192]	[1.134]	[1.066]	[2.766]	[0.822]	[0.737]	[0.302]

See the notes to Table 2.

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

See the notes to Table 2.

role for short-run restrictions is Model (3). Although it did really well for the logarithmic case (p_t), strategies that allow short-run restrictions to play a role 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 (1–5 years ahead). In this case, there are two winners: Models (3) and (4), which are tied as the best strategy 30% of the time each. These are followed closely by Model (5), which is the best 26.67% of the time, and Model (2), which

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

See the notes to Table 2.

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

See the notes to Table 2.

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 Model (3), VECM(HQ-J), or Model (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 Athanasopoulos et al. (2011) and Hecq et al. (2006). They also imply 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 of the short-run dynamics of the system being constrained for stationary data (Vahid & Issler, 2002), and for data that are subject to long- and short-run restrictions (Anderson & Vahid, 2011; Athanasopoulos et al., 2011; Issler & Vahid, 2001). Appropriate common-cycle restrictions improve forecasting because they find linear combinations of the first differences of the data that cannot be forecast based on past information. This embeds natural exclusion restrictions, preventing the estimation of useless parameters, which would otherwise contribute to an increase in 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, found 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 real risk-free rate to be included. 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 Models (3), VECM (HQ-J), and (4), VECM(HQ)Rank, followed closely by the PV model (Model (5)).

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

Finally, when taken as a whole, our results show the usefulness of imposing short-run restrictions when forecasting economic series. It should be stressed that strategies that allow short-run restrictions to play a role performed best 70% of the time at business cycle horizons: Model (4) (VECM(HQ)Rank), the PV model (Model (5)), and Model (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.

Acknowledgments

This paper supersedes that of Hecq and Issler (2012), with two additional coauthors. We gratefully acknowledge the comments and suggestions provided by Ricardo Brito, Rob Hyndman, Søoren 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.

Appendix A. 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 parameter 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 the long-run and short-run parameters jointly as per Centoni et al. (2007).

5 Integrate steps 1, 2 and 3, estimating the lag length of the VAR and the long-run and short-run parameters jointly as per 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:

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

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

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

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 ith 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:

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

11 In 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 $\tilde{\gamma}_1$:

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

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

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

Note 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 proposed here for testing PVMs is to test the rank condition (necessary) first, without imposing any further parameter restrictions yet. As was argued 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 being 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 rejecting the theory, if such 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.

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

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 $\left(\Delta Y_t, \Delta y_t, \hat{S}_{t-1}\right)'$ and the conditioning set W_t'. For example, in the context of Eq. (7), we would have the following moment restrictions:

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

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

$$\begin{aligned}\hat{\theta}_{GIVE}&=\left(\boldsymbol{\Delta}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Delta}\mathbf{X}\right)^{-1}\\ &\quad\times\left(\boldsymbol{\Delta}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Delta}\mathbf{Y}\right),\end{aligned} \tag{36}$$

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

$$\begin{aligned}\hat{\theta}_{GMM}&=\left(\boldsymbol{\Delta}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Delta}\mathbf{X}\right)^{-1}\\ &\quad\times\left(\boldsymbol{\Delta}\mathbf{X}'\mathbf{W}(W'\mathbf{B}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Delta}\mathbf{Y}\right),\end{aligned} \tag{37}$$

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:

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

Table 8
Empirical size (nominal, set to the 5% level).

	VAR(p)	T = 100					
		Levels			Log levels		
		ζ_{LR}	J_1	J_2	ζ_{LR}	J_1	J_2
DGP #1: i.i.d.	$p=2$	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

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 should be used. 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:

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

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 i.i.d. case, the behaviors 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.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2015.02.002.

References

Anderson, H. M., & Vahid, F. (2011). VARs, cointegration and common cycle restrictions. In M. P. Clements & D. F. Hendry (Eds.), *The Oxford handbook of economic forecasting* (Chapter 1). Oxford University Press.

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.

Baillie, R. (1989). Econometric tests of rationality and market efficiency. *Econometric Reviews*, *8*, 151–186.

Busetti, F., & Marcucci, J. (2013). Comparing forecast accuracy: A Monte Carlo investigation. *International Journal of Forecasting*, *29*(1), 13–27.

Campbell, J. Y. (1987). Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis. *Econometrica*, *55*(6), 1249–1273.

Campbell, J. Y., & Deaton, A. (1989). Why is consumption so smooth? *The Review of Economic Studies*, *56*, 357–374.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1996). *The econometrics of financial markets*. Princeton University Press.

Campbell, J. Y., & Shiller, R. (1987). Cointegration tests and present value models. *Journal of Political Economy*, *95*, 1062–1088.

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.

Campbell, J. Y., & Shiller, R. (1988b). Stock prices, earnings and expected dividends. *Journal of Finance*, *43*, 661–676.

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.

Candelon, B., Hecq, A., & Verschoor, W. (2005). Measuring common cyclical features during financial turmoil: Evidence of interdependence not contagion. *Journal of International Money and Finance*, *24*, 1317–1334.

Centoni, M., Cubadda, G., & Hecq, A. (2007). Common shocks, common dynamics and the international business cycle. *Economic Modelling*, *24*(1), 149–166.

Chow, G. (1984). *Econometrics*. New York: McGraw-Hill.

Clements, M., & Hendry, D. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press, 398 pages.

Cochrane, J. H. (2008). The dog that did not bark: a defense of return predictability. *The Review of Financial Studies*, *21*(4), 1533–1575.

Cubadda, G., & Hecq, A. (2001). On non-contemporaneous short-run comovements. *Economics Letters*, *73*, 389–397.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica*, *55*, 251–276.

Engle, R. F., & Kozicki, S. (1993). Testing for common features (with comments). *Journal of Business and Economic Statistics*, *11*, 369–395.

Engsted, T. (2002). Measures of fit for rational expectations models. *Journal of Economic Surveys*, *16*, 301–355.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Goyal, A., & Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, *49*(5), 639–654.

Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.

Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica*, *50*, 1029–1054.

Hansen, L. P., & Sargent, T. J. (1981). A note on Wiener-Kolmogorov prediction formulas for rational expectations models. *Economics Letters*, *8*(3), 255–260.

Hansen, L. P., & Sargent, T. J. (1993). Recursive linear models of dynamic economies. In *Proceedings of the Federal Reserve Bank of San Francisco* (March issue).

Hecq, A., & Issler, J. V. (2012). *A common-feature approach for testing present-value restrictions with financial data*. Economics Working Papers (Ensaios Economicos da EPGE) 728, FGV/EPGE.

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.

Hoffman, D. L, & Rasche, R. H. (1996). Assessing forecast performance in a cointegrated system. *Journal of Applied Econometrics*, *11*, 495–517.

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.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, *59*, 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Johansen, S. (2000). Modelling of cointegration in the vector autoregressive model. *Economic Modelling*, *17*, 359–373.

Johansen, S., & Swensen, A. (1999). Testing rational expectations in vector autoregressive models. *Journal of Econometrics*, *93*, 73–91.

Johansen, S., & Swensen, A. (2004). More on testing exact rational expectations in cointegrated vector autoregressive models: restricted drift terms. *Econometrics Journal*, *7*, 389–397.

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.

Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, *156*(1), 3–20.

Vahid, F., & Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, *8*, 341–360.

Vahid, F., & Engle, R. F. (1997). Codependent cycles. *Journal of Econometrics*, *80*, 199–221.

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.

International Capital Mobility: A panel Data Approach

May 24, 2016

Abstract

In a world with perfect capital mobility, a country can always run current account deficits if its desire to consume and invest cannot be funded domestically. This paper examines if we live in a world with perfect capital mobility following the intertemporal approach to the current account developed in Sachs (1982). We estimate a panel vector autoregressive model (PVAR) and test the restrictions imposed by the the present value relationship yielded by the model solution. We use the econometric techniques developed in Campbell (1987) and Campbell and Shiller (1987) extended for panel data analysis. Our results reject the intertemporal current account model, once we reject some of the model implications. This result give evidences that there isn't perfect international capital mobility. However, in our robustness analysis with subgroups of countries we find some different results, which indicate the model can be partially accepted.

Keywords: Intertemporal current account, Capital mobility, Present-value restrictions, Dynamic Panel analysis

1 Introduction

The assumption that capital is perfectly mobile between countries underlies many discussions in open economy macroeconomics and policy debates. A large empirical literature has investigated how mobile capital is between countries. However, this subject is still controversial and there is divergence in the literature.

Feldstein and Horioka (1980) claims that, even amongst the major industrialized countries, capital mobility was severely limited. They found very high saving-investment correlations for a large cross-section of OECD countries. While many other studies also have found this positive correlation, some authors refuse to accept their interpretation, arguing that the world capital markets are well integrated.

Since the Feldstein-Horioka paper, their approach that it is possible to make inference about international capital mobility (ICM) from saving and investment data alone has been questioned on theoretical and econometric grounds. Sachs (1982) have proposed a formal model to show how today's current account is a function of both current and future economic variables, and how it can be determined under classical assumptions, including perfect capital mobility. If capital is perfect mobile, then it should smooth consumption in the face of shocks to national cash flow. Based on this assumption, the current account would act as as a buffer to smooth consumption in the face of shocks to output, investment, and government expenditure. This approach allows to construct a time series for the optimal current account which should have been observed, given the actual shocks to the economy. Then, the actual current account series can be compared to the optimal series, under the null hypothesis of perfect capital mobility, the two series should be identical. Formally tests can be performed to compare both series and check if the capital has been sufficiently mobile.

Some papers have followed the intertemporal approach to the current account developed in Sachs (1982) to test capital mobility. Otto (1992) has tested the model for the american and canadian series. The model was rejected in both cases. Ghosh (1995) tested the capital mobility for the Unites States, Japan, Germany, Canada and United Kingdom. His results is a mixed bag, as he didn't reject the model for the US but reject for the others. Ghosh and Ostry (1995) has used the model for some developing countries and again have found different results.

On more recent papers, Hoffmann (2004) and Evans et al. (2008) analyze cointegration between savings and investment to analyze international capital mobility but arrive to different conclusions. Costantini and Gutierrez (2013) focus on the effects of global factors on the saving investment relationship, they find that when global shocks are taken into account through common factors, the estimated saving-retention coefficient is close to zero. Holmes and Otero (2014) use data for 25 OECD countries to re-examine the original Feldstein Horioka and Sachs' equations, they find mixed evidence of capital mobility.

Kumar (2015) investigate the impact of regional integration agreements (AFTA, EU, EFTA, CARTAGENA, MERCOSUR and NAFTA) on the international capital mobility. They estimate the cointegrating equation between saving and investments and dynamic adjustments in panel data, finding that international mobility of capital has increased in these countries.

Our study is an extension of these previous work based on Sachs (1982) approach, as we are seeking to find a more conclusive answer. We try to understand the capital mobility in a global way, not only for specific countries. Our strategy is to use a panel of time-series and cross-section data and testing, simultaneously, the implications of the theoretical considerations on more than one country, which provides more powerful statistical tests and precise parameters estimation. Pelgrin and Schich (2008) and Adedeji and Thornton (2008) also works in the panel data framework, nonetheless they follow the saving-invesment approach which have some econometric and economic issues.

Therefore, our approach has two advantages. One is to cope with the saving-investment models issues, as typically savings and investments are non-stationary series whereas current account is stationary. Also, the benchmark value of the current account, gives an estimate what should happen in the absence of capital mobility restrictions, while in the saving-investment approach a positive correlation between them, does not, in itself provide evidence against capital mobility. On top of that, we benefit from the higher power tests and more precise estimates of a panel data analysis.

To test the intertemporal current account model we follow the approach developed by Campbell (1987) and Campbell and Shiller (1987) extended for the panel data analysis. We estimate a panel vector autoregressive model (PVAR) and test the restrictions imposed by the the present value relationship yielded by the model solution.

Our results reject the intertemporal current account model, once we reject some of the model implications. Consequently, the findings for the whole sample suggest that there isn't perfect capital mobility around the world. However, in our robustness analysis with subgroups of countries we find some different results, which indicate the model can be partially accepted.

The rest of the paper is arranged as follows. Section 2 sets out the theoretical model used to construct the benchmark current account series. Section 3 present the econometric approach used to test the model assumptions. Section 4. describes the data. Section 5. provides the empirical results. At last, Section 6. concludes.

2 Theoretical model

In order to estimate the optimal current account we follow the intertemporal approach to the current account developed in Sachs (1982). Its basic hypothesis consist of: a representative infinite-lived agent, a small and open economy, perfect capital mobility. The

main feature of the model is that the agent uses the current account as a buffer to smooth consumption in the face of shocks to output, investment and government expenditure. Therefore, a temporary unanticipated increase in government expenditure or investment, would lead to current account deficit as the country would like to smooth consumption by borrowing in international capital markets. Also, the intertemporal approach to current-account allows a forward looking interpretation of the current account. Suppose, for instance, that private agents expect a temporary increase in government expenditure in the future, then the model would predict an immediate current account surplus as the country saves for higher expenditure in the future.

As said above, the economy is assumed to be populated by a representative agent whose preferences are given by:

$$\sum_{t=1}^{\infty} \beta^t E[u(c_t)] \quad 0 < \beta < 1, \tag{1}$$

where β is the subjective discount rate, $u(\cdot)$ is the quadratic utility function, $u(c_t) = c_t - c_t^2/2$ and c_t is the consumption of a single good.

The solution of the model is achieved working in terms of the social planner's problem, maximizing (1) subject to the economy's dynamic budget constraint:

$$b_{t+1} = (1+r)b_t + q_t - c_t - i_t - g_t, \tag{2}$$

where b is the level of the foreign assets held by the economy, r is the fixed world interest rate, q is the GDP, i is the level of investment and g is the government expenditure. The first order condition is the Euller equation:

$$u'(c_t) = \beta(1+r)E_t[u'(c_{t+1})],$$

Substituting the quadratic utility function in the equation results:

$$1 - c_t = (1+r)\beta(1+r)E_t[1 - c_{t+1}], \tag{3}$$

Moving equation (3) forward j periods and considering that $(1+r)\beta$ is close to one we find:

$$E_t(c_{t+j}) = \frac{c_t - 1}{(1+r)^j \beta^j},$$

Moving equation (2) forward one period and substituting the value for b_{t+1} gives the following result:

$$(1+r)b_t = c_t + g_t + i_t - y_t + \frac{c_{t+1} - g_{t+1} - i_{t+1} - y_{t+1}}{1+r} + \frac{b_{t+2}}{1+r}$$

Repeating this procedure sequentially, we arrives at:

$$\sum_{j=0}^{T} \left(\frac{1}{1+r}\right)^j (c_{t+j}) + \left(\frac{1}{1+r}\right)^T b_{T+1} = (1+r)b_t + \sum_{j=0}^{T} \left(\frac{1}{1+r}\right)^j (y_{t+j} - g_{t+j} - i_{t+j}), \quad (4)$$

with the following "no-Ponzi game" constraint:

$$\lim_{T\to\infty} \left(\frac{1}{1+r}\right)^T b_{T+1} = 0$$

When it takes $\mathbb{E}_t(\cdot)$ from both sides of (4) and T goes to infinity, we obtain:

$$\mathbb{E}_t \sum_{j=0}^{\infty} \left(\frac{1}{1+r}\right)^j (c_{t+j}) = (1+r)b_t + \mathbb{E}_t \sum_{j=0}^{\infty} \left(\frac{1}{1+r}\right)^j (y_{t+j} - g_{t+j} - i_{t+j}) \quad (5)$$

Substituting (4) in (5) we find the approximate equation for the optimal consumption:

$$c_t^* = \frac{r}{\theta}\left(b_t + (1+r)^{-1}\mathbb{E}_t\left[\sum_{j=0}^{\infty}(1+r)^{-j}(q_{t+j} - i_{t+j} - g_{t+j})\right]\right) \quad (6)$$

where $\theta = \beta(1+r)r/[\beta(1+r)^2 - 1]$ and $z_{t+1} \equiv q_{t+j} - i_{t+j} - g_{t+j}$ is the commonly called national cash flow. Therefore, the optimal consumption is proportional to permanent national cash flow, where θ is a constant of proportionality reflecting the consumption-tilting dynamics of consumption. For $\theta < 1$ the country is tilting consumption towards the present as it is consuming more than its current permanent cash flow. For $\theta = 1$ the country is consuming exactly its current permanent cash flow, and if $\theta > 1$ the country is tilting consumption towards the future. In our analysis we will not include the current account movements associated with the consumption-tilting motive.

The optimal consumption-smoothing current account is defined by:

$$CA_t^* \equiv y_t - i_t - g_t - \theta c_t^* \quad (7)$$

where y_t is national income (GNP), equal to GDP (q_t) plus net factor payments (rb_t). Substituting (6) into (7) yields:

$$CA_t^* = y_t - i_t - g_t - r\left\{b_t + (1+r)^{-1}\mathbb{E}_t\sum_{j=0}^{\infty}(1+r)^{-j}(q_{t+j} - i_{t+j} - g_{t+j})\right\}$$

which gives:

$$CA_t^* = -E_t \sum_{j=1}^{\infty} (1+r)^{-j} \Delta(q_{t+j} - i_{t+j} - g_{t+j}) \tag{8}$$

Equation (8) shows that the optimal current equals (minus) the expected present discounted value of changes in the national cash flow $(q_{t+j} - i_{t+j} - g_{t+j})$. The expression for the optimal current account, CA^*, embodies the intertemporal approach to the current account, on what matters for the determination of the current account is agents' expectations of the shocks to the economy, rather than the shocks themselves. Permanent shocks to output, government expenditure, or investment have no effect on the current account since their expected change is zero. However, transitory shocks will have effects on the expected national cash flow leading to changes in the optimal current account.

3 Econometric approach

3.1 Present Value model

First, let's rewrite the equation (8) expliciting the information set of the representative agent, I_t:

$$CA_{i,t}^* = -\sum_{j=1}^{\infty} (1+r)^{-j} \mathbb{E}_t[\Delta(q_{i,t+j} - i_{i,t+j} - g_{i,t+j})|I_t] \tag{9}$$

To create this optimal current account series we need to calculate the expected present discounted value of changes in national cash flow. The equation (9) is equivalent to the present value model(PVM) introduced by Campbell and Shiller (1988), therefore we follow their techniques and first estimate an unrestricted VAR in $CA_{i,t}$ and $z_{i,t} = q_{i,t} - i_{i,t} - g_{i,t}$, where $CA_{i,t}$ is the actual consumption-smoothing component of the current account:

$$CA_{i,t} \equiv y_{i,t} - i_{i,t} - g_{i,t} - \theta_i c_{i,t} \tag{10}$$

Actually, in our case it is a panel VAR, that explains the subscript i in the variables. Nevertheless, the extension of the PVM for the panel case is straightforward and the estimation techniques will be explained in another section. After the estimation, we test the restrictions implied by the PV relationships; see Johansen and Swensen (2011) and Guillén et al. (2015).

The vector autoregression may be written:

$$\begin{pmatrix} \Delta z_{i,t} \\ CA_{i,t} \end{pmatrix} = \begin{pmatrix} a(L) & b(L) \\ c(L) & d(L) \end{pmatrix} \begin{pmatrix} \Delta z_{i,t-1} \\ CA_{i,t-1} \end{pmatrix} + \begin{pmatrix} u^1_{i,t} \\ u^2_{i,t} \end{pmatrix} \quad \forall j = 1, \ldots, N \tag{11}$$

where the polynomials in the lag operator $a(L)$, $b(L)$, $c(L)$, and $d(L)$ are all of order p. Equation (11) can be stacked in a first-order system:

$$\begin{pmatrix} \Delta z_{i,t} \\ \vdots \\ \Delta z_{i,t-p+1} \\ CA_{i,t} \\ \vdots \\ CA_{i,t-p+1} \end{pmatrix} = \begin{pmatrix} a_1 & \ldots & a_p & b_1 & \ldots & b_p \\ 1 & \ddots & & & & \\ & & 1 & & & \\ c_1 & \ldots & c_p & d_1 & \ldots & d_p \\ & & & 1 & \ddots & \\ & & & & & 1 \end{pmatrix} \begin{pmatrix} \Delta z_{i,t-1} \\ \vdots \\ \Delta z_{i,t-p} \\ CA_{i,t-1} \\ \vdots \\ CA_{i,t-p} \end{pmatrix} + \begin{pmatrix} u_{1i,t} \\ \vdots \\ 0 \\ u_{2i,t} \\ \vdots \\ 0 \end{pmatrix}$$

which can be written more succinctly as:

$$X_{i,t} = AX_{i,t-1} + U_{i,t}$$

The matrix A is called the companion matrix of the VAR. A typical term in the optimal current account equation (8) is $\mathbb{E}_t \Delta z_{i,t+j}$ which can be written $h' \mathbb{E}_t X_{i,t+j}$, where h is a column vectors with 2p elements, all of which are zero except for the first element. The information set consisting of current and lagged $\Delta z_{i,t}$ and $CA_{i,t}$ will be written H_t, a subset of agents' information I_t. So that:

$$\mathbb{E}_t[\Delta z_{i,t+j} | H_t] = h' \mathbb{E}_t[X_{i,t+j} | H_t] = h' A^k X_{i,t} \tag{12}$$

so it is straightforward to compute multi-period forecasts once A is estimated.

An important implication of equation (8) for the vector autoregression is that $CA_{i,t}$ must granger cause $\Delta z_{i,t}$. Intuitively, current account is an optimal forecast of future declines in national cash flow, conditional on agents' full information set. Therefore, if agents have more information about the evolution of national cash flow than is contained in its own past values then this additional information should be reflected in the current account. If, for example, a change in administrations portends higher future government expenditure then the country should run a current account surplus. This current account surplus will then Granger cause the change in the national cash flow. This corresponds to the "saving for a rainy day" of Campbell (1987) in his study of the permante income hypothesis.

Projecting equation (8) onto the information set H_t, and noting that the left-hand side is unchanged by the projection since CA is in H_t, we obtain:

$$CA^*_{i,t} = -\sum_{j=1}^{\infty}(1+r)^{-j}\mathbb{E}_t[\Delta z_{i,t+j}|H_t] \tag{13}$$

Stationarity of the variables $CA_{i,t}$ and $\Delta z_{i,t}$ in the VAR is sufficient for the infinite sum on the right-hand side of (13) to converge. Assuming that $z_{i,t}$ is $I(1)$ its first difference will be stationary; since the current account is a discounted sum of $\Delta z_{i,t}$, it will be stationary too. This hypothesis will be formally tested in the next sections. Then, using the results of equation (12),the infinite sum can be written as:

$$CA^*_{i,t} = -\sum_{j=1}^{\infty}(1+r)^{-j}h'A^iX_{i,t}$$

which gives:

$$CA^*_{i,t} = -h'[\frac{A}{(1+r)}][\mathbf{I} - \frac{A}{(1+r)}]^{-1}X_{i,t} \tag{14}$$

To formally compare the optimal current account series $CA^*_{i,t}$ to the observed current account $CA_{i,t}$, we define $CA^* = g'X_t$ where g is a column vector with $2p$ elements, all of which are zero except for the $p+1$st element of g. Rewriting equation (14):

$$g'X_t = -h'[\frac{A}{(1+r)}][I - \frac{A}{(1+r)}]^{-1}X_{i,t}$$

or

$$g' = -h'[\frac{A}{(1+r)}][I - \frac{A}{(1+r)}]^{-1}$$

Posmultiplying both sides by $[I - \frac{A}{(1+r)}]^{-1}$, gives:

$$g'[I - \frac{A}{(1+r)}]^{-1} = -h'[\frac{A}{(1+r)}] \tag{15}$$

From the equation (15) one obtains the following individual restricions on the VAR

companion matrix A:

$$\begin{cases} c_j = a_j, & j = 1, ...p \\ d_1 = (1+r) + b_1 \\ d_j = b_j, & j = 2, ...p \end{cases} \tag{16}$$

which will be tested by the Wald test.

It remains only to describe how to calculate the consumption-tilting parameter θ_i so that the tilting componente of the current account can be removed. The optimal current account series, CA_t^* will be an $I(0)$ process given the stationarity of $\Delta z_{i,t}$. Under the null hypothesis that the actual consumption-smoothing component of the current account is equal to CA_t^*, that one will also be an $I(0)$ process. This means that the definition of CA_t in (10) is $I(0)$ and, therefore θ_i can be obtained as the cointegration parameter between $c_{i,t}$ and $z_{i,t}$. We thus apply the Johansen (1991) procedure for each country to estimate the θ_i components.

Wrapping up the implications of the model which will be studied in the next sections: Existence of unit roots for consumption $c_{i,t}$ and net income $z_{i,t}$, stationarity for $\Delta c_{i,t}$, $\Delta z_{i,t}$ and $CA_{i,t}$, cointegration between $c_{i,t}$ and $z_{i,t}$, $CA_{i,t}$ Granger cause $\Delta z_{i,t}$ and finally the restrictions on the companion matrix A implied by the present value model.

3.2 Panel unit root tests

The following discussion outlines the basics features of panel unit root tests. In the recent literature, it has been discussed a variety of panel unit root tests[1]. Panel-based unit root tests have higher power than unit root tests based on individual time series. In this seciton we discuss the following types of panel unit root tests: Im et al. (2003), Fisher-type tests using Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests (Maddala and Wu (1999) and Choi (2001)), and Hadri (2000)[2].

Panel unit root tests are similar, but not identical, to unit root tests carried out on a single series. We begin by classifying our unit root tests on the basis of whether there are restrictions on the autoregressive process across cross-sections or series. Consider a following AR(1) process for panel data:

$$w_{i,t} = \rho_i w_{i,t-1} + \varepsilon_{i,t}$$

where $i = 1, 2, \ldots, N$ cross-section units or series, that are observed over periods $t =$

[1]See Levin et al. (2002), Breitung (2002), Im et al. (2003), Maddala and Wu (1999), Choi (2001), and Hadri (2000).

[2]While these tests are commonly termed "panel unit root" tests, theoretically, they are simply multiple-series unit root tests that have been applied to panel data structures

$1, 2, \ldots, T$.

The parameters ρ_i's are the autoregressive coefficients, and the errors $\varepsilon_{i,t}$ are assumed to be mutually independent idiosyncratic disturbance process. If $|\rho_i| < 1$, w_i is said to be weakly stationary. On the other hand, if $|\rho_i| < 1$ then $w_{i,t}$ is a unit root process.

In order to test the presence of the unit root precess, there are two natural assumptions that we can make about the ρ_i. First, one can assume that the persistence parameters are common across cross-sections so that $\rho_i = \rho \ \forall \ i$. The Levin, Lin, and Chu (LLC), Breitung, and Hadri tests all employ this assumption. Alternatively, one can allow ρ_i to vary freely across cross-sections. The Im, Pesaran, and Shin (IPS), and Fisher-ADF and Fisher-PP tests are of this form.

The IPS, and the Fisher-ADF and PP tests all allow for individual unit root processes so that ρ_i may vary across cross-sections. The tests are all characterized by the combining of individual unit root tests to derive a panel-specific result.

The IPS begins by specifying a separate ADF regression for each cross section:

$$\Delta w_{i,t} = \alpha w_{i,t-1} + \sum_{j=1}^{p_i} \beta_{ij} \Delta w_{i,t-j} + u_{i,t} \tag{17}$$

The null hypothesis may be written as,

$$H_0 : \ \alpha_i = 0, \quad \forall \ \ i$$

while the alternative hypothesis is given by:

$$H_1 : \begin{cases} \alpha_i = 0, & \text{for } i = 1, \ldots, N_1, \\ \alpha_i < 0, & \text{for } i = N+1, N+2, \ldots, N, \end{cases}$$

(where the i may be reordered as necessary) which may be interpreted as a non-zero fraction of the individual processes is stationary.

After estimating the separate ADF regressions, the average of the t-statistics for α_i individual ADF regressions, $t_{iT}(p_i)$:

$$\overline{t_{NT}} = N^{-1} \sum_{i=1}^{N} t_{iT}(p_i)$$

In the case where the lag order is always zero ($p_i = 0$ for all i), simulated critical values for $\overline{t_{NT}}$ are provided in the IPS paper for different numbers of cross sections N, series lengths T, and for test equations containing either intercepts, or intercepts and linear trends.

In general, where the lag order in (17) may be non-zero for some cross-sections, IPS

show that a properly standardized $\overline{t_{NT}}$ has an asymptotic standard normal distribution:

$$W_{\overline{t_{NT}}} = \frac{N^{\frac{1}{2}}\left[\overline{t_{NT}} - N^{-1}\sum_{i=1}^{N} E\left[t_{iT}(p_i)\right]\right]}{\left[N^{-1}\sum_{i=1}^{N} Var\left[t_{iT}(p_i)\right]\right]^{\frac{1}{2}}} \rightarrow N(0,1)$$

The expressions for the expected mean and variance of the ADF regression t-statistics, $E\left[t_{iT}(p_i)\right]$ and $Var\left[t_{iT}(p_i)\right]$, are provided by IPS for various values of T and p and differing test equation assumptions, and are not provided here. The IPS test statistic requires specification of the number of lags and the specification of the deterministic component for each cross-section ADF equation. In particular, we use the Bayesian Information Criteria (BIC).

An alternative approach to panel unit root tests uses Fisher (1934) results to derive tests that combine the p-values from individual unit root tests. This idea has been proposed by Maddala and Wu (1999), and by Choi (2001). If we define π_i as the p-value from any individual unit root test for cross-section i, then under the null of unit root for all N cross-sections, we have the asymptotic result that

$$-2\sum_{i=1}^{N} log(\pi_i) \rightarrow \chi^2_{2N}(0,1) \tag{18}$$

In addition, Choi demonstrates that:

$$\Xi = N^{-\frac{1}{2}}\sum_{i=1}^{N} \Psi^{-1}(\pi_i) \rightarrow N(0,1) \tag{19}$$

where Ψ^{-1} is the inverse of the standard normal cumulative distribution function. The null and alternative hypotheses are the same as for the as IPS.

3.3 Panel cointegration tests

The extensive interest in and the availability of panel data has led to an emphasis on extending various statistical tests to panel data. We will discuss three types of panel cointegration tests: Pedroni (2004), Kao (1999) and a Fisher-type test using an underlying Johansen methodology Maddala and Wu (1999).

Here, we provide a brief description of the cointegration tests used in our empirical analysis. Pedroni (2004) and Kao (1999) tests are based on Engle and Granger (1987) two-step (residual-based) cointegration tests. The Fisher test is a combined Johansen test.

The Engle and Granger (1987) cointegration test is based on an examination of the

residuals of a spurious regression performed using I(1) variables. If the variables are cointegrated then the residuals should be I(0). On the other hand if the variables are not cointegrated then the residuals will be I(1). Pedroni (1999, 2004) and Kao (1999) extend the Engle-Granger frame-work to tests involving panel data.

Pedroni proposes several tests for cointegration that allow for heterogeneous intercepts and trend coefficients across cross-sections. Consider the following regression

$$w_{i,t} = \alpha_i + \delta_i t + \beta_{1i}x_{1i,t} + \beta_{2i}x_{2i,t} + \ldots + \beta_{1M}x_{1M,t} + e_{i,t} \tag{20}$$

for $t = 1, 2, \ldots, T$; $i = 1, 2, \ldots, N$; $m = 1, \ldots, M$; where $y_{i,t}$ and $x_{mi,t}$ are assumed to be integrated of order one, e.g. I(1). The parameters α_i and δ_i are individual and trend effects which may be set to zero if desired.

Under the null hypothesis of no cointegration, the residuals $e_{i,t}$ will be I(1). The general approach is to obtain residuals from (20) and then to test whether residuals are I(1) by running the auxiliary regression,

$$e_{i,t} = \phi_i e_{i,t-1} + u_{i,t} \tag{21}$$

or

$$e_{i,t} = \phi_i e_{i,t-1} + \sum_{j=1}^{p_i} \psi_{j,i}\Delta e_{i,t-j} + u_{i,t} \tag{22}$$

for each cross-section. Pedroni describes various methods of constructing statistics for testing for null hypothesis of no cointegration ($\phi_i = 01$). There are two alternative hypotheses: the homogenous alternative, $\phi_i = \phi < 1$ for all i (which Pedroni terms the within-dimension test or panel statistics test), and the heterogeneous alternative, $\phi_i < 1$ for all i (also referred to as the between-dimension or group statistics test).

The Pedroni panel cointegration statistic $\aleph_{N,T}$ is constructed from the residuals from either (21) or (22). Pedroni's tests can be classified into two categories. The first set (within dimension) is similar to the tests discussed above, and involves averaging test statistics for cointegration in the time series across cross-section. For the second set (between dimension), the averaging is done in pieces so that the limiting distributions are based on limits of piecewise numerator and denominator terms. The basic approach in both cases is to first estimate the hypothesized cointegration relationship separately for each member of the panel and then pool the resulting residuals when constructing the panel tests for the null of no cointegration. In the second step, the way in which the estimated residuals are pooled will differ among the various statistics, which are defined

as follows[3].

1. Panel ν-Statistic:

$$T^2 N^{\frac{3}{4}} Z_{N,T}^{\hat{\nu}} \equiv T^2 N^{\frac{3}{4}} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1} \right)^{-1} \tag{23}$$

2. Panel ρ-Statistic:

$$TN^{\frac{1}{2}} Z_{N,T-1}^{\hat{\rho}} \equiv TN^{\frac{1}{2}} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1} \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} (\hat{e}_{i,t-1} \Delta \hat{e}_{i,t} - \hat{\lambda}_i) \tag{24}$$

3. Panel PP-Statistic:

$$Z_{N,T-1}^{PP} \equiv \left(\tilde{\sigma}^2 N, T \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1} \right)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} (\hat{e}_{i,t-1} \Delta \hat{e}_{i,t} - \hat{\lambda}_i) \tag{25}$$

4. Panel ADF-Statistic:

$$Z_{N,T-1}^{ADF} \equiv \left(\tilde{s}^{*2} N, T \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1}^{*} \right)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11,i}^{-2} (\hat{e}_{i,t-1}^{*} \Delta \hat{e}_{i,t}^{*}) \tag{26}$$

where

$$\hat{\lambda}_i = \frac{1}{T} \sum_{s=1}^{k_i} \left(1 - \frac{s}{k_i + 1} \right) \sum_{t=s+1}^{T} \hat{u}_{i,t} \hat{u}_{i,t-s}$$

$$\hat{s}_i^2 = \frac{1}{T} \sum_{t=s+1}^{T} \hat{u}_{i,t}^2,$$

$$\hat{\sigma}_i^2 = \hat{s}_i^2 + 2\hat{\lambda}_i,$$

$$\tilde{\sigma}_{N,T}^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{L}_{11,i}^{-2} \hat{\sigma}_i^2,$$

$$\hat{s}_i^{*2} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_{i,t}^{*2},$$

$$\tilde{s}_{N,T}^{*2} = \sum_{i=1}^{N} \hat{s}_i^{*2}, \text{ and}$$

[3] We only report the within dimension statistics.

$$\hat{L}^2_{11,i} = \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}^2_{i,t} + \frac{2}{T}\sum_{s=1}^{k_i}\left(1 - \frac{s}{k_i+1}\right)\sum_{t=s+1}^{T}\hat{\eta}_{i,t}\hat{\eta}_{i,t-s}.$$

and where the residuals $\hat{u}_{i,t}$ and $\hat{u}^*_{i,t}$ from regressions of (21) and (22) respectively, and $\hat{\eta}_{i,t}$ is obtained from the regression:

$$\Delta w_{i,t} = \beta_{1i}\Delta x_{1i,t} + \beta_{2i}\Delta x_{2i,t} + \ldots + \beta_{1M}\Delta x_{1M,t} + \hat{\eta}_{i,t}$$

Pedroni shows that the standardized statistics above are asymptotically normally distributed,

$$\frac{\aleph_{N,T} - \mu N^{\frac{1}{2}}}{\nu^{\frac{1}{2}}} \to N(0,1)$$

where μ and ν are Monte Carlo generated adjustment terms.

The Kao test follows the same basic approach as the Pedroni tests, but specifies cross-section specific intercepts and homogeneous coefficients on the first-stage regressors.

In the bivariate case described in Kao (1999), we have

$$w_{i,t} = \alpha_i + \beta x_{i,t} + e_{i,t}$$

for

$$w_{i,t} = w_{i,t-1} + \varepsilon_{i,t}$$

$$x_{i,t} = x_{i,t-1} + \xi_{i,t}$$

for $t = 1, 2, \ldots, T$; $i = 1, 2, \ldots, N$. More generally, we may consider running the first stage regression (20), requiring the α_i to be heterogeneous, β_i to be homogeneous across cross-sections, and setting all of the trend coefficients γ_i to zero.

Kao then runs either the pooled auxiliary regression,

$$e_{i,t} = \phi e_{i,t-1} + \zeta_{i,t} \tag{27}$$

or the augmented version of the pooled specification,

$$\varepsilon_{i,t} = \phi\varepsilon_{i,t-1} + \sum_{j=1}^{p_i}\psi_{j,i}\Delta\varepsilon_{i,t-j} + \zeta_{i,t} \tag{28}$$

Under the null of no cointegration, Kao shows that following the statistics for the aug-

mented version, (i.e. $p > 0$)

$$ADF = \frac{t_{\hat{\rho}} + \sqrt{6}\hat{\sigma}_{\zeta}/2\sigma\hat{}_{0\zeta}}{\left[\hat{\sigma}_{0\zeta}^2/2\hat{\sigma}_{\zeta}^2 + 3\hat{\sigma}_{\zeta}^2/10\hat{\sigma}_{0\zeta}^2\right]^{\frac{1}{2}}} \tag{29}$$

converge to $N(0,1)$ asymptotically, where the estimated variance is $\hat{\sigma}_{\zeta}^2 = \hat{\sigma}_{\varepsilon}^2 - \hat{\sigma}_{\varepsilon\xi}^2\hat{\sigma}_{\xi}^{-2}$ with estimated long run variance $\hat{\sigma}_{0\zeta}^2 = \hat{\sigma}_{0\varepsilon}^2 - \hat{\sigma}_{0\varepsilon\xi}^2\hat{\sigma}_{0\xi}^{-2}$ The covariance of $\omega_{i,t} = [\varepsilon_{i,t}\ \xi_{i,t}]'$ is estimated as

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{\varepsilon} & \hat{\sigma}_{\varepsilon\xi} \\ \hat{\sigma}_{\varepsilon\xi} & \hat{\sigma}_{\xi} \end{bmatrix} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\omega}_{i,t}\hat{\omega}_{i,t}'$$

and the long run covariance is estimated using the usual Bartlett kernel estimator

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{0\varepsilon} & \hat{\sigma}_{0\varepsilon\xi} \\ \hat{\sigma}_{0\varepsilon\xi} & \hat{\sigma}_{0\xi} \end{bmatrix} = \frac{1}{N}\left[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{w}_{i,t}\hat{w}_{i,t}' + \frac{1}{T}\sum_{\tau=1}^{\infty}\kappa(\tau/b)\sum_{t=\tau+1}^{T}\hat{\omega}_{i,t-\tau}\hat{\omega}_{i,t}' + \hat{\omega}_{i,t-\tau}\hat{\omega}_{i,t-\tau}'\right]$$

where κ is the supported kernel function and b is the bandwidth.

Maddala and Wu (1999) use Fisher (1934) cointegration test and propose an alternative approach to testing for cointegration in panel data by combining tests from individual cross-sections to obtain a test statistic for the full panel. If π_i is the p-value from an individual cointegration test for cross-section i, then under the null hypothesis for the panel,

$$-2\sum_{i=1}^{N} log(\pi_i) \to \chi_{2N}^2$$

We use the χ^2 value based on MacKinnon et al. (1999) p-values for Johansen's cointegration trace test and maximum eigenvalue test.

3.4 Panel VAR

Let $\mathbf{w}_{i,t}$ be an $m \times 1$ vector of random variables for the ith cross-sectional unit at time t and suppose that the $\mathbf{w}_{i,t}$'s are generated by the following Panel Vector Autoregressive (PVAR) model model of order one, PVAR(1):

$$\mathbf{w}_{i,t} = (I_m - \Phi)\boldsymbol{\mu}_i + \Phi\mathbf{w}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where Φ denotes an $m \times m$ matrix of slope coefficients, $\boldsymbol{\mu}_i$ is an $m \times 1$ vector of individual-specific effects, $\boldsymbol{\varepsilon}_{i,t}$ is an $m \times 1$ vector of disturbances,and I_m denotes the identity matrix of dimension $m \times m$. For simplicity we restrict our exposition to first-order PVAR models.

The standard GMM estimator of Arellano and Bond (1991) and Binder et al. (2005)

employs instruments that are orthogonal to the disturbances of the first-differenced form of the model. Such instruments are given by levels of the dependent variables, lagged two or more periods. The resulting orthogonality conditions may be written as

$$E\left[\Delta\mathbf{w}_{i,t} - \Phi\Delta\mathbf{w}_{i,t-1}\mathbf{m}'_{t,i}\right] = 0 \tag{30}$$

where $\mathbf{m}_{i,t}$ is the $m(t-1)\times 1$ vector defined by

$$\mathbf{m}_{i,t} = (\mathbf{w}'_{i,0}, \mathbf{w}'_{i,1}, \ldots, \mathbf{w}'_{i,t-2})'.$$

To derive the standard GMM estimator of Φ based on the moment conditions (30), it will be useful to rewrite these moment conditions in stacked form as

$$E\left[\omega'_i\left(\Delta\mathbf{W}_i - \Delta\mathbf{W}_{i,-1}\Phi'\right)\right] = 0$$

where ω'_i is a matrix of dimension $mT(T-1)/2\times(T-1)$ given by

$$\omega'_i = \begin{pmatrix} \mathbf{m}_{i,2} & 0 & \ldots & 0 \\ 0 & \mathbf{m}_{i,3} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \mathbf{m}_{i,T} \end{pmatrix}$$

and $\Delta\mathbf{W}_i$ and $\Delta\mathbf{W}_{i,-1}$ are $(T-1)\times m$ dimensional matrices,

$$\Delta\mathbf{W}_i = (\Delta\mathbf{w}_{i,2}, \Delta\mathbf{w}_{i,3}, \ldots, \Delta\mathbf{w}_{i,T})'$$

and

$$\Delta\mathbf{W}_{i,-1} = (\Delta\mathbf{w}y_{i,1}, \Delta\mathbf{w}_{i,2}, \ldots, \Delta\mathbf{w}_{i,T-1})'$$

The standard GMM estimator of $\phi = \text{vec}(\Phi)$ is now given by

$$\hat{\phi}_{GMM} = \left(S'_{BC}D_{\hat{f}}^{-1}S_{BC}\right)S'_{BC}D_{\hat{f}}^{-1}S_{C\Upsilon} \tag{31}$$

where

$$S_{BC} = \frac{1}{N}\sum_{i=1}^{N}C'_iB_i, \quad S_{B\Upsilon} = \tfrac{1}{N}\sum_{i=1}^{N}C'_i\iota_i, \quad D_{\hat{f}} = \frac{1}{N}\sum_{i=1}^{N}C'_i\upsilon_iC_i,$$
$$\varsigma_i = f_if'_i, \quad C'_i = M'_i\otimes I_m, \quad B'_i = \Delta\mathbf{W}_{i,-1}\otimes I_m,$$
$$\Upsilon_i = \text{vec}(\Delta\mathbf{W}'_i), \quad f_i = \text{vec}\Delta F'_i, \quad \text{and } \Delta F = \Delta\mathbf{W}_i - \Delta\mathbf{W}_{i,-1}\hat{\Phi}'_{IE},$$

where $\hat{\Phi}'_{IE}$ is an initial consistent estimate of Φ such as the generalized instrumental variables estimator obtained using the formula (31) but with D_f replaced by $\Lambda_M \otimes I_m$ where

$$\Lambda_M = \sum_{i=1}^{N} M_i' \mathbb{V} M_i$$

and $\mathbb{V}$ is the $(T-1) \times (T-1)$ matrix,

$$\mathbb{V} = \begin{pmatrix} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & & 0 & \\ & & & \ddots & & & \\ & 0 & & & 2 & -1 & 0 \\ & & & & -1 & 2 & -1 \end{pmatrix}$$

Because the resultant instrumental variables estimator is invariant to the choice of Ω_f, without loss of generality the estimator may be computed replacing D_f by $\Lambda_M \otimes I_m$. Using the standard formula, a consistent estimate of the variance matrix of GMM can be obtained as

$$\frac{1}{N}\left(S'_{BC} D_{\hat{f}}^{-1} S_{BC}\right)^{-1}$$

The standard GMM estimator is consistent if all eigenvalues of Φ fall inside the unit circle but breaks down if some eigenvalues of Φ are equal to unity. Note that a necessary condition for the GMM estimator (31) to exist is that $\text{rank}(M_i' \mathbf{W}_{i,-1}) = m$ as $N \to \infty$.In the case where $\Phi = I_m$, $\text{rank}(M_i' \mathbf{W}_{i,-1})$ as $N \to \infty$ is less than m,however. This is because when $\Phi = I_m$ for $t = 2, 3, \ldots, T$ we have $\Delta \mathbf{w}_{i,t} = \boldsymbol{\varepsilon}_{i,t}$, and $\mathbf{w}_{i,t} = \mathbf{w}_{i,0} + \mathbf{s}_{i,t}$, with $\mathbf{s}_{i,t} = \sum_{q=1}^{t} \boldsymbol{\varepsilon}_{iq}$ and thus it follows that for $t = 2, 3, \ldots, T$, $l = 0, 1, \ldots, t-2$, as $N \to \infty$.

$$\frac{1}{N}\sum_{i=1}^{N} \Delta \mathbf{w}_{i,t-1} y'_{i,l} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\varepsilon}_{i,t-1}(\mathbf{w}_{i,0} + \mathbf{s}_{i,l}) \xrightarrow{p} 0$$

where $\xrightarrow{p}$ denotes convergence in probability. In other words, when $\Phi = I_m$, the elements of $q_{i,t}$, although still uncorrelated with the equation errors, are also uncorrelated with the regressors. The situation is analogous to that of Phillips (1989) on partial identification.

4 Data

This section briefly describes data sources. The main source of data used in this paper is taken from Penn World Table 7.1[4]. All data are annually series for the period from 1970 to 2010 covering 158 countries. We use PPP converted GNP, consumption, investment and government spending all in per capita at 2005 prices.

For the first robustness exercise performed in this paper we use a trade freedom index data computed by the the heritage foundation, which is an American conservative think tank based in Washington, D.C.. We rank the countries according to their trade freedom value, which is a composite measure of the absence of tariff and non-tariff barriers that affect imports and exports of goods and services. The trade freedom score is based on the trade-weighted average tariff rate and non-tariff barriers. The index uses the most recently reported weighted average tariff rate for a country from official sources. We took an average of the index values between 2011 to 2015 to classify the countries.

We perform a second robustness exercise using data from the World Bank, that revises the classification of the worlds economies based on estimates of gross national income (GNI) per capita for the previous year. We selected the high income countries based on this classification for 2014.

5 Results

The first step in the analysis is to verify the stochastic properties of the series used in the empirical exercise. Before the statistical tests, it's interesting to note the time path of consumption per capita and national cash flow for selected countries. From the Figure 1 below, we can see the charts suggest that c_t and $z_t = y_t - i_t - g_t$ have a trend whereas Δc_t and Δz_t seem to be stationary.

[4]See Heston et al. (2012).

Figure 1: Consumption and national cash flow for selected countries - Real per capita variables at 2005 prices

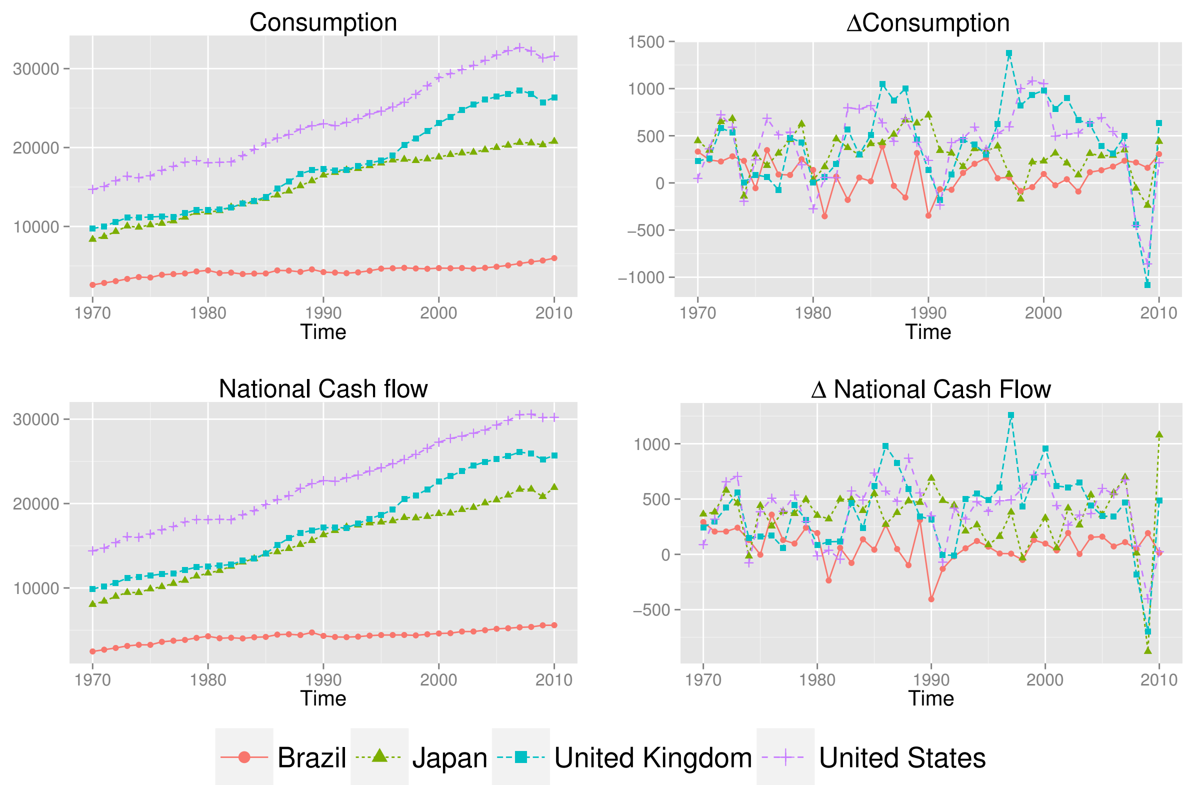


Table 1 presents formal statistical tests for robustness of our previous impressions from the visual analysis. Panel unit root tests for the series in level and first difference are showed.

Three different tests are reported in Table 1: Im, Pesaran and Shin (IPS); Augmented Dickey-Fuller (ADF) - Fisher; Philipps and Perron (PP) - Fisher tests. The IPS test is proposed for dynamic heterogeneous panels based on the mean of individual unit root statistics. It proposes a standard test statistic based on the ADF statistics average across the cross section. Similar to the IPS test, both ADF-Fisher and PP-Fisher aim a combination of the significance of individual independent tests, with the difference between them regarding on how to address the serial correlation issue. However, the crucial difference is that the Fisher tests are based on combining the significance levels of the individual tests, whereas the IPS test is based on combining the test statistics.

Table 1: Panel Unit Root Tests

H_0: Individual Unit Root Process					
	c_t	Δc_t	$y_t - i_t - g_t$	$\Delta(y_t - i_t - g_t)$	CA_t
IPS	14.37	-49,50	10.69	-60.53	-8.85
	(1.00)	(0.00)	(1.00)	(0.00)	(0.00)
ADF - Fisher	216.17	2796.42	288.25	3458,17	828.17
	(1.00)	(0.00)	(0.87)	(0.00)	(0.00)
PP - Fisher	205.9	3163.7	258.59	3937.86	792.54
	(1.00)	(0.00)	(0.99)	(0.00)	(0.00)

Notes: This table reports the panel unit root tests considered in section 3.2 for consumption (c_t, Δc_t), national cash flow ($y_t - i_t - g_t$, $\Delta(y_t - i_t - g_t)$), and the actual consumption smoothing component CA_t. The first and second rows reports the IPS - equation (17) - statistics and p-values in parenthesis respectively. The third and forth rows reports the ADF-Fisher statistics - equation (18) - and p-values in parenthesis respectively. At last, fifth and sixth rows reports the ADF-PP statistics - equation (19) - and p-values in parenthesis respectively.

According to the results in Table 1 we do not reject the null hypothesis of unit root for c_t and z_t in level whereas we reject for first difference, confirming the point made visually that both variables are I(1). Regarding the CA_t serie, we reject the unit root hypothesis indicating stationarity of the variable as expected from the theoretical model. Therefore, we expect that c_t and $y_t - i_t - g_t$ are cointegrated as we found a linear combination of the variables which is stationary. This evidence will be formally tested.

The next step is to verify that c_t and z_t are cointegrated for all the countries. Table 2 reports three cointegration tests: Pedroni (2004), Kao (1999) and Fisher (1934). Pedroni (2004) is a residual-based panel cointegration test which allows heterogeneous cointegration vectors, with the null hypothesis that all of the individuals of the panel are not cointegrated. Kao (1999) is also a residual based test however it imposes a homogenous cointegration vector. The Fisher test was proposed by Maddala and Wu (1999) who combine p-values of individual cross-sectional Johansen (1991) tests for testing cointegration in the full panel.

Table 2: Cointegration Tests - Consumption (c_t) and net income ($y_t - i_t - g_t$)

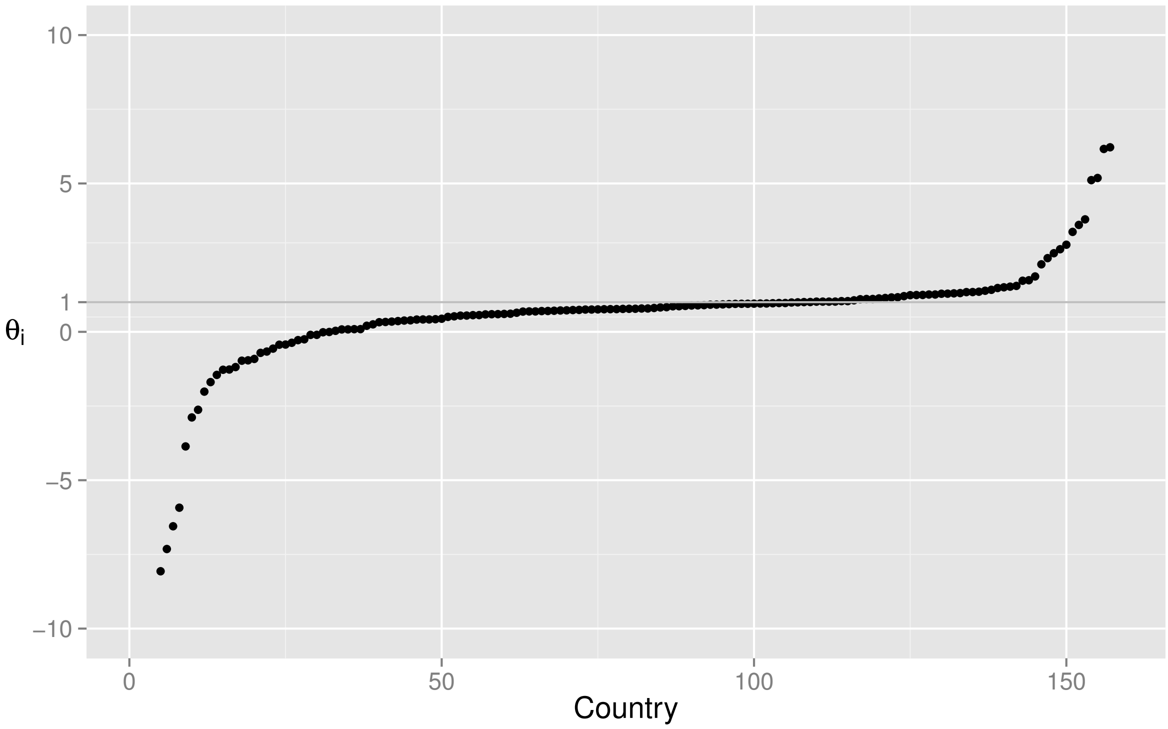
Fisher[1]			Pedroni[1]		Kao	
H_0:Number of Contegrations			H_0:No Contegration (Heterogeneous)		H_0:No Contegration (Homogeneous)	
			Panel v-Statistic	14.18	ADF - Statistic	-5.31
	trace	max-eigenvalue		(0.000)		(0.000)
None	721.60	660.90	Panel rho-Statistic	-19.48		
	(0.000)	(0.000)		(0.000)		
At most 1	345.10	345.10	Panel PP-Statistic	-15.33		
	(0.125)	(0.125)		(0.000)		
			Panel ADF-Statistic	-17.05		
				(0.000)		

Notes: This table reports the panel cointegration tests considered in section 3.3 for consumption (c_t) and net income ($y_t - i_t - g_t$). The column labeled Fisher reports the trace and max-eigenvalue statistics (p-values in parenthesis) to test the number of cointegrations the system has. The column labeled Pedroni reports the cointegration statistics in Pedroni (1999) (p-values in parenthesis) to test cointegration between consumption and net income. At last, the column labeled Kao reports the cointegration statistics in Kao (1999) (p-values in parenthesis) to test cointegration between consumption and net income.

[1] Lags chosen based on SIC: 0 to 9

The results of all tests are reported in Table 2. At 1% significance we reject the null hypothesis of no cointegration for both Pedroni and Kao tests. We also find evidence of cointegration when look at the results of Fisher's test, rejecting at 1% of significance the hypothesis of no cointegration for both trace and maximum eigenvalue statistics. Moreover, the hypothesis of at most one cointegration relation is not rejected.

Figure 2: Consumption Tilting Components



A plot of the estimated values of θ, which represents the degree of consumption tilting of a country, are showed in the Figure 2 below. The large majority of the countries have θ situated between -1,5 and 1,5. Also, we found that for 57% of the sample $-1 < \theta < 1$. As mentioned before, these countries are consuming more than its current permanent cash flow, whereas the others are tilting consumption towards the future. However, it's worth mention that many countries have θ close to 1, value that represents no consumption tilting.

Table 3: PVAR(p) Model Coefficients and statistics

	PVAR(1)		PVAR(2)		PVAR(3)	
Dependent Variables:	Δz_t	CA_t	Δz_t	CA_t	Δz_t	CA_t
Δz_{t-1}	-0.143	0.135**	-0.372**	-0.057	-0.415**	0.038**
	(0.110)	(0.002)	(0.143)	(0.217)	(0.162)	(0.005)
Δz_{t-2}	-	-	-0.405**	-0.510**	-0.520**	0.136
			(0.002)	(0.002)	(0.005)	(0.251)
Δz_{t-3}	-	-	-		-0.466	0.770**
					(0.387)	(0.002)
CA_{t-1}	-0.002	0.620**	0.000	0.610*	0.001	0.663**
	(0.110)	(0.002)	(0.177)	(0.278)	(0.265)	(0.003)
CA_{t-2}	-	-	0.000	-0.121**	-0.005	-0.085
			(0.002)	(0.002)	(0.004)	(0.422)
CA_{t-3}	-	-	-	-	-0.001	-0.061**
					(0.489)	(0.003)
Observations	$N = 158$	$T = 37$	$N = 158$	$T = 36$	$N = 158$	$T = 35$
Instruments	1406		1404		1400	
J-test	31.370		21.498		15.490	
	[1.000]		[1.000]		[1.000]	
Granger Causality	2.412		0.080		2.069	
	[0.120]		[0.961]		[0.558]	
PVM test	22583.2		206621.7		116361.4	
	[0.000]		[0.000]		[0.000]	
BIC	-28420.4		-28369.5		-28274.3	
AIC	-11208.6		-11194.5		-11160.5	
HQ	-59717.4		-59599.7		-59393.1	

Notes: This table reports GMM estimates of the parameters regarding the Panel VAR model considered in section 3.4.

Each column reports a pVAR system of order p with two variables - $\mathbf{w}_{i,t} = [\Delta z_{i,t}\ CA_{i,t}]'$. The first variable is the first difference of cash flow ($\Delta z_{i,t}$) and the later is the actual smoothing current account ($CA_{i,t}$) - computed using the cointegrating vector, α_i. The cointegrating vector is estimated separately for each country by means of likelihood estimation method, Johansen (1991).

The GMM estimates reported are all two step. Asymptotic standard errors are robust to general cross-section and time series heteroskedasticity and they are reported in parentheses.

The row labeled Granger Causality reports a granger causality test which is just Wald test that all coefficients on lagged current account variables in the cash flow equation are jointly zero.

The row labeled PVM test reports a statistical test of the Present Value Model. It is simply a Wald statistic for the joint restrictions on coefficients described on (16) hold. The test has a χ^2 distribution with degrees of freedom equal to the number of restrictions.

The J-test is an identification test that checks whether the errors of the system are orthogonal to all the instruments, Hansen (1982). It has an asymptotic χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions.

The estimation results of the PVAR systems are reported in Table 3. We tested three different specifications for the PVAR model, with one, two and three lags. From the J-test results reported in the table, we check that all the specifications fit the data well, as the null hypothesis that data come close to meeting the GMM restrictions is not rejected. Regarding which model should be chosen among the different specifications, results from the BIC, AIC and SC model selection criteria for GMM estimations, suggest the PVAR(1) is the more appropriate.

Also, we show the Granger causality test results in the table. We don't reject the null hypothesis of no Granger causality for the entire sample. This finding suggests that current account does not cause subsequent changes in national cash flow. Therefore, we reject one of the model propositions.

A final result is a formal statistical test of the model. We test the consumption-smoothing model and the perfect capital mobility hypothesis through a Wald test for the restrictions imposed by the model on the VAR coefficients, see equation (16). In our exercise we assume an international interest rate of 3%[5].The Wald test has a χ^2 distribution with degrees of freedom equal to the number of restrictions. The test confirms the previous finding that the model is statistically rejected for all the model specifications.

We perform two tests of robustness of the results found above. Arguably, the countries in the sample are differents in many aspects, including the degree of openness of the economy. Therefore, we undertake the tests above for subsamples of countries. We try to gather countries where intuitively capital mobility can be less limited.

First, in Table 4 we use an index of trade freedom from the Heritage foundation. This index is a composite measure of the absence of tariff and non-tariff barriers that affect imports and exports of goods and services. We select countries which index are above 80,0 points which results a pool of 41 countries. To avoid lack of power on the coefficients tests we use data only since 2000 to estimate the panel VAR systems, but we use the whole sample to estimate the actual smoothing current account. Overall the results are similar to those found for the whole sample, from the J-test in Table 4 we don't reject the null that the model specification fit the data well. However, one significant difference comes from the Granger causality test, which for this subsample the null hypothesis of no granger causality rejected.

[5]We tested various interest rates between 2 and 6%, however they gave very similar results

Table 4: PVAR(p) Model Coefficients and statistics - Trade of freedom

	PVAR(1)		PVAR(2)		PVAR(3)	
Dependent Variables:	ΔZ_t	CA_t	ΔZ_t	CA_t	ΔZ_t	CA_t
ΔZ_{t-1}	0.026	-0.372**	0.017	-0.757**	0.438	-1.473**
	(0.083)	(0.001)	(0.130)	(0.131)	(0.505)	(0.033)
ΔZ_{t-2}	-	-	0.072**	-0.651**	1.223**	-3.046**
			(0.001)	(0.003)	(0.020)	(0.376)
ΔZ_{t-3}	-	-	-		-0.187	1.066**
					(1.180)	(0.017)
CA_{t-1}	0.002	0.768**	0.003	1.102**	-0.020	0.975**
	(0.060)	(0.001)	(0.108)	(0.102)	(0.469)	(0.010)
CA_{t-2}	-	-	-0.004**	-0.439**	0.026**	-0.470
			(0.001)	(0.001)	(0.005)	(0.344)
CA_{t-3}	-	-	-	-	-0.025	0.086**
					(0.928)	(0.004)
Observations	N = 41	T = 7	N = 41	T = 6	N = 41	T = 5
Instruments						
J-test	12.488		12.743		8.327	
	[1.000]		[1.000]		[1.000]	
Granger Causality	3.256		13.765		20.263	
	[0.071]		[0.001]		[0.000]	
PVM Test	16486.3		44729.8		72549.8	
	[0.000]		[0.000]		[0.000]	
BIC	-1309.5		-1264.7		-1194.9	
AIC	-699.5		-675.3		-639.7	
HQ	-2763.8		-2669.9		-2518.4	

Notes: This table reports GMM estimates of the parameters regarding the Panel VAR model considered in section 3.4.

Each column reports a pVAR system of order p with two variables - $\mathbf{w}_{i,t} = [\Delta z_{i,t}\ CA_{i,t}]'$. The first variable is the first difference of cash flow $(\Delta z_{i,t})$ and the later is the actual smoothing current account $(CA_{i,t})$ - computed using the cointegrating vector, α_i. The cointegrating vector is estimated separately for each country by means of likelihood estimation method, Johansen (1991).

The GMM estimates reported are all two step. Asymptotic standard errors are robust to general cross-section and time series heteroskedasticity and they are reported in parentheses.

The row labeled Granger Causality reports a granger causality test which is just Wald test that all coefficients on lagged current account variables in the cash flow equation are jointly zero.

The row labeled PVM test reports a statistical test of the Present Value Model. It is simply a Wald statistic for the joint restrictions on coefficients described on (16) hold. The test has a χ^2 distribution with degrees of freedom equal to the number of restrictions.

The J-test is an identification test that checks whether the errors of the system are orthogonal to all the instruments, Hansen (1982). It has an asymptotic χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions.

In the second robustness test we follow a classification from the World Bank of the world's economies based on estimates of gross national income (GNI) per capita. In our exercise we select countries classified as "high income" countries, which have a GNI per capita of $12,616 or more. The total number of countries from this group in our dataset is 48, leading us to keep using data only since 2000. The results in Table 5 confirm the findings of the previous robustness test, giving results similar to the whole sample exercise except for the Granger causality test which also reject the null hypothesis that CA_t doesn't Granger causes $\Delta(q_t - i_t - g_t)$.

Table 5: PVAR(p) Model Coefficients and statistics - High Income

	PVAR(1)		PVAR(2)		PVAR(3)	
Dependent Variables:	ΔZ_t	CA_t	ΔZ_t	CA_t	ΔZ_t	CA_t
ΔZ_{t-1}	0.064**	0.135**	0.093**	0.174**	0.054**	0.201**
	(0.001)	(0.002)	(0.007)	(0.007)	(0.013)	(0.005)
ΔZ_{t-2}	-	-	-0.241**	-0.231**	-0.328**	-0.049**
			(0.004)	(0.003)	(0.007)	(0.014)
ΔZ_{t-3}	-	-	-		-0.158**	0.083**
					(0.017)	(0.004)
CA_{t-1}	-0.128**	0.497**	-0.102**	0.514**	-0.094**	0.481**
	(0.003)	(0.002)	(0.005)	(0.004)	(0.010)	(0.007)
CA_{t-2}	-	-	0.034**	0.000	0.064**	-0.079**
			(0.005)	(0.003)	(0.007)	(0.009)
CA_{t-3}	-	-	-	-	-0.031*	0.074**
					(0.013)	(0.005)
Observations	N = 48	T = 7	N = 48	T = 7	N = 48	T = 7
Instruments						
J-test	43.071		41.034		42.242	
	[1.000]		[1.000]		[0.998]	
Granger Causality	3424.057		1176.948		922.782	
	[0.000]		[0.000]		[0.000]	
PVM Test	6492.5		10204.0		8549.7	
	[0.000]		[0.000]		[0.000]	
BIC	-1335.1		-1290.7		-1212.0	
AIC	-668.9		-647.0		-605.8	
HQ	-2851.0		-2755.5		-2591.7	

Notes: This table reports GMM estimates of the parameters regarding the Panel VAR model considered in section 3.4.

Each column reports a pVAR system of order p with two variables - $\mathbf{w}_{i,t} = [\Delta z_{i,t} \ CA_{i,t}]'$. The first variable is the first difference of cash flow ($\Delta z_{i,t}$) and the later is the actual smoothing current account ($CA_{i,t}$) - computed using the cointegrating vector, α_i. The cointegrating vector is estimated separately for each country by means of likelihood estimation method, Johansen (1991).

The GMM estimates reported are all two step. Asymptotic standard errors are robust to general cross-section and time series heteroskedasticity and they are reported in parentheses.

The row labeled Granger Causality reports a granger causality test which is just Wald test that all coefficients on lagged current account variables in the cash flow equation are jointly zero.

The row labeled PVM test reports a statistical test of the Present Value Model. It is simply a Wald statistic for the joint restrictions on coefficients described on (16) hold. The test has a χ^2 distribution with degrees of freedom equal to the number of restrictions.

The J-test is an identification test that checks whether the errors of the system are orthogonal to all the instruments, Hansen (1982). It has an asymptotic χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions.

6 Conclusion

Previous tests of international capital mobility have found differente results according to the approach used. The most common saving and investment approach is arguably flawed, which has been questioned on theoretical and econometric grounds in the literature.

In our work, we follow an alternative criterion for measuring capital mobility. Following the intertemporal current account model developed in Sachs (1982), our test is based on the assumption that if capital is indeed mobile, then it should smooth consumption in the face of shocks to national cash flow. Therefore, the current account would act as a buffer to smooth consumption in the face of shocks to output, investment, and government expenditure. Although some works have followed this approach to test capital mobility (see Otto (1992);Ghosh (1995)),ours is novel in the sense that try to test capital mobility in a global manner, not only for specific countries. To achieve this aim, we perform a panel data analysis with 158 countries, which benefits from higher power tests and more precise estimates.

The solution of the model writes the optimal current account as the expected present discounted value of changes in output, investment, and government expenditure. This result is the typical present value equation(PVM) introduced by Campbell (1987) and Campbell and Shiller (1987), therefore we follow their techniques to estimate an unrestricted VAR (PVAR in our case) and perform statistical tests to validate the model implications on the data.

Considering the whole sample, we don't reject the null hypothesis of unit root for consumption $c_{i,t}$ and national cash flow $z_{i,t}$ in level whereas we reject for first difference. Also, regarding the current account $CA_{i,t}$ serie, we reject the unit root hypothesis and find evidences of cointegration between $c_{i,t}$ and $z_{i,t}$. These results are expected from the hypothesis of the theoretical model. However, we don't reject the null hypothesis of no Granger causality of $CA_{i,t}$ on $\Delta zi, t$ and reject the restrictions imposed by the model in the PVAR coeffcients. Therefore, we reject two of the model propositions. These results suggest absence of perfect capital mobility.

Finally, we perform two tests of robustness with subgroups of countries we believe could have less restrictions to capital mobility.One group gathering high income countries and the other the best ranked countries in trade freedom. The difference in the results from that obtained in the whole sample is that we reject the null hypothesis of no granger causality in both subgroups. This finding supports one of the model implications, which can be interpreted as evidence of capital mobility.

References

Adedeji, O. and J. Thornton (2008). International capital mobility: Evidence from panel cointegration tests. *Economics Letters 99*(2), 349 – 352.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), pp. 277–297.

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory 21*(4), pp. 795–837.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of Econometrics 108*(2), 343 – 363.

Campbell, J. Y. (1987). Does saving anticipate declining labor income? an alternative test of the permanent income hypothesis. *Econometrica 55*(6), pp. 1249–1273.

Campbell, J. Y. and R. J. Shiller (1987). Cointegration and tests of present value models. *Journal of Political Economy 95*(5), pp. 1062–1088.

Campbell, J. Y. and R. J. Shiller (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance 43*(3), pp. 661–676.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance 20*(2), 249 – 272.

Costantini, M. and L. Gutierrez (2013). Capital mobility and global factor shocks. *Economics Letters 120*(3), 513 – 515.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*(2), pp. 251–276.

Evans, P., B.-H. Kim, and K.-Y. Oh (2008). Capital mobility in saving and investment: A time-varying coefficients approach. *Journal of International Money and Finance 27*(5), 806 – 815.

Feldstein, M. and C. Horioka (1980). Domestic saving and international capital flows. *The Economic Journal 90*(358), pp. 314–329.

Fisher, R. A. (1934). *Statistical methods for research workers.* Edinburgh.

Ghosh, A. R. (1995). International capital mobility amongst the major industrialised countries: too little or too much? *The Economic Journal*, 107–128.

Ghosh, A. R. and J. D. Ostry (1995). The current account in developing countries: A perspective from the consumption-smoothing approach. *The World Bank Economic Review 9*(2), 305–333.

Guillén, O. T., A. Hecq, J. V. Issler, and D. Saraiva (2015). Forecasting multivariate time series under present-value model short and long-run co-movement restrictions. *International Journal of Forecasting 31*(3), 862–875.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal 3*(2), 148–161.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), pp. 1029–1054.

Heston, A., R. Summers, and B. Aten (2012). Penn world table version 7.1. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania*.

Hoffmann, M. (2004). International capital mobility in the long run and the short run: can we still learn from saving–investment data? *Journal of International Money and Finance 23*(1), 113–131.

Holmes, M. J. and J. Otero (2014). Re-examining the feldstein–horioka and sachs' views of capital mobility: A heterogeneous panel setup. *International Review of Economics & Finance 33*, 1 – 11.

Im, K. S., M. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics 115*(1), 53 – 74.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59*(6), pp. 1551–1580.

Johansen, S. and A. Swensen (2011). On a graphical technique for evaluating some rational expectations models. *Journal of Time Series Econometrics 3*(1), 1–29.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics 90*(1), 1 – 44.

Kumar, S. (2015). Regional integration, capital mobility and financial intermediation revisited: Application of general to specific method in panel data. *Journal of International Financial Markets, Institutions and Money 36*, 1 – 17.

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

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics 14*(5), 563–577.

Maddala, G. S. and S. Wu (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics 61*(S1), 631–652.

Otto, G. (1992). Testing a present-value model of the current account: Evidence from us and canadian time series. *Journal of International Money and Finance 11*(5), 414 – 430.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and statistics 61*(S1), 653–670.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric Theory*, 597–625.

Pelgrin, F. and S. Schich (2008). International capital mobility: What do national saving–investment dynamics tell us? *Journal of International Money and Finance 27*(3), 331 – 344.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory 5*(2), pp. 181–240.

Sachs, J. (1982). The current account in the macroeconomic adjustment process. *The Scandinavian Journal of Economics 84*(2), pp. 147–159.