Permanent-Transitory Decompositions Under Present Value Model Restrictions

Alain Hecq
Maastricht University

Joao Victor Issler
Getulio Vargas Foundation

June 26, 2011

Abstract

In this paper we revisit present-value (PV) relationships in economics in the context of vector-autoregressive (VAR) models. Whereas the early literature on PV relationships has exploited the existence of long-run restrictions in VAR parameters (cointegration), e.g., Campbell and Shiller (1987), Campbell and Deaton (1989), Campbell (1991), *inter alia*, which is reviewed in Engsted (2002), now added with the interesting contribution of Johansen and Swensen (2011), our focus is the existence of short-run restrictions (common-cyclical features) in VAR parameters. The so-called common-cycle literature has evolved after the initial effort in Engle and Kozicki (1993), Vahid and Engle (1993, 1997) and Cubadda and Hecq (2001), and now includes the more flexible approach in Hecq, Palm, and Urbain (2006) and Athanasopoulos, Guillén, Issler and Vahid (2011).

Our paper has several novel contributions, but two stand out. First, we show that PV relationships entail a weak-form common feature relationship as in Hecq, Palm, and Urbain and in Athanasopoulos, Guillén, Issler and Vahid and also a polynomial serial-correlation common feature relationship as in Cubadda and Hecq. That represents restrictions on dynamic models (VARs) which allow new tests for the existence of PV relationships to be developed. Because these relationships occur mostly with financial data, we propose novel tests based on generalized method of moment (GMM) estimates, where it is straightforward to propose robust tests in the presence of heteroskedasticity and serial-correlation in the data. Their good performance is confirmed in a Monte-carlo exercise. Second, in the context of asset pricing, we propose applying a permanent-transitory (PT) decomposition based on Beveridge and Nelson (1981), which focus on extracting the long-run component of asset prices, a key concept in modern financial theory as discussed in Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Nieuwerburgh, Lustig, Verdelhan (2010).

The techniques discussed here were applied to annual long- and short-term interest rates for the UK, ranging from 1831 to 2008. Unit-root tests do not reject the null that rates in levels are $I(1)$. There is one cointegrating relationship, which does not from $(1-1)$ in statistical tests. Using our proposed robust GMM strategy, we do not reject the existence of a common cyclical feature vector linking these two series. Several exercises along the lines of Issler and Vahid (2001) are performed. Extracting the long-run component of short-run rates show the usefulness of our approach.

JEL: C22, C32

Keywords: present value, common cycles, interest rates

1 Motivation

Since Campbell and Shiller (1987), it is well known that cointegration between the level of the variables is a necessary requirement to asses the empirical validity of the present value model (PVM hereafter). Moreover, from the orthogonality condition and cross-equation restrictions that this type of model entails, there also exist necessary implications about the presence of common cycles (Engle and Kozicki, 1993; Vahid and Engle, 1993). This will be the starting point this article.

In this paper we revisit present-value (PV) relationships in economics in the context of vector-autoregressive (VAR) models. Whereas the early literature on PV relationships has exploited the existence of long-run restrictions in VAR parameters (cointegration), e.g., Campbell and Shiller (1987), Campbell and Deaton (1989), Campbell (1991), *inter alia*, which is reviewed in Engsted (2002), now added with the interesting contribution of Johansen and Swensen (2011), our focus is the existence of short-run restrictions (common-cyclical features) in VAR parameters. The so-called common-cycle literature has evolved after the initial effort in Engle and Kozicki (1993), Vahid and Engle (1993, 1997) and Cubadda and Hecq (2001), and now includes the more flexible approach in Hecq, Palm, and Urbain (2006) and Athanasopoulos, Guillén, Issler and Vahid (2011).

Our paper has several novel contributions, but two stand out. First, we show that PV relationships entail a weak-form common feature relationship as in Hecq, Palm, and Urbain and in Athanasopoulos, Guillén, Issler and Vahid and also a polynomial serial-correlation common feature relationship as in Cubadda and Hecq. That represents restrictions on dynamic models (VARs) which allow new tests for the existence of PV relationships to be developed. Because these relationships occur mostly with financial data, we propose novel tests based on generalized method of moment (GMM) estimates, where it is straightforward to propose robust tests in the presence of heteroskedasticity and serial-correlation in the data. Their good performance is confirmed in a Monte-carlo exercise. Second, in the context of asset pricing, we propose applying a permanent-transitory (PT) decomposition based on Beveridge and Nelson (1981), which focus on extracting the long-run component of asset prices, a key concept in modern financial theory as discussed in Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Nieuwerburgh, Lustig, Verdelhan (2010).

At the model representation level, looking at the VECM representation of the original series instead of the transformed VAR à la Campbell-Shiller we emphasize that the PV Models can also be equivalently tested using the weak form common feature framework (WF hereafter, see Hecq *et al.*, 2000, 2006, and Athanasopoulos, Guillén, Issler and Vahid (2011)). We also show that small modifications in the timing of the present value representation may change the cross-equation restrictions but not the orthogonality condition or the reduced rank properties. This is the reason why we propose to testing for the presence of common cycles first and than look at additional restrictions on parameters later.

In terms of testing for PV restrictions in dynamic models, we discuss two different approaches, which are compared using a small Monte Carlo experiment. The first is a limited information approach (e.g. GMM) and the other is a full information framework (reduced rank regression). It turns out that in the presence of time varying conditional heteroskedasticity (e.g. GARCH) only the robust GMM à la White does not have size distortions. A non-robust GMM and a LR test are strongly oversized while a Newey-West correction is undersized.

In discussing the long-run properties of asset prices, given that we have put forward the numerical equivalence of the transformed VAR framework proposed by Campbell and Shiller (1987) and the VECM representation with of course cointegration but also common cyclical feature restrictions, we propose to apply the multivariate Beveridge-Nelson permanent transitory decomposition.

The techniques discussed here were applied to annual long- and short-term interest rates for the UK, ranging from 1831 to 2008. Unit-root tests do not reject the null that rates in levels are $I(1)$. There is one cointegrating relationship, which does not from $(1-1)$ in statistical tests. Using our proposed robust GMM strategy, we do not reject the existence of a common cyclical feature vector linking these two series. Several exercises along the lines of Issler and Vahid (2001) are performed. Extracting the long-run component of short-run rates show the usefulness of our approach.

2 A present value equation

Let us consider a very simple present value equation

$$Y_t = \theta(1-\delta)\sum_{i=0}^{\infty}\delta^i \mathbb{E}_t y_{t+i}, \tag{1}$$

which states that Y_t is a linear function of the present discounted value of the expected future y_t. In most cases Y_t and y_t are I(1) variables and are respectively long and short-term interest rates, stock prices and dividends, personal consumption and labor disposal income, etc (see Engsted, 2002). We assume constant expected returns with a discount factor $\delta = \frac{1}{1+r}$. θ is a factor of proportionality with for instance $\theta = \delta/(1-\delta)$ in the price/dividend relationship and $\theta = 1$ for the interest rates case (see Campbell and Shiller, 1987; Chow, 1984). The choice of θ will only impact the value of the long-run coefficient in the cointegrating relationship, hence we take for the presentation $\theta = \delta/(1-\delta)$

$$Y_t = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}. \tag{2}$$

Following *inter alia* Campbell and Shiller (1987), the actual spread is defined as

$$S_t \equiv Y_t - \frac{\delta}{1-\delta} y_t, \tag{3}$$

where S_t is I(0) if series Y_t and y_t are cointegrated. Subtracting $\frac{\delta}{1-\delta} y_t$ from both sides of (2) produces the theoretical spread[1] $S_t^{'}$

$$S_t^{'} = \frac{\delta}{1-\delta} \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i}. \tag{4}$$

Further, subtracting from Y_t in (2) $\delta E_t Y_{t+1} = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i+1}$ we obtain

$$Y_t = \delta E_t Y_{t+1} + \delta y_t \tag{5}$$

which leads, if one adds and subtracts in (5) δY_t, to $Y_t = \delta E_t \Delta Y_{t+1} + \delta Y_t + \delta y_t$ or $(1-\delta) Y_t = \delta E_t \Delta Y_{t+1} + \delta y_t$. Finally we obtain

$$S_t^{''} = \frac{\delta}{1-\delta} \mathbb{E}_t \Delta Y_{t+1}. \tag{6}$$

Equation (6) gives the spread as a function of one-step ahead forecasts of ΔY_{t+1}. Without loss of generality we have that

$$\Delta Y_{t+1} = \mathbb{E}_t \Delta Y_{t+1} + \underbrace{(\Delta Y_{t+1} - \mathbb{E}_t \Delta Y_{t+1})}_{u_{1t+1}}. \tag{7}$$

Plugging (7) in (6) and lagging the whole equation by one period we have

$$S_{t-1} = \frac{\delta}{1-\delta} \Delta Y_t + u_t \tag{8}$$

[1] For me to remember: $S_t = E_t(-\frac{\delta}{1-\delta} y_t + \frac{\delta(1-\delta)}{1-\delta} y_t + \frac{\delta^2(1-\delta)}{1-\delta} y_{t+1} + \frac{\delta^3(1-\delta)}{1-\delta} y_{t+2} + \frac{\delta^4(1-\delta)}{1-\delta} y_{t+3} ...)$
$= \frac{\delta}{1-\delta} E_t(-\delta y_t + \delta(1-\delta) y_{t+1} + \delta^2(1-\delta) y_{t+2} ...$
$= \frac{\delta}{1-\delta} E_t(\delta \Delta y_{t+1} + \delta^2 \Delta y_{t+2} + ...)$

or alternatively

$$\Delta Y_t = \frac{1-\delta}{\delta} S_{t-1} + v_t \tag{9}$$

where u_t or v_t is orthogonal to the past. From (8) we also have the Campbell and Shiller (1987) relationships, namely

$$\begin{aligned}
(1-\delta)S_{t-1} &= \delta \Delta Y_t + (1-\delta)u_t \\
S_{t-1} - \delta Y_{t-1} + \delta \frac{\delta}{1-\delta} y_{t-1} &= \delta Y_t - \delta Y_{t-1} + \left\{ \delta \frac{\delta}{1-\delta} y_t - \delta \frac{\delta}{1-\delta} y_t \right\} + (1-\delta)u_t \\
S_{t-1} &= \delta S_t + \delta \frac{\delta}{1-\delta} \Delta y_t + (1-\delta)u_t
\end{aligned}$$

which gives

$$S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta} \Delta y_t + \varepsilon_t \tag{10}$$

with $\varepsilon_t = \frac{(1-\delta)}{\delta} u_t$. Next section makes the link between (9) and (10) on the one hand and the common cyclical feature literature on the other hand.

3 Common cyclical feature restrictions: Model representation

Let us assume that the bivariate system for the I(1) series $(Y_t, y_t)'$ follows a VAR(p) and the $S_t = Y_t - \beta y_t$ with $\beta = \frac{\delta}{1-\delta}$ is the stationary disequilibrium error term. The corresponding VECM representation (see e.g. Johansen, 2006) is given by

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \Gamma_1 \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + ... + \Gamma_{p-1} \begin{pmatrix} \Delta Y_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} S_{t-1} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{11}$$

where we assume that the disturbance terms are multivariate *iid* and that conditions for avoiding I(2)-ness are met. The Γ_i are short-run coefficient matrices, $\acute{\alpha}s$ are the loadings. Looking at Equation (9) we can formulate the following proposition.

Proposition 1 *Equation $\Delta Y_t = \frac{1-\delta}{\delta} S_{t-1} + v_t$ in (9) is a weak form common feature relationship (see Hecq et al., 2006) for the VECM (11) if there exists in the bivariate system a vector $\acute{\gamma}$ with $\acute{\gamma}\Gamma_1 = \acute{\gamma}\Gamma_2 = ... = \acute{\gamma}\Gamma_{p-1} = 0$ but with $\acute{\gamma}\alpha \neq 0$. Moreover to have a trivial common feature relationship corresponding to (9) such that $\acute{\gamma} = (1:0)$, the first row of every Γ_i, $i = 1...p-1$ must be zero.*

The usual cross-equation restriction within the VAR obtain in (10) and proposed by Campbell and Shiller (1987) can also be seen from a transformation of the VECM instead of the usual ad hoc VAR (see also Johansen and Swensen (2011) on this point). Let us write for $X_t = (Y_t, y_t)'$ the following VECM

$$\Gamma(L)\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t, \quad t = 1, ..., T. \tag{12}$$

To obtain the transformed VAR representation, we denote $C = (\beta : M)'$ the $n \times n$ nonsingular matrix where $M = (0_{(n-r)\times r}\ I_{(n-r)})$ is the selection matrix such that $M'\Delta X_t = \Delta y_t$. Premultiplying both sides by C we get $C\Gamma(L)\Delta X_t = C\alpha\beta' X_{t-1} + C\varepsilon_t$ or if we develop and rearrange the first block by putting $-\beta' X_{t-1}$ in the right-hand-side

$$\beta' X_t = (I_r + \beta'\alpha)\beta' X_{t-1} + \beta'\underline{\Gamma}(L)\Delta X_{t-1} + \beta'\varepsilon_t, \tag{13}$$

$$\Delta y_t = (M\alpha)\beta' X_{t-1} + \underline{\Gamma}(L)\Delta y_{t-1} + M\varepsilon_t, \tag{14}$$

where $\underline{\Gamma}(L) = \Gamma(L) - I_n$ and with $S_t = \beta' X_t$ we have

$$\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \underline{B_{11}}(L) & \underline{B_{12}}(L) \\ \underline{B_{21}}(L) & \underline{B_{22}}(L) \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \xi_t \tag{15}$$

where $\underline{B_{11}}(L)$ and $\underline{B_{21}}(L)$ are of order $p-1$ and $\underline{B_{12}}(L)$ and $\underline{B_{22}}(L)$ or order $p-2$. This is the correct VAR derived from the VECM where we see that the last column for the lag value of Δy_t has a coefficient of zero. This gives a VAR of order p both in S_t and in Δy_t in which the coefficients of $\Delta y_{t-p} = 0$. It is in this matrix $B(L)$ or more precisely $\underline{B}(L) = B(L) - I_n$ that people look for cross-equation constraints. This gives rise to the following proposition

Proposition 2 *Equation $S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta}\Delta y_t + \varepsilon_t$ in (10) is a polynomial serial correlation common feature relationship (see Cubadda and Hecq, 2001) for the transformed VAR (15) if there exists in the bivariate system a vector $\tilde{\gamma}_0'$ with $\tilde{\gamma}_0' B_2 = ... = \tilde{\gamma}_0' B_p = 0$ but with $\tilde{\gamma}_0' B_1 = \tilde{\gamma}_1' \neq 0$. Moreover in the PVM $\tilde{\gamma}_0' = (1 : \frac{\delta}{1-\delta})$ and $\tilde{\gamma}_1' = (-\frac{1}{\delta} : 0)$.*

Remark 3 *Due to timing consideration, some authors start with a slight modification of the PVM. Compare for instance Campbell and Shiller (1987) with Campbell, Lo and MacKinlay (1997). If like these latter authors or Johansen and Swensen (2011) one starts to discount after the contemporaneous value of y_t, i.e.*

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

we preserve the reduced rank (orthogonality) conditions associated with the PVM in the VECM or the transformed VAR. However and, using the same kind of algebra we used beneath ,the relation-

ships are respectively

$$WF \quad : \quad \Delta Y_t = -\Delta y_t + \frac{1-\delta}{\delta} S_{t-1} + u_t, \tag{16}$$

$$PSCCF \quad : \quad S_t = -\frac{1}{(1-\delta)} \Delta y_t + \frac{1}{\delta} S_{t-1} + v_t,^2 \tag{17}$$

Once again these two relationships are equivalent; they provide the same likelihood within respectively the VECM and the transformed VAR (if the last column of coefficients is correctly restricted to zero). It is seen that in the PSCCF there is only a modification in the parameter value of Δy_t. For the WF relationship however, there is now the introduction of $-\Delta y_t$ which is associated with a common feature vector $\gamma' = (1 : 1)$ and there is no need to have lines of zeros in the $\Gamma_i, i = 1...p-1$ anymore. Note finally that Johansen and Swensen (2011) compare a Wald test for $\Delta Y_t = -\Delta y_t + \frac{1-\delta}{\delta} S_{t-1} + u_t$ with the restricted version of Campbell/Shiller $S_t = -\frac{\delta}{(1-\delta)} \Delta y_t + \frac{1}{\delta} S_{t-1} + v_t$, which is not strictly speaking correct.

4 Testing for common cycles

In order to test for the WF, one first has to determine the VAR(p) order for the joint I(1) process $(Y_t, y_t)'$. Then a test for cointegration using for instance the trace test (see Johansen, 1996) or the Engle and Granger (1987) regression is carried out and $\hat{S}_t = Y_t - \hat{\beta} y_t$ is computed. Finally a likelihood ratio (reduced rank regression) test for the weak form common feature in the VECM of lag order $p-1$ can be undertaken using the canonical correlation

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

where $CanCor\{X, W \mid R\}$ denotes the computation of canonical correlation between the two sets X and W, concentrating out the effect of R (deterministic terms and disequilibrium error term) by multivariate least squares. The previous program is equivalent in order to

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

which is more convenient to obtain the coefficient of $\hat{S}_{t-1}$ in (9). The likelihood ratio (LR) test for the null hypothesis that there exist at least s common feature vectors is

$$LR = -T\sum_{i=1}^{s}\ln(1-\hat{\lambda}_i), \qquad s = 1,\ldots,n. \tag{18}$$

The unrestricted model having $4(p-1)+2$ parameters to estimate, the restricted $2(p-1)+2+1$, the number of restrictions when testing that there exists one WF common feature is $2(p-1)-1 = 2p-3$ for $p > 1$.[3] We can obtain the same statistics by computing twice the difference between the log-likelihood in the VECM and in the pseudo structural form estimated by FIML (See e.g. Vahid and Issler (2001)) is

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

For the transformed VAR the restriction underlying the restricted PSCCF might be tested using

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

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

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

$$\begin{aligned}
&\begin{pmatrix} 1 & -\tilde{\gamma}_{0,1} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} S_t \\ \Delta y_t \end{pmatrix} \\
&= \begin{pmatrix} \tilde{\Gamma}_{1a} \\ \tilde{\Gamma}_{1b} \end{pmatrix}\begin{pmatrix} S_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} 0 \\ \tilde{\Gamma}_{p-1} \end{pmatrix}\begin{pmatrix} S_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \tilde{\Gamma}_{p,p} & 0 \end{pmatrix}\begin{pmatrix} S_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + v_t
\end{aligned}$$

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

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

Note that this set of restrictions are close to the ones in Campbell and Shiller (1987). The difference is that these authors do not impose zeros in the last matrix and that they add further conditions on the parameters coming from the theory to link γ and ϑ_1. However if our LR test rejects the null the additional parameter restrictions won´t be accepted. For instance the presence of measurement errors, a definition of the timing or data revisions can lead to a determination of the correct rank but with parameters that might differ from the theory. Hence we first propose to test for the rank reduction before imposing the constraint on the parameters in order to split both effects.

There is a GMM counterpart to these relationships. The common feature null hypothesis consists of an orthogonality condition between a combination of variables and the past of the series. Consequently 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 condition for γ' being a common feature vector corresponds in the WF to the orthogonality of the linear combination $\Delta Y_t - \gamma_1 \Delta y_t - \gamma_2 \hat{S}_{t-1}$ with the past information of the process. In terms of moment conditions, it can be expressed as

$$g_T(\theta; \Delta Y_t, \Delta y_t, \hat{S}_{t-1}, W_t) = g_T(\theta, .) = E([\Delta Y_t - \gamma_1 \Delta y_t - \gamma_2 \hat{S}_{t-1}] \otimes W_t') = 0. \tag{19}$$

This GIVE estimator is simply the 2SLS or the IV estimator when the instruments are the past of the series, namely

$$\hat{\theta}_{GIVE} = \left(\mathbf{\Delta X'W}(\mathbf{W'W})^{-1}\mathbf{W'\Delta X}\right)^{-1}\left(\mathbf{\Delta X'W}(\mathbf{W'W})^{-1}\mathbf{W'\Delta Y}\right), \tag{20}$$

with $\mathbf{\Delta X}_t = (\Delta y_t, \hat{S}_{t-1}, 1)'$. The validity of the orthogonality condition and consequently the presence of a common feature vector is obtained via an overidentification test *à la* Hansen, namely

$$J_1(\theta) = Tg_T(\theta;.)' P_T^{-1} g_T(\theta;.),$$

whose empirical counterpart is

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

The variance covariance matrix of the orthogonality condition has under usual regularity properties 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 $\mathbf{i} = (1...1)'$) because we do not want to impose that the common feature vector also annihilates the constant vector.

Both GIVE and LR estimators assume homoskedasticity, i.e. that the variance is constant through time. This is clearly unsatisfactory. Consequently we also use a robust GMM test statistics and we propose to extend the GIVE estimator by using the White's H.C.S.E. estimator such that (see Hamilton, 1994 or Hayashi, 2001)

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

where the only difference with the usual $\hat{\theta}_{GIVE}$ estimation is the presence of an additional matrix $\mathbf{B}$ constructed such that

$$\mathbf{B} = \begin{pmatrix} u_1^2 & 0 & \cdots & 0 \\ 0 & u_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & u_T^2 \end{pmatrix},$$

with $u_t = \Delta Y_t - \hat{\gamma}_{1IV}\Delta y_t - \hat{\gamma}_{2IV}\hat{S}_{t-1}$, $t = 1 \ldots T$, the residuals obtained under homoskedasticity using the GIVE estimation in a first step. We may then 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 test robust to heteroscedasticity

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

A third version apply the Newey-West principle and gives a $J_3(\theta_{GMM})$.

A small Monte Carlo simulation might help to advise the use of one of these tests in our framework. We use $T = 100$ and 500 observations; 2000 replications are run for respectively $p = 2, 3$ and 5 in the estimated model. The lag length in the DGP is $p = 3$. The DGP that ensure to have $\gamma' = (1 : 0)$ is

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

where in DGP#1 the disturbance terms are bivariate normal with a unit variance and a correlation of 0.5; in DGP#2 the disturbance terms are governed by a bivariate GARCH process with a yesterday news coefficient of 0.25, a coefficient of persistence of 0.74 and a long run variance equals 0.01. Note that the yesterdays news coefficient might be larger than what is encountered empirically (between 0.10 and 0.15). It is seen that the theoretical coefficient in the relationship like $\Delta Y_t = -\gamma_1 \Delta y_t + \gamma_2 S_{t-1} + u_t$ are $\gamma_1 = 0$ and $\gamma_2 = 1$. Table 1 reports the frequency with which the null is rejected at a 5% significance level (nominal size). It emerges that although in the iid case the behavior of the four tests is rather similar except that the LR is slightly oversized in small samples (see also Hecq et al. (2006)) and the Newey-West type of correction is undersized for $T = 100$. Things get much more worse in the presence of time varying conditional variance however. Only test J_2, namely a robust-White GMM, must be advocated. The other three have very bad properties. As far as the estimation of the coefficients are concerned now we do not notice any systematic bias. For DGP# the estimated coefficients for the mean value of γ_1 over 2000 replications ranges over the different specifications on p and T from -0.001 to 0.008 in the worst case while those from the mean value of γ_2 are between 0.978 ($p = 5, T = 100$) and 0.999. There are no major differences in the GARCH case as expected. About the standard deviation over the simulations, it lies for DGP#1 between 0.021 ($p = 3, T = 500$) and 0.054 ($p = 5, T = 100$) for γ_1 and between 0.032 and 0.077 for γ_2 for the same specifications. In the presence of GARCH for the DGP#2 these standard deviations are in general higher.[4]

5 PT decomposition under cointegration and common cycles

5.0.1 PV Theory and PT decomposition

We start with the seminal approach contained in Campbell and Shiller (1987), where, for ease of exposition, we focus on the PV model for stock prices and dividends, although our results are much

[4] Detailed results can be obtained upon request.

DGP	$VAR(p)$	LR		J_1		J_2		J_3	
		100	500	100	500	100	500	100	500
DGP#1:iid	2	5.2	5.75	5.05	5.75	5.05	5.65	4.25	5.15
	3	7.15	5.60	6.25	5.45	5.65	5.45	3.75	5.2
	5	8.5	6.5	6.4	6.15	5.3	5.85	0.75	4.75
DGP#2:GARCH	2	11.8	23.6	11.5	23.5	4.65	4.45	2.25	3.6
	3	14.4	32.9	13.5	32.6	4.7	5.3	2.0	2.95
	5	17.8	41.1	14.9	40.4	3.7	4.15	0.05	1.4

Table 1: Monte Carlo simulations: Common feature tests statistics

broader and indeed apply to any PV model involving two economic series. In terms of notation, let $\theta = \frac{\delta}{1-\delta}$,where δ is a discount factor of next period's utility function in deffinig welfare for a representative consumer. Variable Y_t represents the price of a given stock and y_t represents the stock dividend. The present-value model is:

$$Y_t = \theta (1-\delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t \ y_{t+i} + c \tag{21}$$

Subtracting θy_t from (21), we obtain:

$$\begin{aligned} Y_t - \theta y_t &\equiv S_t \\ &= \theta (1-\delta) \ y_t + \theta (1-\delta) \ \delta E_t \ y_{t+1} + \theta (1-\delta) \ \delta^2 \mathbb{E}_t y_{t+2} + \cdots - \theta y_t \\ &= \theta y_t - \theta \delta y_t + \theta \delta E_t y_{t+1} - \theta \delta^2 \mathbb{E}_t y_{t+1} - \theta \delta^2 \mathbb{E}_t y_{t+2} - \theta \delta^3 \mathbb{E}_t y_{t+2} + \cdots - \theta y_t \\ &= \theta \delta \ \mathbb{E}_t (y_{t+1} - y_t) + \theta \delta^2 \mathbb{E}_t (y_{t+2} - y_{t+1}) + \theta \delta^3 \mathbb{E}_t (y_{t+3} - y_{t+2}) + \cdots \end{aligned}$$

or,

$$S_t = \theta \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i} + c. \tag{22}$$

Forward (21) one period and take $\mathbb{E}_t$ to obtain:

$$\mathbb{E}_t \ Y_{t+1} = \theta (1-\delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t \ y_{t+1+i} + c. \tag{23}$$

Subtract now (21) from (23) to get:

$$\begin{aligned}\mathbb{E}_t\left(Y_{t+1}-Y_t\right) &= \mathbb{E}_t\ \Delta Y_{t+1} \\ &= \theta\left(1-\delta\right)\sum_{i=1}^{\infty}\delta^{i-1}\mathbb{E}_t\ \Delta y_{t+i} \\ &= \frac{1-\delta}{\delta}\left(S_t-c\right). \end{aligned} \tag{24}$$

This shows that we can express S_t in a regression equation setup as:

$$\begin{aligned} S_t &= \frac{\delta}{1-\delta}\ \mathbb{E}_t\Delta Y_{t+1}+c \\ &= \frac{\delta}{1-\delta}\ \Delta Y_{t+1}+c+\underbrace{\left(\Delta Y_{t+1}-\mathbb{E}_t\ \Delta Y_{t+i}\right)}_{\text{Regression Unforcastable Error}}. \end{aligned}$$

In what follows, we discuss a Beveridge and Nelson (1981) (BN) Decomposition for stock prices, which aims to separate the long-run price from its transient fluctuations. We start by defining the permanent Y_t^p and the transient Y_t^c component of stock prices, which obviously must add up to the current price Y_t:

$$Y_t = Y_t^p + Y_t^c.$$

The basis of the BN decomposition is that the permanent component is defined to be a random variable to which the price will converge to, taking the current period t as a starting point. In mathematical terms, we have:

$$Y_t^p = Y_t + \lim_{k\to\infty}\mathbb{E}_t\sum_{i=1}^{k}\Delta Y_{t+i} = Y_t + \mathbb{E}_t\sum_{i=1}^{\infty}\Delta Y_{t+i}. \tag{25}$$

Obviously, in defining the permanent component or trend, we are not implicitly or explicitly assuming any specific data-generating process for stock prices, although in the empirical implementation we will assume that Y_t contains one unit root in its auto-regressive polynomial. Therefore, at least in this section, the discussion is very general and applies to any type of stock prices, regardless of its time-series properties.

We now focus on the transitory or cyclical component of stock prices using the trend definition in (25):

$$Y_t^c = Y_t - \left(Y_t + \mathbb{E}_t \sum_{i=1}^{\infty} \Delta Y_{t+i} \right) \tag{26}$$

$$= -\sum_{i=1}^{\infty} \mathbb{E}_t \ \Delta Y_{t+i}. \tag{27}$$

We now recall (24) and combine it with (27) to obtain an intuitive expression for the transitory component of asset prices:

$$Y_t^c = -\frac{1-\delta}{\delta} \sum_{i=0}^{\infty} \mathbb{E}_t \left(S_{t+i} - c \right). \tag{28}$$

Equation (28) links the transitory component with expected current and future imbalances in the long-run relationship between asset prices and dividends summarized in $S_t - c$. As long as these centered expected imbalances are nil, so will be the transitory component, therefore asset prices will be identical their permanent component or long-run prices. As far as we know, this is the first paper to link these two variables, although a lot of work has been done lately on the issue of long-run prices in modern financial theory; see, *inter-alia*, Lettau and Ludvigson (2001, 2011), Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Nieuwerburgh, Lustig, and Verdelhan (2010).

Using (28) is straightforward to calculate the cyclical innovation:

$$Y_t^c - \mathbb{E}_{t-1} Y_t^c = -\frac{1-\delta}{\delta} \sum_{i=0}^{\infty} \left(\mathbb{E}_t - \mathbb{E}_{t-1} \right) S_{t+i}.$$

As shown above, there exists a VAR on $\binom{S_t}{\Delta y_t}$, that can be put in companion form:

$$W_t = \mathbf{A} \ W_{t-1} + V_t,$$

the vector W_t stacks the variables of interest S_t and Δy_t at different lags, and the matrix $\mathbf{A}$ is the companion matrix of the original VAR, where some restrictions must apply to some of its coefficients as discussed above. Using simple PV formulas, we are now ready to compute the cyclical component:

$$Y_t^c = -\frac{1-\delta}{\delta} \left(\sum_{i=0}^{\infty} e_1^{'} \mathbf{A}^i \right) W_t \tag{29}$$

$$= -\frac{1-\delta}{\delta} e_1^{'} \left(\mathbf{I} - \mathbf{A} \right)^{-1} W_t \tag{30}$$

where e_1 is a selection vector, compatible with W_t, with unity in its first component and zeros elsewhere, and $\mathbf{I}$ is an identity matrix whose order matches the number on rows of W_t. Long-run asset prices are given by:

$$Y_t^p = Y_t - Y_t^c = Y_t + \frac{1-\delta}{\delta} e_1^{'} (\mathbf{I} - \mathbf{A})^{-1} W_t. \tag{31}$$

Form (31) and (30) we can compute the respective innovations to trends and cycles, respectively:

$$\begin{aligned} Y_t^p - \mathbb{E}_{t-1} Y_t^p &= Y_t - Y_t^c = (Y_t - \mathbb{E}_{t-1} Y_t) + \frac{1-\delta}{\delta} e_1^{'} (\mathbf{I} - \mathbf{A})^{-1} W_t, \text{ and,} \\ Y_t^c - \mathbb{E}_{t-1} Y_t^c &= -\frac{1-\delta}{\delta} e_1^{'} (\mathbf{I} - \mathbf{A})^{-1} V_t. \end{aligned}$$

Finally, it is worth comparing Y_t^c with S_t. While from (22), $S_t - c$ is:

$$S_t - c = \theta \delta e_2^{'} A (I - \delta A)^{-1} W_t,$$

where e_2 is a selection vector, compatible with W_t, with unity in its second component and zeros elsewhere, Y_t^c equals $-\frac{1-\delta}{\delta} e_1^{'} (\mathbf{I} - \mathbf{A})^{-1} W_t$. They are both linear combinations of the data, but with different weights.

5.1 Econometrics

This section shows how to obtain a permanent-transitory (PT)decomposition of multiple time series generated by finite order Gaussian VAR(p) models with both long-run (cointegration) and short-run (common cycle) restrictions coming from the present value model. Given that we have seen in Section 2 that restrictions within the transformed VAR are equivalent to a WF common feature model for the VECM, we are able to use the multivariate Beveridge-Nelson (BN) decomposition proposed in Hecq et al. (2000). Lettau and Ludvigson (2001) produce a PT for the PVM using a Gonzalo-Granger decomposition. This decomposition have many drawbacks however and it corresponds to the BN decomposition if and only if there exist in an n dimensional model r cointegrating vectors and $n - r$ SCCF vectors (see Proietti, 1997; Hecq et al. 2000). A situation that is impossible under the PVM because an S_{t-1} term enters in the relationship. Given that the results are obtained there we only summarize them here.

This being said we define a strong common PT decomposition of $X_t = (Y_t, y_t)^{'}$ with both common trends and common cycles, the pair of process (μ_t, ψ_t), such that μ_t is a random walk process while ψ_t is a covariance stationary process. $X_t = \mu_t + \psi_t$ with $\beta^{'} \mu_t = 0$ (cointegration) and

$\gamma^{'}\psi_t = 0$ (SCCF). In the BN decomposition $X_t = \mu_t + \psi_t$ where

$$\mu_t = X_t + \left\{ \lim_{l \to \infty} \sum_{i=1}^{l} \Delta \tilde{X}_{t+i|t} - E(\Delta X_t) \right\}$$

is the trend component, that is the value the series would take if it were on its long-run path. $\Delta \tilde{X}_{t+i|t}$ denotes the ith-step ahead forecast. Components μ_t and ψ_t are derived in Proietti (1997) and Hecq et al. (2000). Now in the WF case, only a part of the cycle is annihilated by $\gamma^{'}\psi_t$. Using the companion form of the VECM these components are obtained in Hecq et al. (2000) derived an observable decomposition such that $\psi_t = \psi_t^A + \psi_t^B$ with $\gamma^{'}\psi_t^A = 0$ but $\gamma^{'}\psi_t^B \neq 0$ in

$$X_t = \mu_t + \psi_t^A + \psi_t^B.$$

This can be interpreted either as a decomposition of X_t that includes also WF common cycles or a BN decomposition of $X_t^* = X_t - \psi_t^B$. The specific where

$$\begin{aligned} \psi_t^A &= -\gamma_{\perp}(\gamma_{\perp}^{'}\gamma_{\perp})^{-1}\gamma_{\perp}^{'}(I-P)(\Gamma(1)-\alpha\beta^{'})^{-1}\Gamma^*(L)\Delta X_t + \gamma_{\perp}(\gamma_{\perp}^{'}\gamma_{\perp})^{-1}\gamma_{\perp}^{'}PX_t \\ \psi_t^B &= -\gamma(\gamma^{'}\gamma)^{-1}\gamma^{'}\alpha[\beta^{'}(\Gamma(1)-\alpha\beta^{'})^{-1}\alpha]^{-1}\beta^{'}(\Gamma(1)-\alpha\beta^{'})^{-1}\Gamma^*(L)\Delta X_t + \\ &\quad \gamma(\gamma^{'}\gamma)^{-1}\gamma^{'}\alpha[\beta^{'}(\Gamma(1)-\alpha\beta^{'})^{-1}\alpha]^{-1}\beta^{'} + \alpha\beta^{'}]X_t \end{aligned}$$

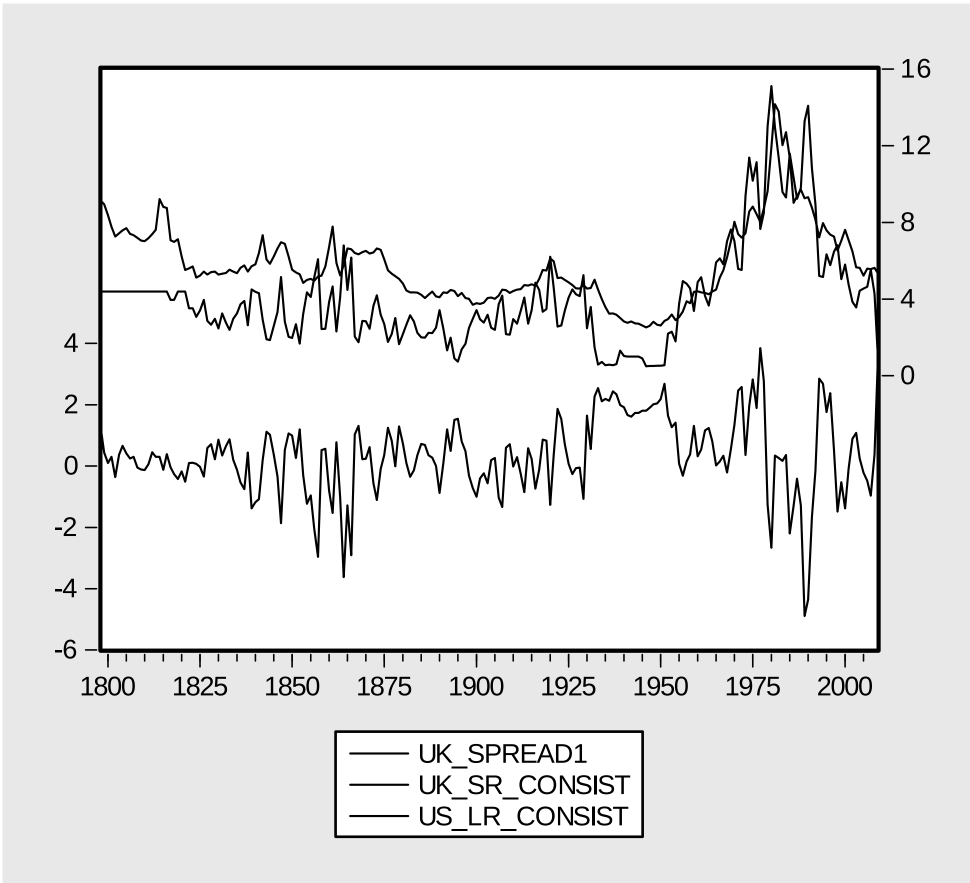
where

6 Application to interest rates

To be finished

We consider annual long and short-term interest rates for the UK. The series ranges from 1831 to 2008. ADF tests do not reject the null that rates in levels are I(1). There is a cointegrating relationship from both Johansen tests and EG with a cointegrating vector that do not differ from $(1 - 1)$.

VAR(2) or VAR(6) dynamic systems are detected by different information criteria. We show results from these two systems although VAR(2) is probably misspecified because it omits some significant dynamics at some lags. Using a robust GMM strategy we don´t reject the null that there is a common cyclical feature vector.



UK: Short Rate, Long Rate and Spread

DGP	$VAR(p)$	LR	J_1	J_2	J_3
WF	2	5.88 (0.016)	21.7 (0.000)	6.34 (0.012)	4.58 (0.032)
	6	28.5 (0.000)	39.6 (0.000)	12 (0.214)	11.5 (0.240)
PSCCF	2	5.88 (0.015)	21.7 (0.000)	6.34 (0.012)	4.58 (0.032)
	6	28.5 (0.000)	39.6 (0.000)	12 (0.214)	11.5 (0.240)

Table 2: Common feature tests statistics)

7 Conclusion

To be finished

References

[1] Alvarez, F. and Jermann, U., 2005, "Using asset prices to measure the persistence of the marginal utility of wealth," *Econometrica*, Vol. 73(6), pp. 1977–2016.

[2] Athanasopoulos, G., Guillen, O.T.C., Issler, J.V., and Vahid, F. (2011), "Model Selection, Estimation and Forecasting in VAR Models with Short-run and Long-run Restrictions," forthcoming, *Journal of Econometrics*.

[3] Beveridge, S. and Nelson, C.R., 1981, "A New Approach to Decomposition of Economic Time Series into a Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151-174.

[4] Campbell, J.Y. (1991). A variance decomposition of stock returns. *Economic Journal*, 405, 175-179.

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

[6] Campbell, John Y. and Deaton, Angus 1989, "Why is Consumption So Smooth?" *The Review of Economic Studies* 56:357-374.

[7] Cubadda, G. and A. Hecq (2001), On Non-Contemporaneous Short-Run Comovements, *Economics Letters*, 73, 389-397.

[8] Engle, R.F. and Kozicki, S. (1993). "Testing for Common Features", *Journal of Business and Economic Statistics*, 11(4): 369-80.

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

[10] Hansen, Lars Peter and José A. Scheinkman (2009), "Long-Term Risk: An Operator Approach," *Econometrica*, 77(1), pp. 177 - 234.

[11] Hecq, A., Palm, F.C., Urbain, J.-P., 2006, "Testing for common cyclical features in VAR models with cointegration," *Journal of Econometrics*, Volume 132(1), pp. 117-141.

[12] Issler, J.V. and Vahid. F. (2001), "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," *Journal of Monetary Economics*, vol. 47(3), pp. 449-475.

[13] Johansen, Soren, and Anders Rygh Swensen, 2011, On a numerical and graphical technique for evaluating some models involving rational expectations. Mimeo., University of Copenhagen and CREATES, University of Aarhus.

[14] Lettau, M., and S. C. Ludvigson (2001): "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815—849.

[15] Lettau, M., and S. C. Ludvigson (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94(1), 276—299.

[16] Lettau, M., and S. C. Ludvigson (2010): "Measuring and Modeling Variation in the Risk-Return Tradeoff," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia, and L. P. Hansen, vol. 1, pp. 617—690. Elsevier Science B.V., North Holland, Amsterdam.

[17] Lettau, Martin, and Sydney C. Ludvigson, 2011, "Shocks and Crashes," NBER Working Paper # 16996.

[18] Vahid, F. and Engle, R. F., 1993, "Common trends and common cycles," *Journal of Applied Econometrics*, vol. 8, 341–360.

[19] Vahid, F., Engle, R. F., 1997, "Codependent cycles," *Journal of Econometrics*, vol. 80, 199–221.

[20] Van Nieuwerburgh, Stijn, Lustig, Hanno N. and Verdelhan, Adrien, 2010, "The Wealth-Consumption Ratio." NYU Working Paper No. FIN-08-045.