Structural Time Series Models with Common Trends and Common Cycles

Christoph Schleicher*†

First version: August 2002
This version: October 2002
Compiled: January 21, 2003

Abstract

This paper models and estimates the Beveridge-Nelson decomposition of multivariate time series in an unobserved components framework. This is an alternative to standard approaches based on VAR and VECM models. The appeal of this method lies in its transparency and structural character. The basic model parsimoniously nests a large set of common trend and common cycle restrictions. It is found that if the cyclical component has a sufficiently rich serial correlation pattern, all covariance terms of the trend and cycle innovations are identified. Tests for common trends are based on a method developed by Nyblom and Harvey (2000), while hypotheses on common cycles are tested using likelihood ratio statistics with standard distributions. This testing framework is used to assess the implications of common trend-common cycle restrictions for the income-consumption relationship in U.S. data. The presence of a common cyclical component yields a rejection of the permanent income hypothesis and evidence is found for the stylized fact that permanent shocks play a more important role for consumption than for income. Out-of-sample forecasts show that common trend and common cycle restrictions improve predictive accuracy.

Keywords: business cycles, common trends, common cycles, unobserved components models, Beveridge-Nelson decomposition, Kalman filter

JEL classification: C15, C22, C32, E32

* University of British Columbia, email: christoph@iam.ubc.ca.

† I would like to thank Francisco Barillas, Paul Beaudry, Lilia Karnizova, Chang-Sik Kim, James Nason and Geneviève Verdier for helpful comments. All remaining mistakes are my own.

1 Introduction

It is well known in business cycle research[1] that trend-cycle decompositions based on unobserved component (UC) models tend to be very different than those based on the Beveridge-Nelson decomposition. While the former typically produce smooth trends and highly persistent cycles of large amplitude, the latter yields uneven trends and cycles that are small and mean-reverting. In a recent paper Morley, Nelson and Zivot (2002, MNZ henceforth) take a closer look at this apparent inconsistency and note that unobserved components models tend to impose the restriction that trend and cycle innovations are orthogonal. This restriction limits the parameter space of the underlying ARIMA representation of the UC model. MNZ show that if the correlation between the innovations of the unobserved components is unrestricted, the UC model gives a trend-cycle decomposition that is identical to that of the Beveridge-Nelson decomposition. Moreover, they find that the zero-correlation restriction is rejected by U.S. GDP data.

This paper expands the analysis of MNZ to a multivariate setting. This aim is motivated on a variety of grounds. First, it is generally believed that a multivariate framework provides a superior modeling environment for macroeconomic variables, because it offers important insights in the dynamic relations between variables as well as the identification of innovation sources. Second, as Durbin and Koopman (2001) argue, the key advantage of unobserved components models and the underlying state-space approach is the structural analysis of the problem that contrasts with ARIMA modeling methods. Individual pieces like the trend, cycle, seasonal and possible exogenous and endogenous explanatory variables can be modeled separately and subsequently combined in the state-space model. Third, the unobserved components approach has the particulary appealing property that restrictions imposed by common factors can be modeled in a transparent way.

For the multivariate Beveridge-Nelson decomposition, possible common factor restrictions include long-run restrictions imposed by common trends (Engle and Granger, 1987, Stock and Watson, 1988) and short-run restrictions imposed by common cycles

[1] See e.g. Canova (1998a).

(Vahid and Engle, 1993). The theoretical framework of both Stock and Watson and Vahid and Engle is based on a structural unobserved components model derived from the Wold representation of a differenced multivariate time series. However, virtually all empirical work[2] has been based on finite order vector autoregressions. While VARs are simple to estimate, this approach has the disadvantage that the Beveridge-Nelson decomposition can only be retrieved indirectly, except in the case where the number of common trends and common cycles adds up to the dimension of the system, as shown by Proietti (1997).

Naturally, several issues arise in this context. First, it is not *a priori* clear whether and under what circumstances the covariance terms of trend and cycle innovations are identified in a multivariate setting. Second, estimation needs to be based on a Kalman filter maximum likelihood method. While this approach is more involved than the standard OLS and IV methods used for the estimation of VARs and structural VARs, it provides a more flexible framework that can be adjusted for the inclusion of additional components such as seasonals and exogenous variables. Recent innovations (see Durbin and Koopman, 2001) have reduced the computational burden and improved the capacity of the Kalman filter to deal with non-stationary problems. As in the VECM framework a goal is to use the estimated model to generate out-of-sample forecasts and forecast error variance decompositions. Third, the standard routines to test for common trends and common cycles (Johansen, 1988 and Vahid and Engle, 1993) are not applicable, therefore we need an alternative testing framework. A contribution of this paper is to shed light on these issues and offer answers.

The general multivariate unobserved components model parsimoniously nests more specific models with a restricted number of common trends and common cycles. The issue of selecting the best model among possible alternatives is interesting for several reasons. First, as Stock and Watson (1988) and later Vahid and Engle (1993) note, the existence of common trends and common cycles may be predicted by theoretical models, such that testing the implied restrictions is equivalent to a test of the theory itself. Second, if a simpler model with fewer parameters is the correct data-generating process, its use improves forecast accuracy. Third, misspecification

[2]Examples include King et al. (1991), Engle and Issler (1995) and Issler and Vahid (2001).

of common stochastic trends components leads to biased estimates or loss of efficiency.

Restrictions imposed by common factors can be interpreted as a reduction of the rank of the covariance of trend and cycle innovations. If the number of trends is held constant under both alternatives, likelihood ratio tests have standard limiting distributions. This is the case for tests for common cycles. The limiting distribution for likelihood ratio tests for common trends is nonstandard. Thus, this paper employs an alternative test for common trends recently developed by Nyblom and Harvey (2000).

The rest of the paper is organized as follows. Section 2 describes the Beveridge-Nelson decomposition and extensions to common trends and common cycles. The standard VECM approach is compared with the state-space framework. It is also shown that a simple correspondence between the VECM and UC models exists in the special case when the number of trends and cycles is equal to the dimension of the system. Section 3 outlines a testing framework based on nested models. Section 4 includes an empirical application for U.S. income and consumption data, a comparison of out-of-sample forecasts, and forecast error variance decompositions. Section 5 concludes.

2 A structural approach to cointegration and common cycles

This section derives the multivariate Beveridge-Nelson decomposition with common trend and common cycle restrictions and gives an overview of the standard VAR/VECM-based estimation approach. Subsequently a state-space model for direct estimation of the unobserved components in the trend-cycle decomposition is introduced.

Let y_t denote an n-vector of I(1) variables such that its first difference has a Wold representation[3]

$$\Delta y_t = C(L)u_t \tag{1}$$

where $C(L)$ is a polynomial matrix with the properties $\sum_{j=1}^{\infty} j|C_j| < \infty$ and $C(0) = I_n$ and u_t is multivariate white noise. By defining $C^*(L) = (1-L)^{-1}(C(L) - C(1))$[4] this process can be rewritten as

$$\Delta y_t = C(1)u_t + \Delta C^*(L)u_t. \tag{2}$$

Integrating both sides then gives the trend-cycle decomposition

$$\begin{aligned} y_t &= C(1)\sum_{s=0}^{\infty} u_{t-s} + C^*(L)u_t \\ &= \tau_t + c_t, \end{aligned} \tag{3}$$

which is Stock and Watson's (1988) multivariate extension of the decomposition proposed by Beveridge and Nelson (1981). Beveridge-Nelson showed that any ARIMA(p,1,q) process can be decomposed into an exactly identified stochastic trend τ_t plus a transitory part with a cyclical interpretation c_t. The trend can alternatively be defined as the limit of the forecast of the time series as the horizon approaches infinity, adjusted for the mean rate of growth μ (which is set equal to zero in our case)

$$\tau_t^{BN} = \lim_{\kappa \to \infty} E[y_{t+\kappa} - \kappa\mu | \Omega_t]. \tag{4}$$

Consequently the cyclical component has no long-run effect. More generally, Stock and Watson build on earlier work by Engle and Granger (1987), to allow for common trends in (3). We will use the following definition:

(*Common Trends*) An n-vector of $I(1)$ variables y_t is said to have $k = n - r$ common trends if there exists an $n \times r$ matrix α of rank r such that

$$\alpha' y_t \sim I(0). \tag{5}$$

Consider further an $n \times k$ matrix γ that lies in the left null-space of α such that $\alpha'\gamma = \mathbf{0}$. The Stock-Watson trend-cycle decomposition is then given by

$$y_t = \gamma\tau_t + c_t \tag{6}$$

[3] In the following we will assume without loss of generality that the mean of Δy_t equals zero, which implies the absence of a linear trend in levels.

[4] See, for example Engle and Granger (1987) for this result.

where τ_t is a k-vector of random walks (the common trends). Note that this implies that $C(1)$ can be factored as $\gamma\delta'$ for some $k \times n$ matrix δ in equation (3). A similar, albeit more restricted, version of this model is given in a *seemingly unrelated time series equations* (SUTSE) context by Harvey (1989).

The Stock and Watson common trends model was further refined with common cycle restrictions proposed Vahid and Engle (1993)[5]. Analogously to the definition of common trends, common cycles (in the sense of Vahid and Engle) are defined as follows:

(*Common Cycles*) An n-vector of $I(0)$ variables Δy_t is said do have $l = n - s$ common cycles if there exists an $n \times s$ $(s + r \leq n)$ matrix $\tilde{\alpha}$ of rank s such that

$$\tilde{\alpha}'\Delta y_t \sim \text{WN}. \tag{7}$$

By defining a $n \times l$ matrix $\tilde{\gamma}$ that lies in the null-space of the cofeature vectors $\tilde{\alpha}$ $(\tilde{\alpha}'\tilde{\gamma} = \mathbf{0})$ and an $n \times l$ polynomial $\tilde{\delta}(L)$ such that $\tilde{\gamma}\tilde{\delta}'(L) = C^*(L)$, we can extend the Stock-Watson-Beveridge-Nelson decomposition to include common cycles, which is then given by the structural model

$$\begin{aligned} y_t &= \gamma\tau_t + \tilde{\gamma}c_t \\ \tau_t &= \tau_{t-1} + \delta' u_t \\ c_t &= \tilde{\delta}'(L)u_t. \end{aligned} \tag{8}$$

The k common trends τ_t follow a multivariate random walk, the l common cycles c_t are usually modelled as an ARMA(p,q) process. Also note that the contemporaneous trend and cycle innovations are perfectly correlated[6], which is an often cited feature of the BN decomposition.

[5]Building on the common features notion of Engle and Kozicki (1993).

[6]In the original paper by Beveridge and Nelson (1981) the trend-cycle decomposition was defined as $y_t = \tau_t - c_t$, such that trend and cycle innovations are perfectly negatively correlated.

2.1 The standard VECM approach

In the related literature[7] the usual approach to estimate models with common trends and common cycles is based on finite order vector autoregressive (VAR) models as an approximation to the more general class of models given by (1). Consider therefore the p-th order VAR model in levels

$$\Pi(L)y_t = u_t, \tag{9}$$

where $\Pi(L) = I - \Pi_1 L - \Pi_2 L^2 - \ldots - \Pi_p L^p$. Since $y_t \in I(1)$ some of the roots of $|\Pi(z)| = 0$ fall on or outside the unit circle. The VAR in levels can be reparametrized to yield the interim multiplier representation (see Banerjee et al. (1993))

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{t-p+1} \Delta y_{t-p+1} + \Pi y_{t-1} + u_t, \tag{10}$$

where $\Pi(L) = \Pi(1) + \Delta\Gamma(L)$, $\Gamma(L) = I - \Gamma_1 L - \ldots - \Gamma_{p-1} L^{p-1}$ and $\Gamma_j = -\sum_{i=j+1}^{p} \Gamma_i$.

Engle and Granger (1987) show that if there are cointegrating relationships, the rank of Π equals $r < n$, such that Π can be factored as the product of two $n \times r$ matrices ($\Pi = -\beta\alpha'$). Here α includes the r cointegrating vectors which span the cointegration space, while β is called the matrix of adjustment coefficients that are the factor loadings in the vector error correction model (VECM)

$$\Gamma(L)\Delta y_t = -\beta z_t, \tag{11}$$

where $z_t = \alpha' y_t$ is the error correction term. It is clear that the assumption of common trends imposes cross-equation restrictions on the VAR in levels, since the sum $\sum \Pi_i$ must then have a reduced rank. It is also evident that the VAR in levels parsimoniously encompasses the VECM, since the former has $n^2 p$ parameters, while the latter has only $n^2(p-1) + 2nr - r^2$ parameters in the conditional mean after accounting for free parameters in α.

Vahid and Engle (1993) show that the existence of common cycles presents additional cross-equation restrictions on the VAR model. Premultiplication of the system in

[7] King et al. (1991), Vahid and Engle (1994), Proietti (1997), Issler and Vahid (2001), amongst others.

differences by the cofeature matrix $\tilde{\alpha}$ produces multivariate white noise

$$\tilde{\alpha}'\Delta y_t = \tilde{\alpha}' u_t. \tag{12}$$

The linear combinations of the cofeature vectors therefore remove serial correlation from the first difference of the data. Equation (12) eliminates $s(np + r) - s(n - s)$ additional parameters, so that the model with common cycles is parsimoniously nested in the VECM.

For estimation usually Johansen's (1988) full information maximum likelihood method is used to determine the number of cointegrating relationships and estimate the VECM. Conditional on the cointegrating vectors the structural reduced form implied by the existence of common cycles can be estimated by standard simultaneous equations techniques (see Vahid and Engle (1993) and Issler and Vahid (2001) for details).

A main attraction of the approach based on finite order vector autoregression lies in the fact that it is relatively straightforward to estimate even if the dimension of the system n and the number of lags p necessary to describe dynamic interaction is large. On the downside, it does not directly give the trend-cycle decomposition described in the previous section (equation (8)). An important exception, as Vahid and Engle note, is the case where the sum of common trends and common cycles is equal to the dimension of the system ($k + l = n$). In this case the cointegrating vectors and the cofeature vectors together span $\mathbb{R}^n$. Using the notation of the previous section we have

$$\begin{aligned} \tilde{\alpha}' y_t &= \tilde{\alpha}\gamma\tau_t \\ \alpha' y_t &= \alpha\tilde{\gamma} c_t \end{aligned}$$

such that the trend-cycle decomposition is given by

$$y_t = \begin{bmatrix} \tilde{\alpha}' \\ \alpha' \end{bmatrix}^{-1} \left(\tilde{\alpha}\gamma\tau_t + \alpha\tilde{\gamma} c_t\right). \tag{13}$$

For the more general case in which k and l do not necessarily add up to n, Proietti (1997) derived explicit formulae for the multivariate Beveridge-Nelson decomposition based on the state-space representation of (10). Proietti's framework also sheds light on the connection between the Beveridge-Nelson decomposition and an alternative factor based decomposition proposed by Gonzalo and Granger (1995). In the particular case where $k + l = n$, the two decompositions are identical.

2.2 The unobserved components model

Instead of following the VAR-based approach discussed in the previous section, the aim of this paper is to approach the model specified in equation (8) directly. This section presents a state-space model that serves as a basis to estimate the Beveridge-Nelson decomposition as the sum of two unobserved components, which consist of k common stochastic trends, $\gamma\tau_t$, and l common cycles, $\tilde{\gamma}c_t$. This approach is in line with the unobserved components models proposed by Harvey (1985), Watson (1986) and Clark (1987), except that no restrictions are imposed on the covariances of the error terms. It is also assumed that the cyclical component is described by a stationary VAR(p) process. This structure yields the model

$$\begin{aligned} y_t &= \gamma\tau_t + \tilde{\gamma}c_t, && (14)\\ \tau_t &= \tau_{t-1} + \eta_t, && (15)\\ \mathbf{\Phi}(L)c_t &= \varepsilon_t, && (16) \end{aligned}$$

where η_t and ε_t are the trend and cycle innovations, and $\mathbf{\Phi}(L)$ is a l-dimensional lag polynomial of order p. The model ((14)-(16)) can be cast into state space form by defining the measurement equation as (14) and the state vector as (15) with present and past values of the cycle being generated by (16). For a model with a VAR(2) cycle a due state space representation is

$$\begin{aligned} y_t &= Z\alpha_t, && (17)\\ \alpha_t &= T\alpha_{t-1} + R\nu_t, \quad \nu_t \sim NID(0, Q), && (18)\\ \alpha_1 &\sim NID(\alpha_{1|0}, P_{1|0}), && (19) \end{aligned}$$

where

$$Z = \begin{bmatrix} \gamma & \tilde{\gamma} & \mathbf{0} \end{bmatrix}, \quad \alpha_t = \begin{bmatrix} \tau_t \\ c_t \\ c_{t-1} \end{bmatrix}, \quad T = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times l} & \mathbf{0}_{k\times l} \\ \mathbf{0}_{l\times k} & \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \\ \mathbf{0}_{l\times k} & \mathbf{I}_l & \mathbf{0}_l \end{bmatrix},$$

$$R = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times l} \\ \mathbf{0}_{l\times k} & \mathbf{I}_l \\ \mathbf{0}_{l\times k} & \mathbf{0}_l \end{bmatrix}, \quad \nu_t = \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}, \quad Q = E[\nu_t'\nu_t].$$

The parameters can be estimated by maximum likelihood using the prediction error decomposition (Harvey 1981). Based on available information at time t and suitable initial conditions, the Kalman filter generates mean square error predictions of the unobserved components, as well as their contemporaneous estimates[8]. A special feature of the state-space system under consideration is that the transition matrix T has k unit roots (corresponding to the k stochastic trends). Therefore care has to be taken with regards to the initialization of the state vector. A consistent way to deal with this problem is the exact initialization method developed by Koopman (1997) and refined in Durbin and Koopman (2001)[9].

The state-space model has a VARIMA reduced form (the canonical form), which can be found by expressing the unobserved components as a function of their innovations and then substituting them into the observation equation:

$$\begin{aligned} \tau_t &= \Delta^{-1}\eta_t \\ c_t &= \Phi(L)^{-1}\varepsilon_t \\ \Delta y_t &= \gamma\eta_t + \tilde{\gamma}\Phi(L)^{-1}\Delta\varepsilon_t. \end{aligned}$$

By defining $\bar{\gamma} \equiv \tilde{\gamma}(\tilde{\gamma}'\tilde{\gamma})^{-1}$, we get the structural model

$$\bar{\gamma}\Phi(L)\bar{\gamma}'\Delta y_t = \bar{\gamma}\Phi(L)\bar{\gamma}'\gamma\eta_t + \bar{\gamma}\Delta\varepsilon_t, \tag{20}$$

which, by Granger and Newbold's theorem[10] and because the right hand side has zero autocorrelations for lags greater than p, has the VARIMA(p,1,p) reduced form

$$\tilde{\Phi}(L)\Delta y_t = \Theta(L)v_t. \tag{21}$$

[8] Details about the estimation by the Kalman filter are given in appendix D.

[9] See appendix D for a brief description.

[10] Granger and Newbold 1986, p. 29.

Here $\tilde{\Phi}(L) \equiv \bar{\gamma}\Phi(L)\bar{\gamma}'$ and $\Theta(L)$ are lag polynomials of rank l, the dimension of the common cycle, and order p. The error process v_t is a linear combination of $\gamma\eta_t$ and $\bar{\gamma}\varepsilon_t$.

As Morley et al. show for the univariate case, in the absence of restrictions on the covariance matrix of η_t and ε_t, the Beveridge-Nelson decomposition of the VARIMA reduced form is identical to the contemporaneous estimates of the unobserved components, $\tau_{t|t}$ and $c_{t|t}$, generated by the Kalman filter. This observation is the converse of Proietti's derivation of the Beveridge-Nelson decomposition for VECM models.

It is important in this context to distinguish between the innovations of the unobserved components η_t and ε_t and the observed innovations v_t. The correlation between the former is not restricted in any way. However, since the observed components constitute a Beveridge-Nelson decomposition, their innovations are perfectly correlated. Returning to the structural model at the beginning of this section, we can state that if the Wold representation of the reduced VARIMA model (21) is identical to the Wold representation in (1), the contemporaneous ML estimates of the Kalman filter of the structural unobserved components model (14)-(16) will be the FIML estimates of the Beveridge-Nelson decomposition in (8).

2.3 A special case: $k + l = n$

In the case when the number of trends and cycles add up to the dimension of the system, the cointegration and cofeature vectors exactly span the space $\mathbb{R}^n$, which simplifies the analysis and gives rise to a direct relationship between the VECM and the UC representation. Consider the following simple example from Engle and Granger (1987)[11]:

$$\begin{aligned} x_t + by_t &= u_t \\ x_t + ay_t &= e_t \\ u_t &= u_{t-1} + \varepsilon_{1t} \end{aligned} \tag{22}$$

[11]Banerjee et al. (1993) also use this example as a demonstration for equivalent model representations.

$$
\begin{aligned}
e_t &= \rho e_{t-1} + \varepsilon_{2t} \\
|\rho| &< 1 \\
[\varepsilon_{1t}, \varepsilon_2]' &\sim \mathrm{NID}(\mathbf{0}, \mathbf{\Sigma}).
\end{aligned}
$$

It is easy to see that this model contains one common stochastic trend u_t (with co-integrating vector $\alpha = [1, a]'$) and one common cycle e_t that follows an AR(1) process (with co-feature vector $\tilde{\alpha} = [1, b]'$). It is straightforward to modify the model into an unobserved component framework (assuming that $a \neq b$)

$$
\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 1 & b \\ 1 & a \end{bmatrix}^{-1} \begin{bmatrix} u_t \\ e_t \end{bmatrix} \\
&= \frac{1}{a-b} \begin{bmatrix} a \\ -1 \end{bmatrix} u_t + \frac{1}{a-b} \begin{bmatrix} -b \\ 1 \end{bmatrix} e_t \\
&= \gamma \tau_t + \tilde{\gamma} c_t,
\end{aligned}
$$

where

$$
\gamma = \frac{1}{a-b} \begin{bmatrix} a \\ -1 \end{bmatrix} \quad \text{and} \quad \tilde{\gamma} = \frac{1}{a-b} \begin{bmatrix} -b \\ 1 \end{bmatrix}
$$

are the loading matrices for the stochastic trend τ_t $(= u_t)$ and the cyclical component c_t $(= e_t)$. It is easy to check that γ and $\tilde{\gamma}$ lie in the null space of α and $\tilde{\alpha}$, respectively. Also $\alpha'\tilde{\gamma} = \tilde{\alpha}'\gamma = 1$, such that

$$
\begin{aligned}
(\tilde{\alpha}'\gamma)^{-1}\tilde{\alpha}' y_t = \tilde{\alpha}' y_t &= \tau_t \\
(\alpha'\tilde{\gamma})^{-1}\alpha' y_t = \alpha' y_t &= c_t.
\end{aligned}
$$

This, and the fact that the number of trends and cycles add up to the dimension of the system, are be used to derive the error-correction form of the model. Define

$$
\mathbf{A} \equiv \begin{bmatrix} \tilde{\alpha}' \\ \alpha' \end{bmatrix}
$$

Apply this definition to the previous equations to find

$$
\mathbf{A} y_t = \begin{bmatrix} \tau_t \\ c_t \end{bmatrix} \quad \text{and} \quad \mathbf{A}\Delta y_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} 0 \\ (1-\rho) \end{bmatrix} \alpha' y_{t-1}
$$

such that the interim multiplier representation is given by

$$\Delta y_t = \Pi y_{t-1} + v_t.$$

The impact matrix Π has rank 1 and can be factored as $\Pi = -\beta\alpha'$, where

$$\begin{aligned} \beta &= \mathbf{A}^{-1} \begin{bmatrix} 0 \\ (1-\rho) \end{bmatrix} \\ v_t &= \mathbf{A}^{-1} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \end{aligned}$$

Define the error-correction term to be $z_t = \alpha' y_t$, to produce the vector-error correction representation

$$\Delta y_t = -\beta z_{t-1}.$$

The following proposition generalizes this result for the more general case of a VAR(p) cycle.

Proposition 1 (VECM Representation) *If the sum of common cycles and common trends equals the dimension of the system ($k + l = n$), the reduced form of a multivariate unobserved component model with p autoregressive lags in the cycle has a vector error correction representation with p lags.*

The interim multiplier representation is given by

$$\Delta y_t = \mathbf{\Gamma}_1 \Delta y_{t-1} + ... + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + \mathbf{\Pi} y_{t-1} + \nu_t, \tag{23}$$

where

$$\begin{aligned} \mathbf{\Gamma}_j &= -\mathbf{A}^{-1} \begin{bmatrix} \mathbf{0}_{k\times l} \\ \sum_{i=j+1}^{p} \mathbf{\Phi}_i \end{bmatrix} \alpha' \\ \mathbf{\Pi} &= -\mathbf{A}^{-1} \begin{bmatrix} \mathbf{0}_{k\times l} \\ \mathbf{I} - \sum_{i=1}^{p} \mathbf{\Phi}_i \end{bmatrix} \alpha' = -\mathbf{A}^{-1} \begin{bmatrix} \mathbf{0}_{k\times l} \\ \mathbf{\Phi}(1) \end{bmatrix} \alpha' \\ \nu_t &= \mathbf{A}^{-1} \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \end{aligned}$$

and $\mathbf{A}$ is defined as before (see appendix A for a derivation). The VECM representation is given by factoring the long-run multiplier matrix $\mathbf{\Pi} = -\beta\alpha'$, which is restricted in the sense that it has $(p-1)nr + 2nr - r^2$ variables in the conditional mean, compared to the VAR in levels, which has pn^2 variables[12].

The converse of proposition 1 is not true, because not every VECM(p) model has a UC representation with a VAR(p) cycle. However, it follows from Proietti's (1997) derivation of the BN decomposition that every VECM model has a UC representation with a VARMA(p,q) cycle. On the other hand, as was shown in the previous section, every UC model has a VARIMA reduced form.

3 Testing for common trends and common cycles

The state space framework allows for a direct and transparent comparison between nested models, which provides a convenient background for testing for common trends and common cycles. Most existing tests in the literature, in particular the common trend tests by Stock and Watson (1988) and Johansen (1988) and the common cycle test by Vahid and Engle (1993), are based on a VECM framework. These approaches cannot be applied in our case. Since the unobserved components model is estimated by maximum likelihood, it is intuitive to ground a test on a comparison of the likelihoods of a restricted and an unrestricted model. The likelihood ratio (LR) test is a leading example. For the case where the number of stochastic trends is equal under both the null and the alternative hypothesis (such as a test for common cycles), the LR test statistic has a standard χ^2 distribution based on the number of restrictions imposed. If, on the other hand, the number of stochastic trends differs under the alternative hypotheses, the limiting distribution of the LR test becomes non-standard and is unknown. A viable alternative in this case is a test recently developed by Nyblom and Harvey (2000).

This section presents a general to specific procedure of model testing and a subsequent discussion of LR tests for common cycles and the Nyblom/Harvey test for common

[12] Issler and Vahid 2001 show this in their appendix A.

trends. Define model $M(k,l)$, to have at most k and l common trends and cycles, respectively, such that the set of all possible models $\mathcal{M} = \{M(k,l) : 0 < k \leq n, 0 < l \leq n, n \leq k + l\}$[13] has cardinality $n - 1 + \sum_{j=1}^{n} j$. We can also see that two models are nested

$$M(k_1, l_1) \subset M(k_2, l_2)$$

if (and only if) $k_1 \leq k_2$ and $l_1 \leq l_2$.

To show that the matrices γ and $\tilde{\gamma}$ do not alter the inherent structure of the model (in the sense that submodels can be nested) we can define $\tilde{\tau} \equiv \gamma\tau$ and $\tilde{c} = \tilde{\gamma}c$ and rewrite the unobserved components model as

$$\begin{aligned} y_t &= \tilde{\tau}_t + \tilde{c}_t \\ \Delta\tilde{\tau}_t &= \gamma\eta_t \\ \bar{\gamma}\Phi(L)\bar{\gamma}' &= \tilde{\gamma}\varepsilon_t, \end{aligned} \tag{24}$$

which gives an alternative interpretation in terms of reduced rank components.

The proposed strategy for model selection and testing for common cycles can be described as follows: (i) Start with the most detailed model (n,n) and estimate its likelihood. (ii) Estimate the models $(n-1,n)$ and $(n,n-1)$ and compute the relevant test statistics. As is argued below, tests for common cycles (comparing models (k,l) and $(k,l-1)$) have a limiting χ^2 distribution, while tests for common trends (comparing models (k,l) and $(k-1,l)$) can be based on the method developed by Nyblom and Harvey. (iii) Repeat these steps until the less general model is rejected by the tests. For the cases $n = 2$ and $n = 3$ the model selection tree is shown in figure 7. For $n = 2$ the most parsimonious model has one common trend and one common cycle. For $n = 3$ the most parsimonious models have once cycle and two trends and two trends and one cycle, respectively. If there were only one trend and one cycle the resulting model would be stochastically singular and the data could be represented by a two-dimensional system. Note that the selection scheme may not provide a unique result, since tests are only possible on a vertical level among nested models, but not on a horizontal level among non-nested models. In the case where

[13] The last inequality stems from Issler and Vahid's (1993) observation that the space spanned by the co-integrating and common cycle vectors α and $\tilde{\alpha}$ must be at least of dimension n.

several potential models remain, the ultimate choice has to be left to the discretion of the researcher.

Tests for common cycles:

Since there are no misspecified nonstationary components under either H_0: $M(k,l)$ and H_1: $M(k,l+1)$ the asymptotic distribution of the LR test has a $\chi^2(f)$ distribution, where f is the number of restrictions imposed by H_0. In order to compute f, we need first verify that the model is not under-identified. Otherwise there might be unrestricted parameters under the alternatives.

It is common practice to set the covariance between trend and cycle innovations equal to zero to avoid under-identification of unobserved components models. An example is the univariate local level model, which is a special case of the unobserved components model considered in this paper in which $n = 1$ and $p = 0$. This model is under-identified unless $E[\eta_t \varepsilon_t] = \sigma_{\eta\varepsilon}$ is fixed. On the other hand, Morley et al. show that in the case when $n = 1$, $p = 2$, the model is exactly identified. The following proposition indicates that in most cases of interest a sufficient condition for identification is that there are at least 2 autoregressive lags in cycle. The intuition behind this result is that the serial correlation induced by the cyclical component increases the complexity of the autocovariance function of the VARIMA reduced form which is directly observable.

Proposition 2 *If $k = l = n$, the parameters in the UC model defined in equations (14) to (16) with with a VAR(p) cycle are identified if (and only if) $p \geq 1 + \frac{1}{n}$. The model is exactly identified if $p = 1 + \frac{1}{n}$. The only integer solution for exact identification is $p = 2$ and $n = 1$.*

If $k < n$ and/or $l < n$ the condition $p \geq 1 + \frac{1}{n}$ is sufficient, but not necessary for identification.

See appendix B for a proof. Given that the model is identified under both alternatives, the number of restrictions can be determined by comparing the VAR polynomial

$$\tilde{\Phi}(L) = \bar{\gamma}\Phi(L)\bar{\gamma}',$$

and the covariance matrix of the trend and cycle innovations

$$\tilde{\Sigma} = E[\tilde{v}\tilde{v}'] = \begin{bmatrix} \gamma\Sigma_\eta\gamma' & \tilde{\gamma}\Sigma'_{\eta\varepsilon}\gamma' \\ \gamma\Sigma_{\eta\varepsilon}\tilde{\gamma}' & \tilde{\gamma}\Sigma_\varepsilon\tilde{\gamma}' \end{bmatrix}$$

under the alternatives. A Monte-Carlo experiment (described in appendix F) indicates that empirical critical values of LR tests on simulated data are close to their theoretical counterparts.

Tests for common trends:

Because the limiting distribution of the LR test for alternative hypothesis about the numbers of stochastic trends is unknown, Nyblom and Harvey (2000) consider a locally best invariant (LBI) test. Their framework operates on the multivariate local level model, which is a basic building block of many structural time series models:

$$\begin{aligned} \tau_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \xi_t. \end{aligned} \tag{25}$$

Here $\epsilon_t \sim NID(0, \Sigma_\epsilon)$, $\xi_t \sim NID(0, \Sigma_\xi)$, and $E[\epsilon\xi'] = 0$. In our context the multivariate local level model describes the estimated trend component τ_t after accounting for the cyclical component c_t ($\hat{\tau}_t = y_t - \hat{c}_t$). We are interested in the hypotheses $M(k,l)$ and $M(k+1,l)$, which in Nyblom and Harvey's test correspond to $H_0 : \text{rank}(\Sigma_\xi) = k$ against $H_1 : \text{rank}(\Sigma_\xi) > k$. The test statistic is given by

$$\zeta_{k,n} = \lambda_{k+1} + ... + \lambda_n, \tag{26}$$

which is the sum of the $(n-k)$ smallest eigenvalues of $S^{-1}C$, where C is an estimator of the second moments of partial sums of the time series

$$C = T^{-2}\sum_{j=1}^{T}\left[\sum_{t=1}^{j}(\tau_t - \bar{\tau})\right]\left[\sum_{t=1}^{j}(\tau_t - \bar{\tau})\right]'$$

and S is an estimator of the spectral density at zero frequency

$$S = T^{-1}\sum_{t=1}^{T}(\tau_t - \bar{\tau})(\tau_t - \bar{\tau}).$$

The limiting distribution of $\zeta_{k,n}$ depends on functionals of Brownian motion and is given in appendix E, critical values are tabulated by Nyblom and Harvey. The test can be adjusted for serial dependence in ϵ_t by substituting S with a non-parametric estimator of the spectral density at the zero frequency.

An interesting observation is that this test moves in the opposite direction as Johanson's (1988) test based on canonical correlations. The Nyblom and Harvey test starts with the null hypothesis of no stochastic trends, while at the outset of Johansens's testing framework is the assumption of n stochastic trends. As a consequence a researcher rejecting alternatives at a low tail probability level will more likely adopt a model with fewer common trends using the Harvey/Nyblom test than with Johansen's test.

4 The permanent income example

A reoccurring theme in macroeconomic research is Hall's (1978) assertion[14] that under certain assumptions rational representative agents with time-separable utility function will maximize their life-time utility by consuming their permanent income in each period. An important implication of the permanent income hypothesis (PIH) is that it would allow identification of the unobserved permanent component of income by setting it equal to observed consumption. While Beveridge and Nelson (1981) demonstrate that a decomposition of income into a stochastic trend and a stationary cyclical component is always possible, from a statistical perspective there is no guarantee for uniqueness and competing decompositions may be unidentifiable (Watson, 1986).

Contrary to the earlier belief that consumption would violate the PIH by being too volatile, Deaton (1987) shows that consumption is in fact excessively smooth. Noting that the first difference of GDP is positively autocorrelated, Deaton deduces that the variance of income must be smaller than that of permanent shocks to income, and therefore the variance of consumption. However, for U.S. data the variance of

[14]Based on Friedman's (1957) earlier work.

innovations in income is larger or approximately equal to the variance of the first difference of consumption.

With the aim to reconcile theory with empirical evidence, Campbell and Mankiw (1989 and 1990) and Flavin (1981 and 1993) develop alternative consumption models. As Vahid and Engle show, these models imply the unobserved components form

$$\begin{bmatrix} y_t \\ c_t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} y_t^P + \begin{bmatrix} 1 \\ \lambda \end{bmatrix} y_t^T, \tag{27}$$

where y_t^P is permanent income following a random walk and y_t^T is a stationary cyclical component (transitory income). Note that this model is a special case of a two dimensional unobserved components model with a common and a common cycle where $\gamma = [1,1]'$ and $\tilde{\gamma} = [1,\lambda]'$. The model nests the PIH as the case $\lambda = 0$. In Campbell and Mankiw's model the economy is populated by two types of agents. Rational individuals consume their permanent income in each period, while their myopic counterparts consume their present income. λ is defined as the ratio of income that belongs to the myopic rule-of-thumb consumers. In Flavin's model λ has a conceptually similar interpretation as the marginal propensity to consume out of transitory income.

Compared to earlier investigations of the PIH based on univariate models[15], a bivariate system provides a superior modeling environment to test for stationary interaction (or the lack thereof) between income and consumption. Moreover, Flavin (1993) notes that the relative size of income and consumption innovations, which are the variables of interest in Deaton's paradox, may be sensitive to whether we allow for contemporaneous correlation or not.

The data are quarterly series of U.S. per capita GNP and private consumption in the period 1949:1-1988:4 taken from the dataset of King et al. (1991) that was also used later by Proietti (1997) and Issler and Vahid (2001). A plot of the data (figure 1) indicates that the two series share similar long-run and short-run movements. It is also evident that consumption has a smoother appearance than income. Because

[15] Examples include Hall (1978), Watson (1986) and Deaton (1987).

the unobserved components model in this paper abstracts from a drift term in the stochastic trend, a linear trend was subtracted from the data prior to estimation.

Based on the testing framework discussed in the previous section we can compare the following four statistical models

- Model 1, $M(2,2)$: 2 trends and 2 cycles
- Model 2, $M(1,2)$: 1 trend and 2 cycles
- Model 3, $M(2,1)$: 2 trends and 1 cycle
- Model 4, $M(1,1)$: 1 trend and 1 cycle.

The corresponding model selection chart is shown in the top row of figure (7). Following a tradition in the unobserved components literature the cyclical component is modeled as an AR(2) process, since this is the most parsimonious representation that allows for an interior maximum in the spectral density[16].

The parameter estimates of the four competing model are summarized in table 1 and the corresponding trend-cycle decompositions are shown in figures 2 to 5. Standard errors are computed from the inverse of the Hessian at the maximized log-likelihood. Table 1 reveals that all estimates of the autoregressive lag polynomial $\phi_{i,j}$, as well as the loading matrices γ and $\tilde{\gamma}$ are significantly different from zero. Most standard errors of the covariance terms are of the same magnitude as the parameter estimates themselves. We can also rank the the log-likelihoods of the models as -438.26 $>$ -441.10 $>$ -443.07 $>$ -445.19, going from model 1 to model 4. Restrictions imposed by common cycles therefore have a stronger impact on the likelihood function than restrictions imposed by a common trend. This observation comes as a surprise because the trend-cycle decompositions with common trend restrictions are qualitatively more different from the most general model with two trends and cycles in the sense that the cyclical component has a much larger amplitude.

This is clearly visible in figures 2 to 5, which are drawn on the same scale. Another salient feature of the graphs is that the estimated cycles drop during each NBER

[16]The choice of two autoregressive parameters also facilitates the constraint that the roots of the AR polynomial stay outside the unit circle during ML estimation.

recession. For the two models with 2 trends (model 1 and model 3), the spectra of the cycles (shown in figure 6) peak at a periodicity of about one year. If, on the other hand, the trends are restricted to cointegrate, the peak of the cyclical spectra moves to the zero frequency for income in model 2 (2 cycles) and for both income and consumption for model 4 (1 cycle). For the two models with a common cycle the quadrature-spectra are equal to zero, since (by definition) there is no correlation between phase-shifted components at any frequencies.

A more rigorous comparison between the 4 different statistical representations can be drawn using the tests among nested alternatives discussed in section 3. It was maintained that LR tests for the number of common cycles have standard distributions, while a comparison of hypotheses about the number of common trends can be based on the test developed by Nyblom and Harvey (2001). Test statistics and critical values are given in table 2. To remove the effect of serial correlation in the estimated trend component, the estimator of the long-run variance in the Nyblom-Harvey test S is corrected by a Bartlett window with 8 lags. The null hypothesis of model 2 ($M(1,2)$: 1 trend and 2 cycles) versus the alternative of model 1 ($M(2,2)$: 2 trends and 2 cycles) is not rejected at the 10 percent level using the critical values of the Cramér-von Mises distribution tabulated by Nyblom and Harvey. A similar result holds for the case of model 4 ($M(1,1)$: 1 trend and 1 cycle) against model 3 ($M(2,1)$: 2 trends and 1 cycle).

Tests for common cycles yield somewhat weaker evidence. Since model 3 has 5 fewer parameters than model 1, the distribution of the LR test is asymptotically $\chi^2(5)$. The test statistic is 9.608, therefore the null hypothesis of a common cycle can be rejected at the 10 percent level, but not at the 5 percent level. Model 4 ($M(1,1)$: 1 trend and 1 cycle) has 4 fewer parameters than model 2, implying a $\chi^2(4)$ distribution. Again we can reject the null at the 10 percent level, but not at the 5 percent level.

These results confirm the earlier impression that common trend restrictions are better supported by the data than common cycle restrictions. Note also that the estimates of the common trend parameter γ and the common cycle parameter $\tilde{\gamma}$ are both significantly different from zero at the 5 percent level[17].

[17] With some abuse of notation the loading matrix of the common trend components is defined

The statistical results have several implications for the consumption models mentioned earlier. It is fair to assume that income and consumption share a common stochastic trend and we cannot reject the hypothesis that the cointegrating vector is equal to $[1, -1]$, since the estimate of γ is not significantly different from unity. In all specifications, except model 2, the variance of the innovations of the stationary component of consumption ($\sigma^2_{\varepsilon_2}$ in models 1 and 2 and $\tilde{\gamma}^2\sigma^2_{\varepsilon_1}$ in models 3 and 4) is significantly different from zero, which implies a rejection of the PIH. If we accept the hypothesis of a common cycle (which is not rejected at the 5 percent level), we can use the modeling environment (27) of Campbell and Mankiw and Flavin, which statistically encompasses Hall's model. In this case the estimate of λ is 0.5 (parameter $\tilde{\gamma}$ in model 4) and significantly different from zero. Therefore the PIH is rejected in favor of the new-Keynesian consumption models. An interesting observation is that the estimated value of λ is identical to the estimate of Campbell and Mankiw (1990). In the sample period half of the U.S. economy's income therefore accrues to rule-of-thumb consumers. Similar estimates are also obtained by Flavin (1993) and Vahid and Engle (1993).

Table 1 shows that the covariances between trend and cycle innovations are negative in all cases, which coincides with the univariate example of MNZ (2002). MNZ interpret their finding as evidence for the relative importance of real shocks, in the sense that positive shocks to the trend will have a negative effect for the cycle. This view is questioned by Proietti (2002), who argues that the direction of causality may be well reversed, such that positive cyclical shocks have a negative impact on the long-run trend. At the crux of the problem lies the fact that it is impossible to pin down the particular orthogonal decomposition of the trend and cycle errors based on *a priori* grounds. Proietti notes that permanent-transitory decompositions based on orthogonalized errors are close to the spirit of Blanchard and Quah (1989), such that the permanent component follows a VARIMA process rather than a random walk. Proietti further finds that (i) unobserved components models with correlated trend and cycle innovations may be observationally equivalent to alternative UC represen-

as $\gamma = [\gamma_1, \gamma_2]' \equiv [1, \gamma]'$. Since α and γ are only defined up to a nonsingular transformation, we can normalize $\gamma_1 = 1$. The same definition is used for $\tilde{\gamma}$, the loading matrix of the common cycle components.

tations that have orthogonal innovations and (ii) negative correlation between the innovations implies that the future carries more information for the signal extraction than the past. The first point is of little relevance for our results, since it only concerns the case where innovations are either positively correlated or the variance of trend innovations is relatively small compared to the variance of cycle innovations. The second point is important, however, and hints at the possibility that the cyclical components extracted by the Kalman filter are too small in amplitude. A remedy for this problem is to compute smoothed estimates that make use of past as well as future information.

The variances of the income trend and cycle innovations are both larger than the corresponding values for consumption, which matches Deaton's challenging result that consumption fails the PIH because of excessive smoothness. In order to assess the relative importance of permanent and transitory shocks, tables 5 and 6 provide forecast error variance decompositions for income and consumption, respectively (see appendix D for details on the computation of the FEVDs). Because the orthogonalization is based on a Cholesky decomposition, the FEVDs should be interpreted with caution. The variation of FEVDs between the four estimated models and earlier results by King et al. (1991) and Issler and Vahid (2001) is quite large. It is, however, possible to draw comparisons between the relative importance of permanent and transitory shocks of income and consumption. All models indicate that the relative importance of permanent shocks at business cycle horizons is much larger for consumption than for income. One can conclude that although the PIH in its purest form is not supported by the data, permanent income has a strong impact on consumption even over a short horizon.

4.1 Out-of-sample forecasts

A potential advantage of common factor restrictions is that they lead to a more parsimonious representation of the data and, if correctly imposed, will improve the predictive accuracy of the model. Since the Kalman filter is built on a Markovian first order difference equation (the state equation), it provides an ideal forecasting framework (see appendix D for more details). To produce out-of-sample forecasts, the four statistical models were estimated by setting the last 22 observations as missing. A comparison between the forecasts and actual values of income and consumption is given in figure 8. The confidence sets indicate that the standard error of the forecast is about twice as high for consumption as for income. Table 3 provides mean squared forecast errors to compare the predictive accuracy of the individual models. A striking feature is that for both time-series the models with a common trend (model 2 and model 4) outperform the other models. For income the best predictor is model 2 and for consumption the best predictor is model 4. If we use the determinant of the mean squared error matrix as a measure of the overall predictive ability of the system, the most parsimonious representation, model 4, is the clear winner. In order to verify the statistical significance of the variations in predictive accuracy, table 4 gives p-values of White's (2000) reality-check test, based on 50.000 stationary bootstrap[18] resamples. The test is applied to the determinant of the mean squared forecast error matrix, and shows that model 4 consistently outperforms all other models, except model 2. These findings are in line with those of Issler and Vahid (2001), who find that a VECM with common cycle restrictions provides more accurate forecasts than an unrestricted VECM.

5 Conclusions

The theoretical framework on the Beveridge-Nelson decomposition and its extensions to common trends and common cycles are usually based on unobserved components models. However, the tradition is to estimate VECMs. This paper shows that direct estimation of the unobserved components model is a viable alternative which may be preferred from an econometric perspective. The UC model can be cast into a state-space framework and estimated by Kalman filter maximum likelihood. The appeal of this approach is that restrictions imposed by common factors are transparent. A further advantage is that the trend-cycle decomposition is immediately available, also in cases when the sum of common trends and common cycles is greater than the dimension of the system.

[18] Politis and Romano (1994).

The unobserved components model always has a VARIMA reduced form which can be used to verify the identification of the model parameters. In the case where the cyclical component follows a VAR(p) process, the model is identified whenever p is greater than one. The structure of the unobserved components model facilitates testing for common trends and common cycles. It is found that tests for common cycles can be based on the likelihood ratio principle, where the number of restrictions depends on the reduction of the rank of the VAR polynomial of the cyclical component and the covariance matrix of the trend and cycle innovations. Likelihood ratio tests for common trends have an unknown nonstandard limiting distribution. A possible alternative is the test developed by Nyblom and Harvey (2000).

As Vahid and Engle (1993) show, the common trend - common cycle model provides a suitable testing framework for hypotheses on consumption models. For U.S. data the unobserved components model finds strong evidence that income and consumption follow the same stochastic trend. The existence of a common cycle finds support at the 5 percent level. Furthermore, the cofeature vector is significantly different from $[1, 0]$, thereby rejecting the permanent income hypothesis. However, forecast error variance decompositions show that consumption is dominated by permanent shocks, even in the short-run. Out-of-sample forecasts support the assertion that common trend and common cycle restrictions provide a more efficient representation of the data-generating process, thus improving predictive accuracy.

Proietti (2002) notes that correlation between innovations of unobserved trend and cycle components may be explained by a variety of data-generating processes. If the correlation is negative, he further shows that the future carries more information for signal extraction than the past, such that the Kalman filter underestimates the cyclical component. In this case the two-sided Kalman filter (Kalman smoother) is a preferred alternative.

For the sake of parsimony and comparability to the existing unobserved components literature the cyclical component is modelled as a simple AR(2) process, however it is likely that a more complex model will give a better description of the data. Further research should concentrate on more realistic models for the unobserved components, such as the inclusion of seasonal dummies and higher order autoregressive processes.

Another interesting application are rational expectations models. The Markovian nature of the state-space approach often allows to express expectations conditional on past values of driving forces in a compact way, which provides a potential alternative to GMM estimation of Euler equations.

Appendix

A Derivation of Proposition 1

By choosing α, $\tilde{\alpha}$, γ and $\tilde{\gamma}$ so that $\alpha'\tilde{\gamma} = \mathbf{I}_r$ and $\tilde{\alpha}'\gamma = \mathbf{I}_s$ (which is possible, since these matrices are defined only up to a nonsingular transformation), we can write

$$\tilde{\alpha}' y_t = \tau_t \tag{A-1}$$
$$\alpha' y_t = c_t \tag{A-2}$$

and

$$\tilde{\alpha}' \Delta y_t = \eta_t \tag{A-3}$$
$$\alpha' \Delta y_t = \varepsilon_t + \mathbf{\Phi}_1 \alpha' y_t + ... + \mathbf{\Phi}_p \alpha' y_{t-p} - \alpha' y_{t-1}. \tag{A-4}$$

The last expression can be rewritten as

$$\begin{aligned}
& \varepsilon_t + (\mathbf{\Phi}_1 - \mathbf{I})\alpha' y_{t-1} + ... + (\mathbf{\Phi}_{p-1} + \mathbf{\Phi}_p)\alpha' y_{t-p+1} - \mathbf{\Phi}_p \alpha' \Delta y_{t-p+1} \\
= \; & \varepsilon_t + (\mathbf{\Phi}_1 - \mathbf{I})\alpha' y_{t-1} + ... + (\mathbf{\Phi}_{p-2} + \mathbf{\Phi}_{p-1} + \mathbf{\Phi}_p)\alpha' y_{t-p+2} \\
& -(\mathbf{\Phi}_{p-1} + \mathbf{\Phi}_p)\alpha' \Delta y_{t-p+2} - \mathbf{\Phi}_p \alpha' \Delta y_{t-p+1} \\
= \; & \varepsilon + (\sum_{i=1}^{p} \mathbf{\Phi_i} - \mathbf{I})\alpha' y_{t-1} - \sum_{i=2}^{p} \mathbf{\Phi}_i \alpha' \Delta y_{t-1} - ... - \mathbf{\Phi}_p \alpha' \Delta y_{t-p+1}
\end{aligned} \tag{A-5}$$

B Proof of Proposition 2

Identification of the model parameters can be verified by comparing the structural VARIMA representation of the unobserved component model

$$\tilde{\Phi}(L)\Delta y_t = \tilde{\Phi}(L)\gamma\eta_t + \bar{\gamma}\Delta\varepsilon_t, \tag{B-1}$$

and its reduced form

$$\tilde{\Phi}(L)\Delta y_t = \Theta(L)u_t, \tag{B-2}$$

where $\tilde{\Phi}(L) = \bar{\gamma}\Phi(L)\bar{\gamma}'$.

The parameters in the autoregressive lag polynomial and $\bar{\gamma}$ (and consequently $\tilde{\gamma}$ up to a nonsingular transformation) are clearly identified by equating the LHS of (B-1) and (B-2). Similarly, identification of the parameters in the covariance matrix

$$\tilde{\Sigma} = E[\tilde{v}\tilde{v}'], \qquad \tilde{v} \equiv [\gamma\eta', \bar{\gamma}\varepsilon']',$$

and γ depends on equating the information of the RHS of the structural and the reduced form. This information is contained in the structural and reduced form autocovariance functions $\Gamma_s(j)$ and $\Gamma_r(j)$, with $\Gamma_s(j) = \Gamma_r(j) = 0$ for $|j| > p$. Assume first that $k = l = n$ such that $\gamma = \tilde{\gamma} = I_n$. Then we have $n^2(p+1)$ equations $(\Gamma(j), j = 0, ..., p)$ equations in $2n^2 + n$ unknowns (the parameters in $\tilde{\Sigma}$). Hence the model parameters are (over-) identified if (and only if) $n^2(p+1) \geq 2n^2 + n$, or $p \geq 1 + \frac{1}{n}$. Define $M = p + 1$. Then the model is exactly identified if (and only if) $MN = 1$. Clearly there are no positive integer solutions other than $M = n = 1$.

For the cases $k \neq n$ and $l \neq n$ there are still $n^2(p+1)$ equations generated by the autocovariances, since the space spanned by $\tilde{v}$ must have at least dimension n, because $k + l \geq n$. However, the number of parameters to be identified decreases. Hence the model parameters are (over-) identified if $p \geq 1 + \frac{1}{n}$.

C Implicit restrictions on the covariance structure imposed by common trends and common cycles

It is shown in (24) that the UC model with common trends and common cycles can be alternatively expressed in terms of n-dimensional components with reduced-rank innovations. This representation has the advantage that nested models are directly comparable. This appendix demonstrates how common trends and common cyles implied by rank reductions of the cointegration and cofeature matrices α and $\tilde{\alpha}$ (and consequently γ and $\tilde{\gamma}$) implicitly lead to (partly) nonlinear restrictions on the covariance matrix of $v_t = [\eta_t, \varepsilon_t]'$. This is shown by comparing the parameters of model 1 (2 trends and 2 cycles) and model 3 (2 trends and 1 cycle) from the empirical part (section 4).

The parameters of model 1 are: $\phi_{1,1}, \phi_{1,2}, \phi_{2,1}, \phi_{2,2}$ plus the covariance terms

$$\mathrm{E}[(\eta_{\mathrm{t}}', \varepsilon_{\mathrm{t}}')'(\eta_{\mathrm{t}}', \varepsilon_{\mathrm{t}}')] = \begin{bmatrix} \sigma^2_{\eta_1} & & & \\ \sigma_{\eta_1\eta_2} & \sigma^2_{\eta_2} & & \\ \sigma_{\eta_1\varepsilon_1} & \sigma_{\eta_2\varepsilon_1} & \sigma^2_{\varepsilon_1} & \\ \sigma_{\eta_1\varepsilon_2} & \sigma_{\eta_2\varepsilon_2} & \sigma_{\eta_1\eta_2} & \sigma^2_{\varepsilon_2} \end{bmatrix}$$

The restrictions imposed by model 3 are:

$$\phi_{2,1} = \phi_{1,1} \qquad \phi_{2,2} = \phi_{1,2}$$

$$\begin{aligned} \sigma^2_{\varepsilon_2} &= \tilde{\gamma}^2 \sigma^2_{\varepsilon_1} & \sigma_{\eta_1\varepsilon_2} &= \tilde{\gamma}\sigma_{\eta_1\varepsilon_1} \\ \sigma_{\eta_2\varepsilon_2} &= \tilde{\gamma}\sigma_{\eta_2\varepsilon_2} & \sigma_{\varepsilon_1\varepsilon_2} &= \tilde{\gamma}\sigma^2_{\varepsilon_1}. \end{aligned}$$

After using one restriction to eliminate $\tilde{\gamma}$, 5 (partly nonlinear) restrictions remain. It is easy to see that the rank of the covariance matrix of η_t and ε_t is reduced from full rank to rank 3. The other cases are similar, except that model 4 (1 trend, 1 cycle) has two additional parameters (γ and $\tilde{\gamma}$) such that the number of restrictions decreases by two.

D Maximum likelihood estimation and the Kalman filter

Given the state-space framework (17)-(18)

$$\begin{aligned} y_t &= Z\alpha_t & \text{(D-1)} \\ \alpha_t &= T\alpha_{t-1} + R\nu_t, \quad \nu_t \sim NID(0, Q), & \text{(D-2)} \end{aligned}$$

and initial conditions (19),

$$\alpha_1 \sim NID(\alpha_{1|0}, P_{1|0}), \qquad \text{(D-3)}$$

the Kalman filter (Kalman, 1960) computes minimum mean squared error estimates a_t of the state vector α_{t+1} and its mean square error matrix P_{t+1} conditional on available information at time t ($\Omega_t = \{y_1, y_2, ..., y_t\}$) using the recursive algorithm

$$\begin{array}{rclrcl} v_t & = & y_t - Z\alpha_t, & F_t & = & ZP_tZ', \\ K_t & = & TP_tZ'F_t^{-1}, & L_t & = & T - K_tZ, \\ a_{t+1} & = & Ta_t + K_tv_t, & P_{t+1} & = & TP_tL_t' + RQR'. \end{array} \tag{D-4}$$

The contemporaneous filtering equations give the real time or concurrent estimates of the state vector and its mean square error matrix

$$a_{t|t} \;=\; a_t + P_tZ'F_t^{-1}v_t, \qquad P_{t|t} \;=\; P_t - P_tZ'F_t^{-1}ZP_t'. \tag{D-5}$$

In the context of this study, the trend-cycle decomposition is based on the corresponding components of $a_{t|t}$, which is identical to the Beveridge-Nelson decomposition of the underlying reduced form VARIMA model.

If the system is stationary in the sense that all roots of the transition matrix T are inside the unit disk, the log-likelihood of the estimated model is given by

$$\log L(y) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}(\log|F_t| + v_t'F_t^{-1}v_t). \tag{D-6}$$

The exact initial Kalman filter

For the UC model defined in section 2.2 matters are complicated by the fact that the transition matrix of the state equation T has k unit roots, corresponding to k stochastic trends. Therefore the covariance matrix of the initial value of the state vector, $\alpha_{1|0}$ will be unbounded. A consistent approach to this problem is the *exact initial Kalman filter* developed by Koopman (1997) and refined by Koopman and Durbin (2001). For the UC model the *diffuse* initial state vector can be defined as

$$\alpha_{1|0} = \begin{bmatrix} \delta_{k\times 1} \\ 0_{2l\times 1} \end{bmatrix} + \nu_0, \tag{D-7}$$

where $\delta \sim N(0, \kappa I_k)$, $\kappa \to \infty$. The covariance matrix of the initial state vector, $P_{1|0}$ can be split into an unbounded component κP_∞ pertaining to the stochastic trends and a bounded component associated with the stationary component P_*

$$P_{1|0} = \kappa P_\infty + P_*, \qquad \kappa \to \infty, \tag{D-8}$$

where

$$P_\infty = \begin{bmatrix} I_k & 0_{k\times 2l} \\ 0_{2l\times k} & 0_{2l} \end{bmatrix} \quad \text{and} \quad P_* = RQR'. \tag{D-9}$$

The stationary component P_* can be initialized as the steady state value

$$vec(P_*) = (I_{(k+2l)^2} - T^* \otimes T^*)^{-1} vec(RQR'), \tag{D-10}$$

where $T^* \equiv T - P_\infty$ is the stationary component of the transition matrix.

Univariate treatment of the multivariate series

Koopman's original solution approach to the exact initialization problem can be simplified by using the univariate treatment of multivariate series originally suggested by Anderson and Moore (1979). Apart from avoiding a complicated factorization problem, the univariate treatment also leads to considerable computational efficiency gains, especially if the dimension of the system n is large (see Durbin and Koopman (2001), ch. 6.4 for more details).

Forecasting

Forecasts are obtained by setting the innovations v_t equal to zero during the forecast window. 95% confidence intervals for variable j can be based on $\pm 1.96 F_t(jj)$. This method is equivalent to the treatment of missing observations.

Forecast error variance decompositions

The forecast variance decomposition can be computed as follows: Using Cholesky-factorization we can write $RQR' = RCC'R'$. Setting $P_t = 0$ and using a recursive argument, it can be shown that

$$P_{t+j} = \sum_{i=1}^{j} T^j CC'T'^j. \tag{D-11}$$

The forecast variance at period j is then given by $F_{t+j} = ZP_{t+j}Z'$. The forecast variance that is attributed to the k-th orthogonalized innovation is then obtained by computing

$$\begin{aligned}\bar{P}_{t+j,k} &= \sum_{i=1}^{j} T^j C E_k C' T'^j &\text{(D-12)}\\ \bar{F}_{t+j,k} &= Z\bar{P}_{t+j}Z', &\text{(D-13)}\end{aligned}$$

where E_k is a square matrix of zeros with a 1 in the k-th diagonal entry.

E Nyblom and Harvey (2000), A Test for Common Stochastic Trends

Based on the multivariate local level model

$$\begin{aligned}y_t &= \tau_t + \epsilon_t &\text{(E-1)}\\ \tau_t &= \tau_{t-1} + \eta_t, \qquad \eta_t \sim NID(0, \Sigma_\eta),\end{aligned}$$

we test for the null hypothesis H_0 : rank$(\Sigma_\eta) = k$ $(k < n)$ versus the alternative hypothesis H_1 : rank$(\Sigma_\eta) > k$.

Let $\lambda_1 \leq ... \leq \lambda_n$ be the ordered eigenvalues of $S^{-1}C$, where

$$\begin{aligned}C &= T^{-2}\sum_{j=1}^{T}\left[\sum_{t=1}^{j}(y_t - \bar{y})\right]\left[\sum_{t=1}^{j}(y_t - \bar{y})\right]'\\ S &= T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y}).\end{aligned}$$

Then the test statistic is given by the sum of the $(n-k)$ smallest eigenvalues

$$\zeta_{k,n} = \lambda_{k+1} + ... + \lambda_n. \tag{E-2}$$

Nyblom and Harvey (Theorem 4) show that under the null hypothesis the limiting distribution of $\zeta_{k,n}$ is given by

$$\zeta_{k,n} \xrightarrow{d} tr\left[C_{22}^* - C_{11}^{*\,-1}C_{12}^*\right] \tag{E-3}$$

$$
\begin{aligned}
C_{11}^{*} &= \int_0^1 \left[\int_0^u W^*(s)ds\right]\left[\int_0^u W^*(s)ds\right]' du \\
C_{12}^{*} &= C_{21}^{*}{}' = \int_0^1 \left[\int_0^u W^*(s)ds\right] B(u)du \\
C_{22}^{*} &= \int_0^1 B(u)B(u)du.
\end{aligned}
$$

where $B(u)$ is an r-dimensional standard Brownian bridge ($r = n - k$), $W^*(u) = W(u) - \int_0^1 W(s)ds$ ($u \in [0,1]$) and $W(u)$ is a K-dimensional standard Brownian motion.

Nyblom and Harvey call the distribution in (E-3) the Cramér-von Mises (CvM(k,n)) distribution and provide critical values obtained by simulation.

To account for serial correlation in ϵ_t, S can be substituted by a non-parametric estimate of the spectral density at the zero frequency

$$
S(m) = \sum_{j=-m}^{m} w_j \Gamma_j, \tag{E-4}
$$

where w is some weighting function (e.g. $w = 1 - j/(1+m)$) and $\Gamma_j = T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y})$.

F Monte Carlo evidence

In section 3 it is asserted that likelihood ratio tests for the comparison of hypotheses with different numbers of common cycles, but the same number of common trends have limiting χ^2 distributions. A more rigorous proof of this statement, based on the autocovariance generating function of the Kalman filter, is work in progress. This appendix provides evidence from a Monte-Carlo simulation of tests for the null $H_0 : M(1,1)$ (1 trend and 1 cycle) versus the alternative $H_1 : M(1,2)$ (1 trend and 2 cycles). This is equivalent to a comparison between model 4 and model 3 in the empirical part (section 4). The data was generated under the null with parameters $\sigma_\eta^2 = 0.64$, $\sigma_\varepsilon^2 = 3$, $\sigma_{\eta\varepsilon} = -0.3$, $\phi_{1,1} = 0.7$, $\phi_{1,2} = -0.2$, $\gamma = [1, 1.4]'$, and $\tilde{\gamma} = [1, 0.5]'$. The sample size was set to $T = 100$. Because there are 4 additional parameters under H_1 compared to H_0 (see table 1), the expected limiting distribution is $\chi^2(4)$. A comparison shows that empirical and theoretical critical values are close, however the empirical distribution is slightly leptokurtic. Since the fat tails may be caused by the relatively small sample size, $\chi^2_{100}(4)$ provides critical values of a "small sample χ^2" distribution, generated from squared sums of random variables drawn from a t-distribution with 100 degrees of freedom[19].

significance level	$\chi^2(4)$	$\chi^2_{100}(4)$	empirical
10 %	7.78	7.94	7.80
5 %	9.49	9.72	9.77
1 %	13.28	13.73	13.84

Number of replications: 15.000.

References

Ahn, S. K., and G. C. Reinsel (1990): "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

Anderson, B. D. O., and J. B. Moore (1979): *Optimal Filtering*. Prentice Hall, Englewood Cliffs.

Banerjee, A., J. J. Dolado, J. W. Galbraith, and D. F. Hendry (1993): *Co-integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, New York, NY.

Beveridge, S., and C. Nelson (1981): "A new aproach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle," *Journal of Monetary Economics*, 7, 151–174.

Blanchard, O., and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655–673.

[19]This distribution was generated using 100.000 Monte-Carlo iterations.

Campbell, J., and N. Mankiw (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomic Annual*, ed. by O. Blanchard, and S. Fischer, pp. 185–216. MIT Press.

——— (1990): "Permanent Income, Current Income and Consumption," *Journal of Business and Economic Statistics*, 8(3), 265–279.

Clark, P. (1987): "The cyclical component of U.S. economic activity," *Quarterly Journal of Economics*, 102, 797–814.

Deaton, A. (1987): "Life cycle models of consumption: is there evidence consistent with the theory," in *Advances in Econometrics: Fifth World congress, vol. 2*, ed. by T. Bewley. Cambridge University Press, New York.

Durbin, J., and S. Koopman (2001): *Time Series Analysis by State Space Methods*. Oxford University Press, New York.

Engle, R., and C. Granger (1987): "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 35, 251–276.

Engle, R. F., and J. V. Issler (1995): "Estimating Common Sectoral Cycles," *Journal of Monetary Economics*, 35, 83–113.

Engle, R. F., and S. Kozicki (1993): "Testing for Common Features," *Journal of Business and Economics Statistics*, 11, 369–380.

Flavin, M. (1981): "The Adjustment of Consumption to Changing Expectiation about Future Income," *Journal of Political Economy*, 89, 974–1009.

——— (1993): "The excess smoothness of consumption: identification and interpretation," *Review of Economic Studies*, 60(3), 651–666.

Friedman, M. (1957): *A Theory of the Consumption Function*. Princeton University Press, Princeton, NJ.

Gonzalo, J., and C. Granger (1995): "Estimation of Long-Memory Components in Cointegrated Systems," *Journal of Business and Economic Statistics*, 13, 27–35.

Granger, C. W. J., and P. Newbold (1986): *Forecasting Economic Time Series, Second Edition*. Academic Press.

Hall, R. (1978): "Stochastic Implications of the Life-Cycle Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971–987.

Harvey, A. (1981): *Time Series Models*. Oxford University Press, Oxford, UK.

Harvey, A. (1985): "Trends and Cycles in Macroeconomic Time-Series," *Journal of Business and Economic Statistics*, 3, 216–227.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, UK.

Issler, J. V., and F. Vahid (2001): "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," *Journal of Monetary Economics*, 47, 449–475.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics & Control*, 12, 231–254.

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Kalman, R. E. (1960): "A new approach to linear filtering and prediction problems," *J. Basic Engineering, Transactions ASMA, Series D*, 82, 35–45.

King, R. B., C. I. Plosser, J. H. Stock, and M. W. Watson (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

Koopman, S. J. (1997): "Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models," *Journal of the American Statistical Association*, 92, 1630–1638.

Morley, J. C., C. Nelson, and E. Zivot (2002): "Why are Beveridge-Nelson and Unobserved Component Decompositions of GDP so Different?," *Review of Economics and Statistics (forthcoming).*

Nyblom, J., and A. Harvey (2000): "Tests of Common Stochastic Trends," *Econometric Theory*, 16, 176–199.

Peña, D., and P. Poncela (2002): "Dimension Reduction in Multivariate Time Series," manuscript.

Politis, D., and J. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Proietti, T. (1997): "Short-Run Dynamics in Cointegrated Systems," *Oxford Bulletin of Economics and Statistics*, 59, 405–422.

——— (2002): "Some Reflections on Trend-Cycle Decompositions with Correlated Components," Dip. di Scienze Statistiche, University of Udine.

Stock, J., and M. Watson (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

Vahid, F., and R. F. Engle (1993): "Common Trends and Common Cycles," *Journal of Applied Econometrics*, 8, 341–360.

Watson, M. W. (1986): "Univariate detrending methods with stochastic trends," *Journal of Monetary Economics*, 18, 49–75.

White, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68, 1097–1126.

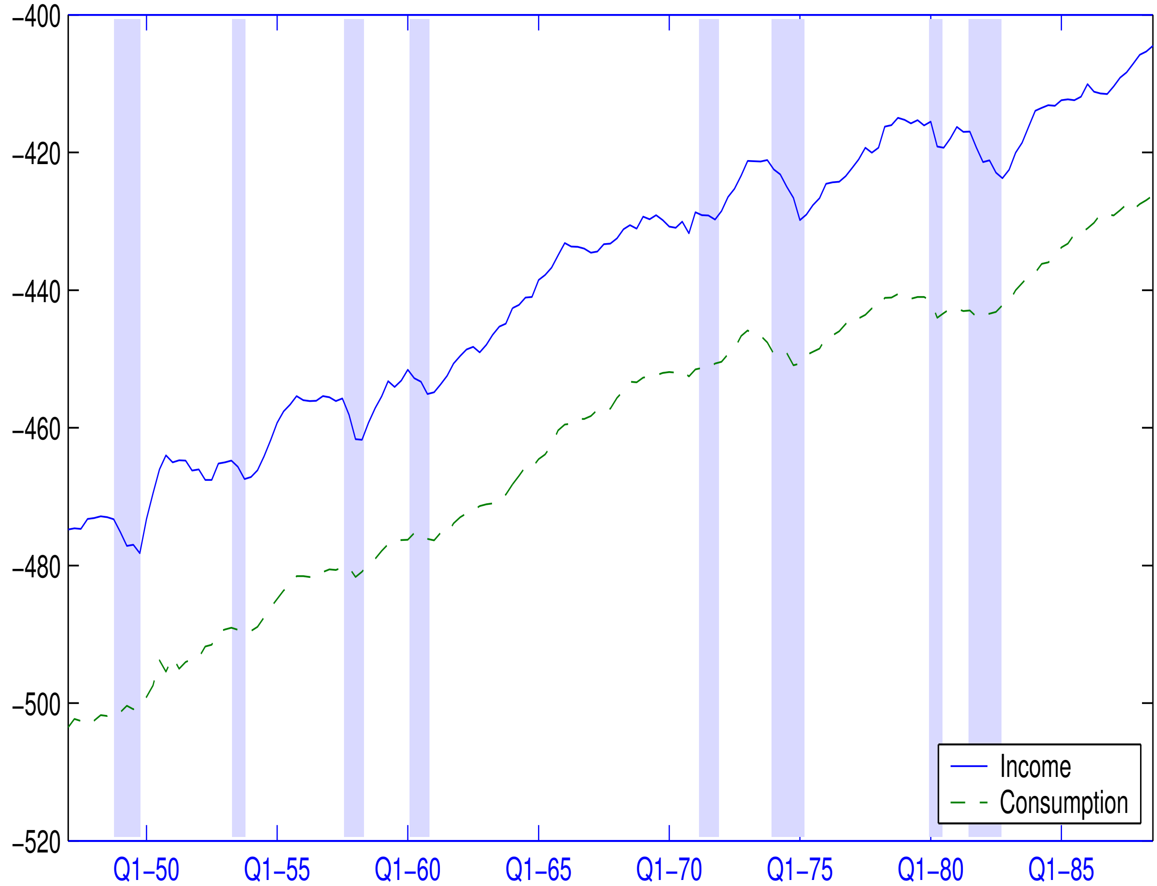


Figure 1: Per-capita private GDP and private consumption from the dataset of King et al. (1991), NBER recessions are shaded.

Table 1: Regression Results for Models 1 to 4

	Model 1 $M(2,2)$		Model 2 $M(1,2)$		Model 3 $M(2,1)$		Model 4 $M(1,1)$	
cycle								
$\phi_{1,1}$	1.55	*(0.03)*	1.06	*(0.07)*	1.40	*(0.11)*	1.13	*(0.06)*
$\phi_{1,2}$	-0.76	*(0.09)*	-0.26	*(0.07)*	-0.59	*(0.13)*	-0.26	*(0.06)*
$\phi_{2,1}$	1.47	*(0.06)*	1.23	*(0.10)*	-	-	-	-
$\phi_{2,2}$	-0.72	*(0.12)*	-0.49	*(0.16)*	-	-	-	-
$\sigma^2_{\varepsilon_1}$	0.52	*(0.33)*	2.35	*(0.76)*	0.92	*(0.45)*	2.93	*(0.79)*
$\sigma^2_{\varepsilon_2}$	0.20	*(0.23)*	0.36	*(0.40)*	-	-	-	-
$\sigma_{\varepsilon_1\varepsilon_2}$	0.29	*(0.18)*	0.82	*(0.56)*	-	-	-	-
trend								
$\sigma^2_{\eta_1}$	1.12	*(0.71)*	0.92	*(0.54)*	1.10	*(0.60)*	0.64	*(0.42)*
$\sigma^2_{\eta_2}$	0.91	*(0.43)*	-	-	0.82	*(0.26)*	-	-
$\sigma_{\eta_1\eta_2}$	0.43	*(0.42)*	-	-	0.53	*(0.32)*	-	-
cross-terms								
$\sigma_{\eta_1\varepsilon_1}$	-0.27	*(0.55)*	-0.91	*(0.65)*	-0.41	*(0.51)*	-1.04	*(0.58)*
$\sigma_{\eta_2\varepsilon_1}$	-0.36	*(0.45)*	-	-	-0.59	*(0.36)*	-	-
$\sigma_{\eta_1\varepsilon_2}$	-0.17	*(0.23)*	-0.47	*(0.45)*	-	-	-	-
$\sigma_{\eta_2\varepsilon_2}$	-0.36	*(0.45)*	-	-	-	-	-	-
γ	-	-	1.15	*(0.17)*	-	-	1.41	*(0.28)*
$\tilde{\gamma}$	-	-	-	-	0.54	*(0.10)*	0.50	*(0.07)*
parameters	14		11		9		7	
$\log L$	-438.26		-441.00		-443.07		-445.19	

Standard errors are given in parentheses.

Table 2: Tests for Common Trends and Common Cycles

H_0	H_1	Type	Stat	Dist	10% val.	5% val.	1% val.
$M(1,2)$	$M(2,2)$	NH	0.136	CvM(1,2)	0.162	0.218	0.383
$M(1,1)$	$M(2,1)$	NH	0.117	CvM(1,2)	0.162	0.218	0.383
$M(2,1)$	$M(2,2)$	LR	9.608	$\chi^2(5)$	9.236	11.070	15.086
$M(1,1)$	$M(1,2)$	LR	8.375	$\chi^2(4)$	7.779	9.488	13.277

NH = Test by Nyblom and Harvey (2000), LR = Likelihood Ratio Test

Table 3: Out of Sample Mean Squared Prediction Errors

	Model 1	Model 2	Model 3	Model 4
	$M(2,2)$	$M(1,2)$	$M(2,1)$	$M(1,1)$
income	6.396	1.185	3.917	5.256
consumption	9.290	4.586	5.022	0.486
|MSE|	13.686	4.938	8.505	1.494

Table 4: Evaluation of Forecast Performance

	Model 2	Model 3	Model 4
	$M(1,2)$	$M(2,1)$	$M(1,1)$
Model 1 $M(1,2)$	0.0016	0.0544	0.0000
Model 2 $M(1,2)$		0.9104	0.0676
Model 3 $M(1,2)$			0.0043

p-values based on the boostrap test by White (2000).
Null hypothesis: Model i outperforms model j ($i < j$).
Evaluation metric: |MSE|, 50000 stationary bootstrap resamples.

Table 5: Forecast Error Variance Decompositions for Income

	Proportion of Permanent Shocks					
Horizon	*M(2,2)*	*M(1,2)*	*M(2,1)*	*M(1,1)*	*K*	*IV*
1	67	0	59	16	*45*	*32*
4	48	3	50	17	*58*	*38*
8	68	29	67	16	*68*	*51*
12	77	50	78	25	*73*	*58*
16	82	63	83	38	*77*	*67*
20	85	70	86	48	*79*	*76*
24	87	76	88	56		
28	89	79	90	62		
32	90	82	91	67		
36	91	84	92	71		
∞	100	100	100	100	*100*	*100*

Percentage of the forecast error variance attributed to the orthogonalized trend innovation.
K and IV represent the results of King et al. (1991) and Issler and Vahid (2001), respectively.

Table 6: Forecast Error Variance Decompositions for Consumption

	Proportion of Permanent Shocks					
Horizon	*M(2,2)*	*M(1,2)*	*M(2,1)*	*M(1,1)*	*K*	*IV*
1	85	77	67	42	*88*	*65*
4	73	79	51	42	*89*	*77*
8	92	93	79	68	*83*	*85*
12	96	96	88	81	*83*	*88*
16	97	97	91	87	*85*	*90*
20	97	98	93	91	*87*	*93*
24	98	98	94	93		
28	98	99	95	94		
32	98	99	96	95		
36	99	99	96	96		
∞	100	100	100	100	*100*	*100*

Percentage of the forecast error variance attributed to the orthogonalized trend innovation.
K and IV represent the results of King et al. (1991) and Issler and Vahid (2001), respectively.

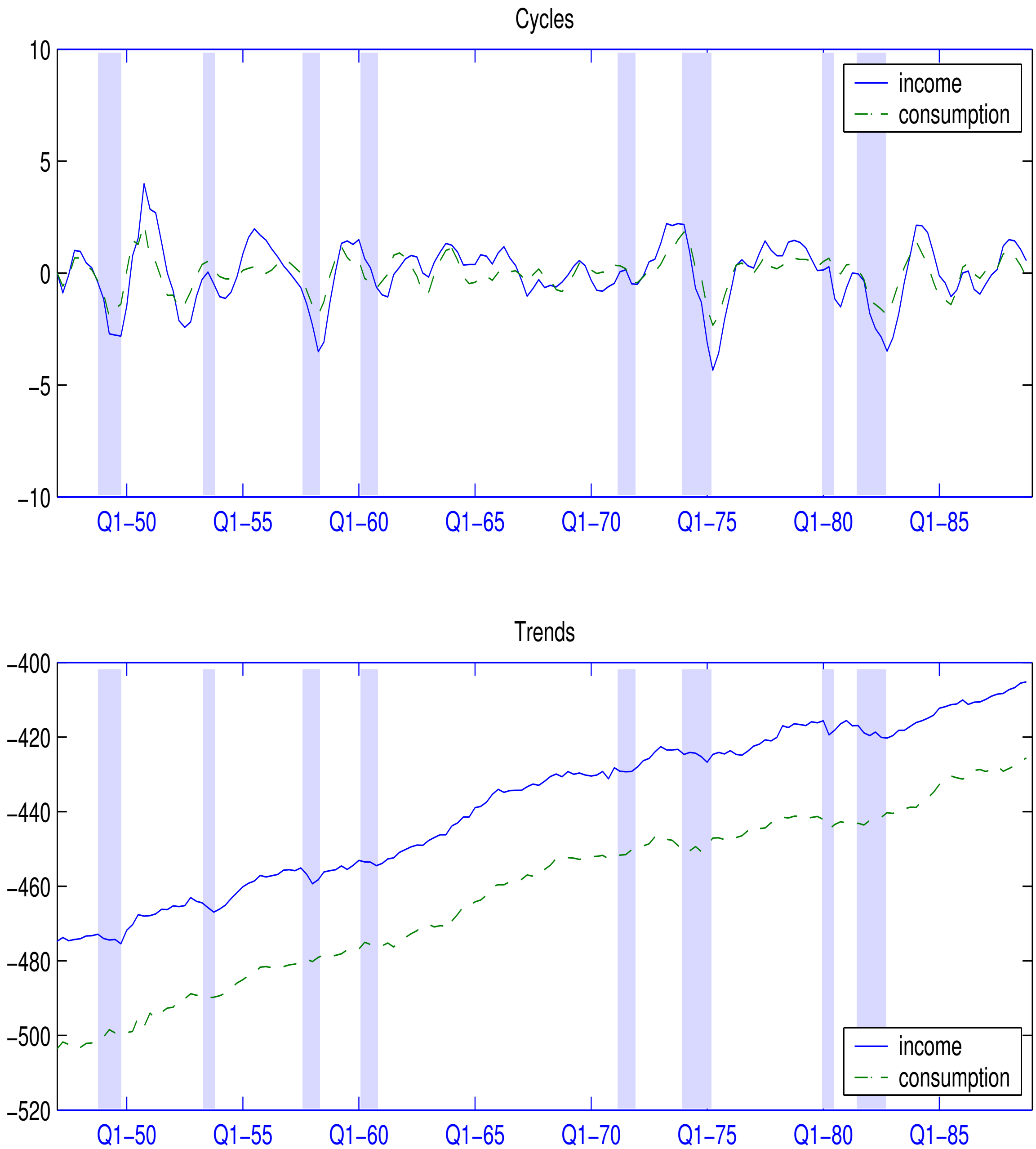


Figure 2: Model 1 (2 trends and 2 cycles).

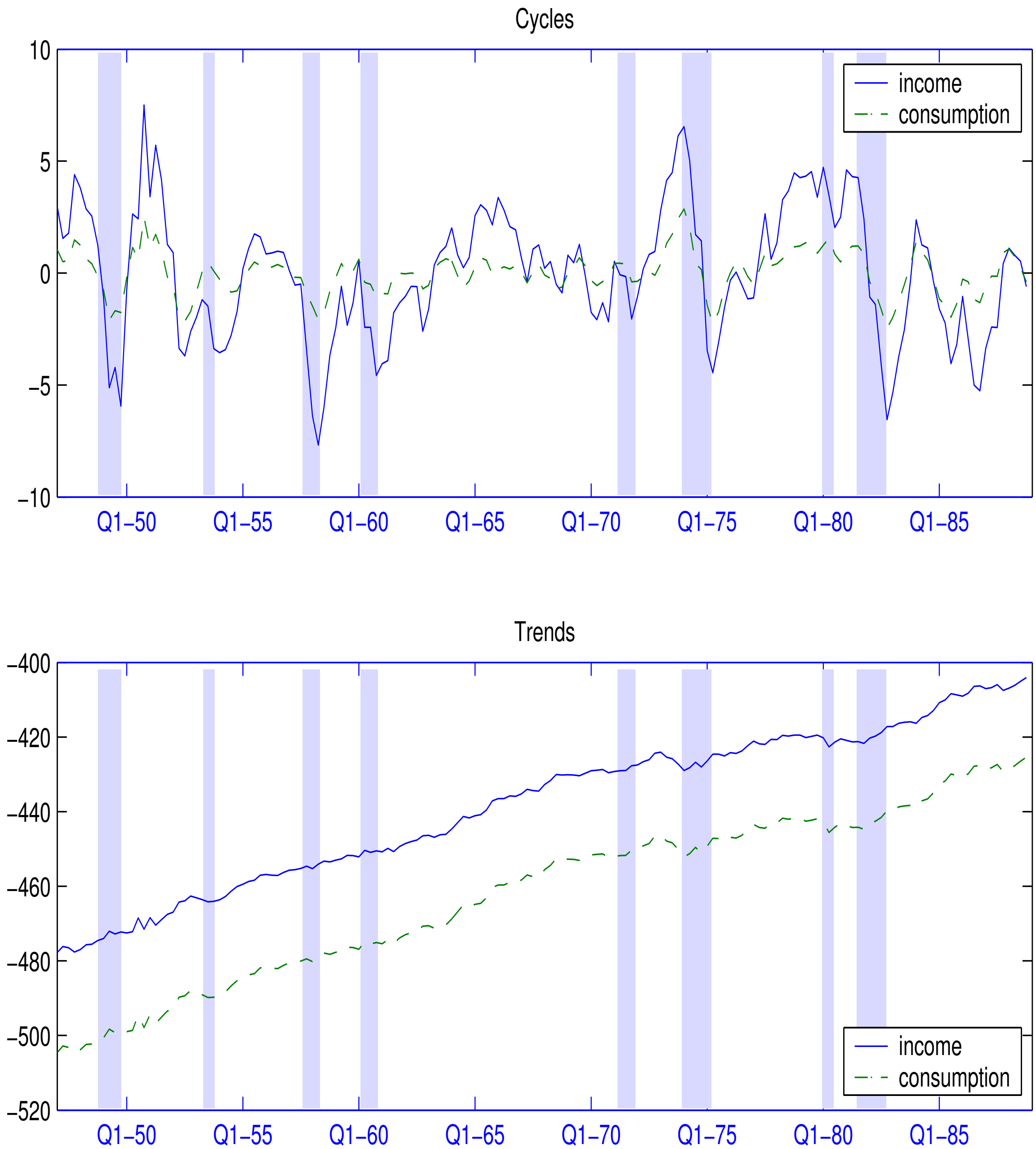


Figure 3: Model 2 (1 trend and 2 cycles).

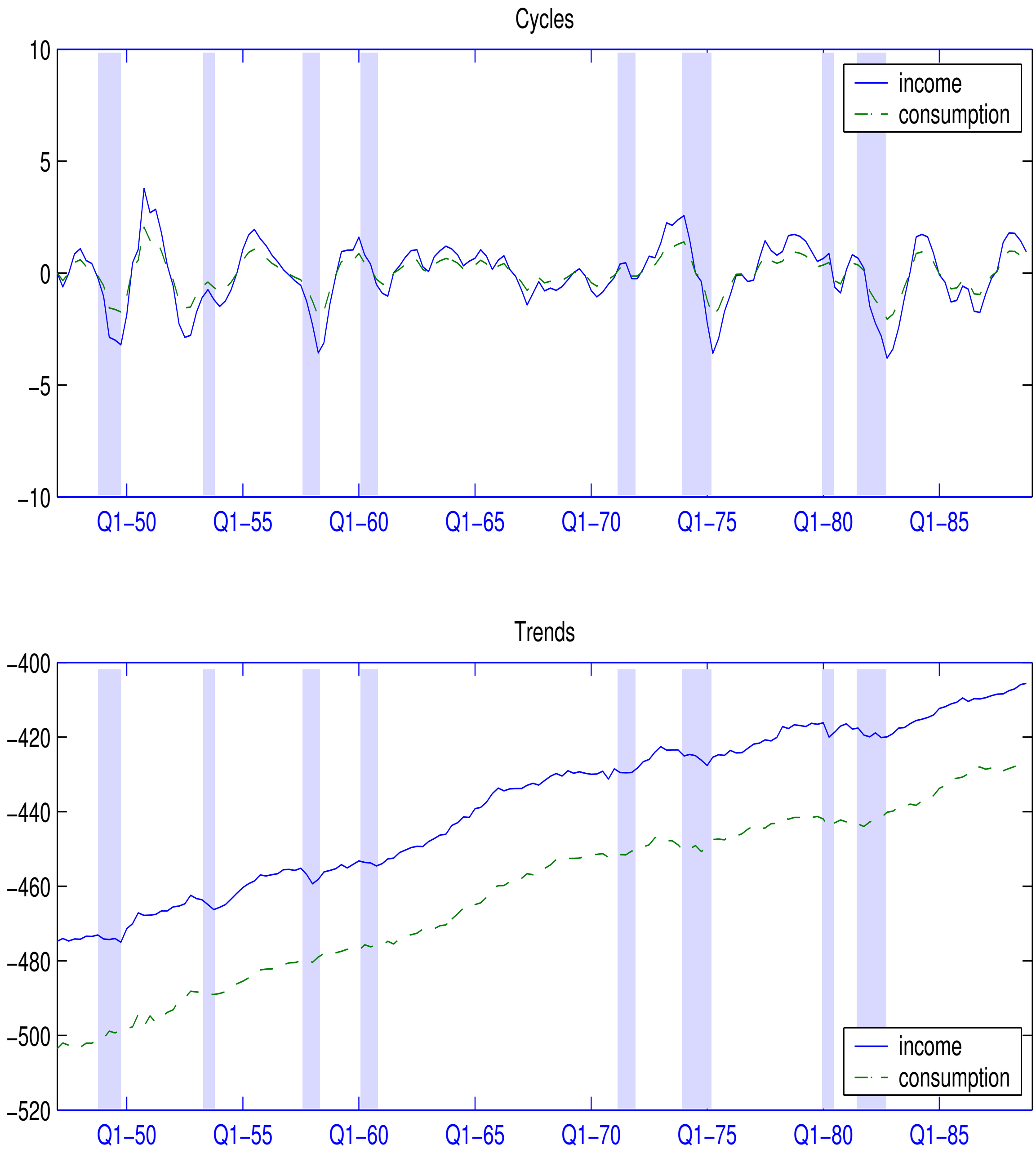


Figure 4: Model 3 (2 trends and 1 cycle).

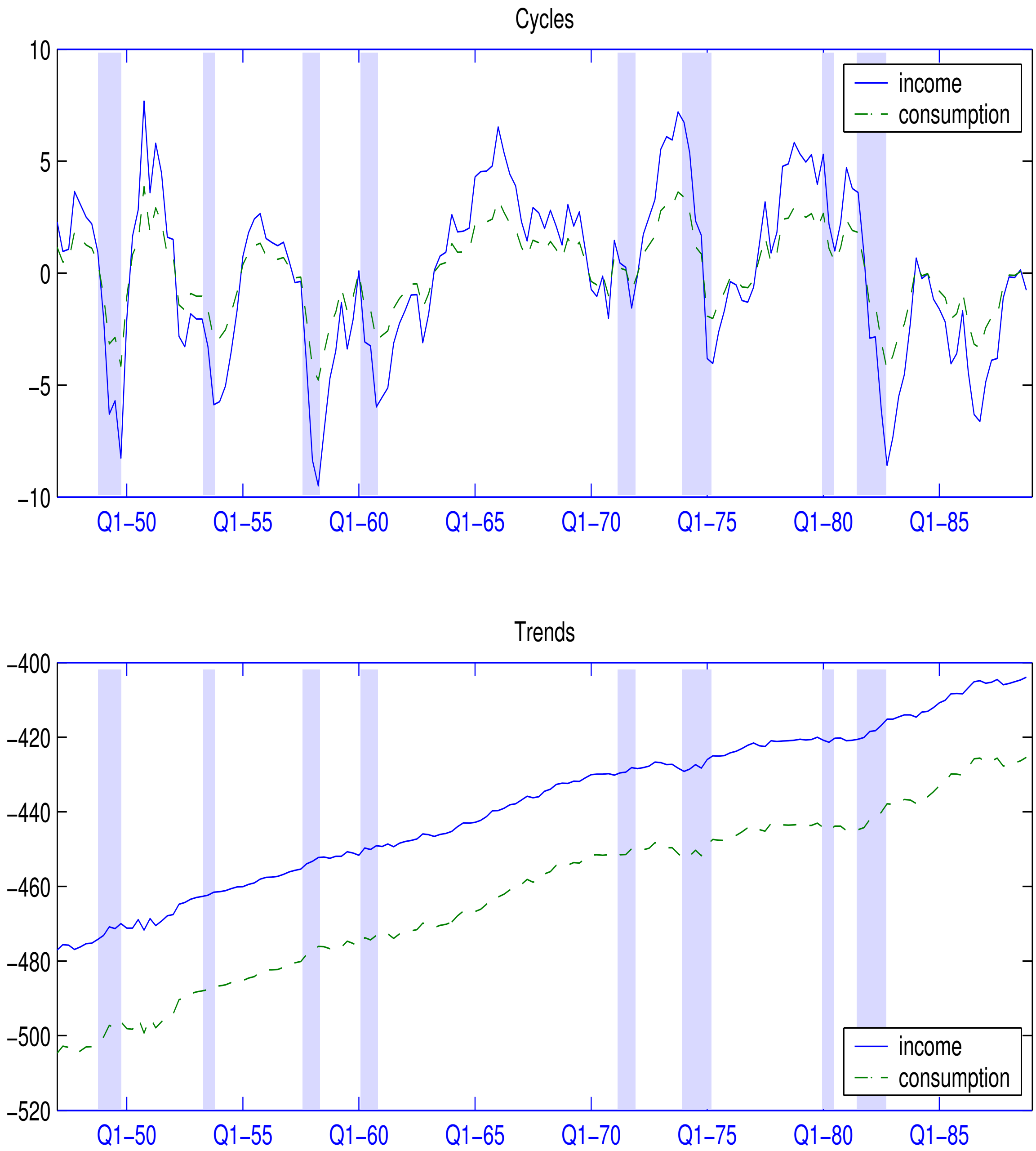


Figure 5: Model 4 (1 trend and 1 cycle).

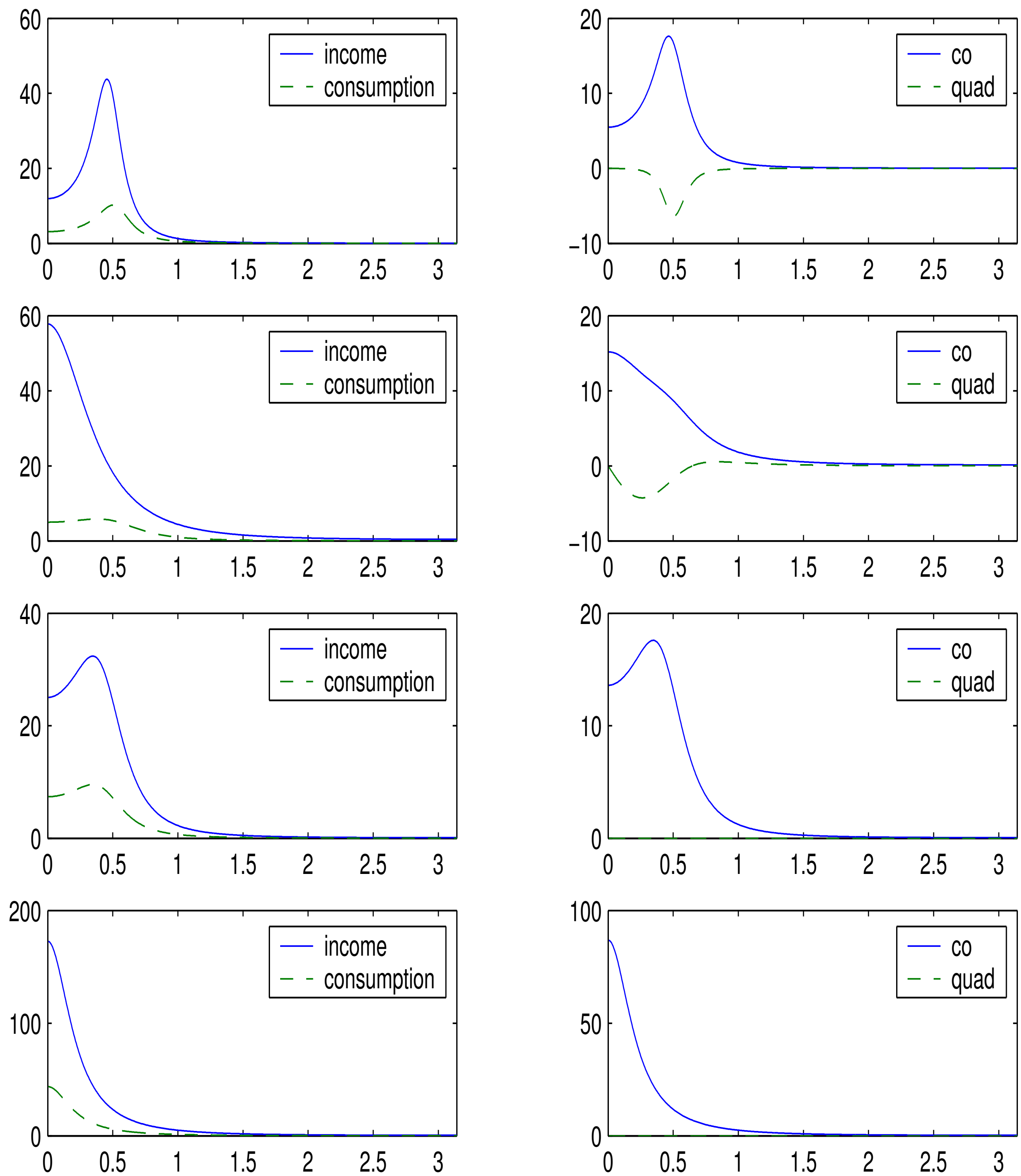
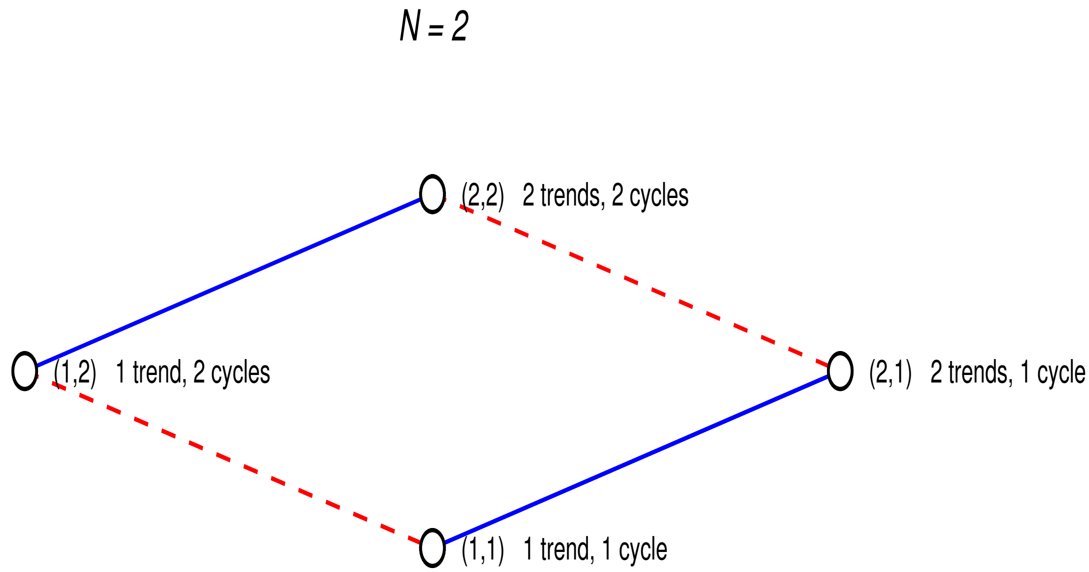


Figure 6: Spectra (*left*) and cross-spectra (*right*) of cyclical component (rows 1-4 correspond to models 1-4).



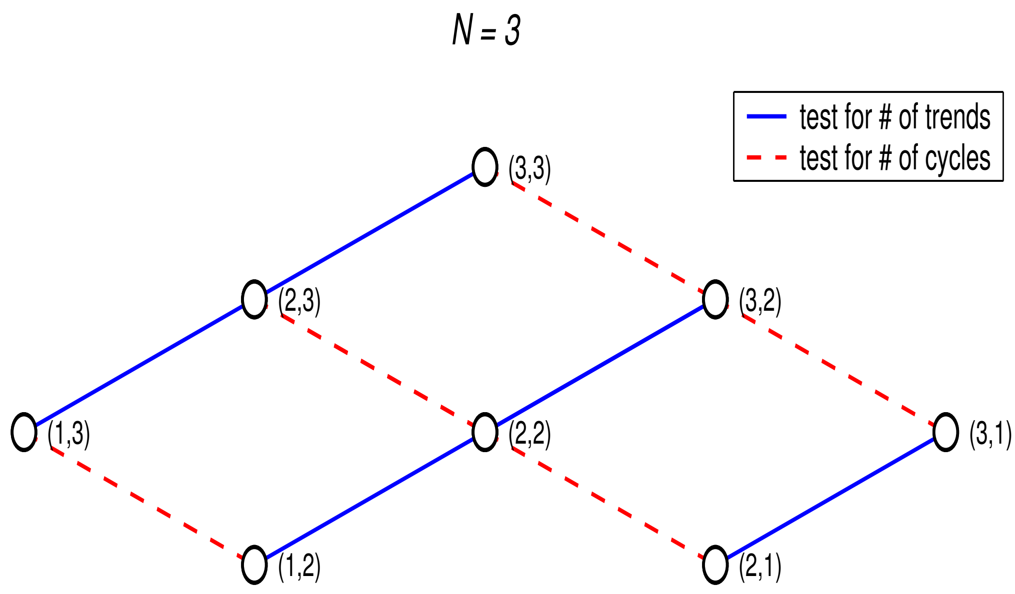


Figure 7: Model selection for a time series of dimension 2 (*top*) and 3 (*bottom*). Tests between nested alternatives are only possible on a vertical level (indicated by connecting lines), but not on a horizontal level.

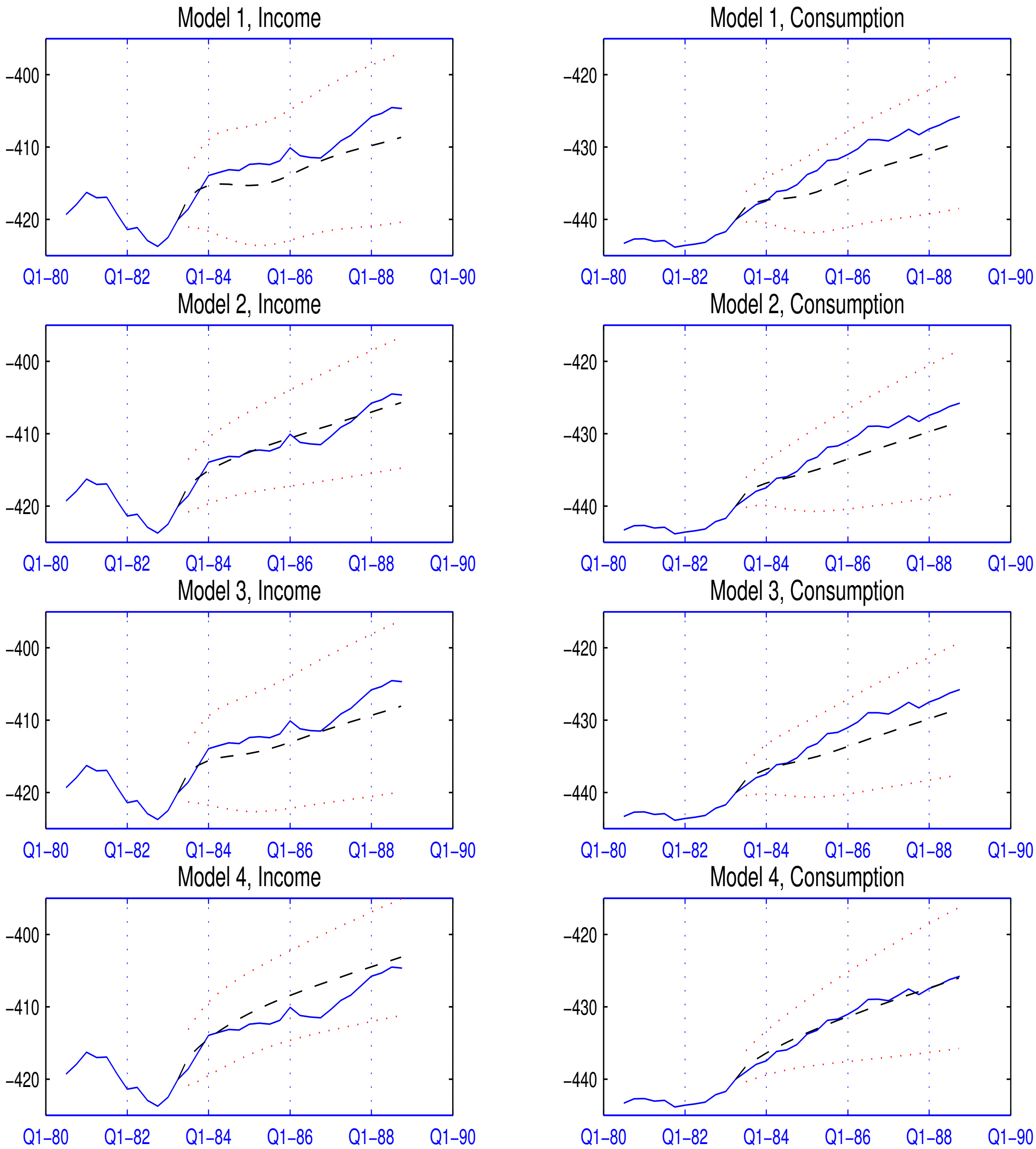


Figure 8: Out of sample forecasts with 95 percent confidence intervals.