# Nonlinear Permanent -Temporary Decompositions in Macroeconomics and Finance

**Richard H. Clarida**

Columbia University
and
National Bureau of Economic Research

**Mark P. Taylor**

University of Warwick
and
Centre for Economic Policy Research

**Abstract**

We suggest a method of decomposing univariate and multivariate nonlinear processes into their permanent and temporary components, extending the analysis of Beveridge and Nelson (1981) and Stock and Watson (1987). We provide an application in the univariate nonlinear case to recent work on nonlinearities in the US business cycle, and in the multivariate nonlinear case to recent work on asymmetric nonlinear adjustment in the US term structure of interest rates.

**JEL Classifications :** C22, C32, E32, E43.
**Keywords:** nonlinearity, decomposition, business cycle, interest rate term structure.

## 1. Introduction[1]

For both theoretical and empirical reasons, it is frequently desirable to decompose an economic time series into the sum of unobservable permanent and temporary components that are thought to generate the observed series. For integrated processes that have a linear autoregressive integrated moving average (ARIMA) representation, a popular method of obtaining this decomposition is the Beveridge-Nelson (1981) decomposition.[2] The Beveridge-Nelson method decomposes an ARIMA process into the sum of a random walk and a stationary stochastic process with a linear ARMA representation. The Beveridge-Nelson permanent component is defined as the long-horizon forecast of the level of the time series (suitably adjusted for drift) and is in practice often calculated in just this way, using an estimated ARIMA time series model for the series to be decomposed.[3] Stock and Watson (1987) subsequently generalized the Beveridge-Nelson decomposition to the multivariate case. In this paper, we propose a further extension of the Beveridge-Nelson decomposition by showing how it may be applied to a particular class of univariate and multivariate nonlinear processes, while still yielding the univariate linear Beveridge-Nelson (1981) and multivariate Stock-Watson (1987) decompositions as special cases. We also provide illustrative applications of the proposed generalized decomposition in both the univariate and multivariate cases.

The remainder of the paper is set out as follows. In the next section we show how the Beveridge-Nelson decomposition may be applied in the nonlinear case, what restrictions must be satisfied by the nonlinear process under consideration for it to be applicable and how to check whether these restrictions are satisfied. Although there turn out to be important practical difficulties in deriving an analytical solution for the permanent-temporary decomposition in the nonlinear case, we show how these may be overcome in practice using Monte Carlo integration. In Section 3 we provide an application of the technique, in a univariate nonlinear setting, to the US business cycle. In Section 4 we provide a multivariate nonlinear application to the US term structure of interest rates. A final section concludes.

## 2. Applying the Beveridge-Nelson decomposition to nonlinear processes

Beveridge and Nelson (1981) propose a method for decomposing a univariate time series into its permanent and temporary components where the permanent component is defined as the long-horizon level forecast of the series, or the part that remains after all transitory dynamics have worked themselves out. This was generalized to the multivariate case by Stock and Watson (1987) and applied, for example, by Cochrane (1994). For a multivariate, Nx1 vector Z(t), the trend or permanent component, $Z^{\pi}(t)$, may be defined:

$$Z^{p}(t) = \lim_{k\to\infty} E\left[Z(t+k)\middle|\Omega(t)\right] - k\left\{E\left[Z(t+k)\middle|\Omega(t)\right] - E\left[Z(t+k-1)\middle|\Omega(t)\right]\right\}. \qquad (1)$$

Where $\Omega(t) \equiv \{Z(t), Z(t-1), Z(t-2), \ldots\}$ is the information set consisting of current and lagged values of the time series vector. The transitory or cycle component, $Z^{\tau}(t)$, is defined as the complement to the permanent component :

$$Z^{t}(t) = Z(t) - Z^{p}(t)\,. \qquad (2)$$

The interpretation of (1) is quite intuitive. The permanent component is simply the level that is expected to prevail in the long run, minus that part of the expected long-run level which is due to a purely deterministic trend component between now and then. If the vector time series is not explosive, then for large enough k, Z(t+k) will enter steady state such that it is expected to increase each period only by a constant drift of, say, μ per period:

$$\lim_{k\to\infty}\left\{E\left[Z(t+k)\middle|\Omega(t)\right] - E\left[Z(t+k-1)\middle|\Omega(t)\right]\right\} = \mu\,, \qquad (3)$$

where μ is a vector of finite constants. Hence, the last bracketed term in (1) extracts the deterministic trend component that has accumulated in the long-horizon prediction. Further insight into the Beveridge-Nelson decomposition may be obtained by substituting (3) into (1) and rearranging :

$$Z^{p}(t) = Z(t) + \lim_{k\to\infty} \sum_{j=1}^{k} \left(E\left[(1-L)Z(t+j)\middle|\Omega(t)\right] - \mu\right), \qquad (4)$$

where L is the lag operator. Thus, the Beveridge-Nelson permanent or trend component of Z(t) is just Z(t) itself plus all expected future growth in Z(t) above the deterministic growth rate. If an element of Z(t) is expected to grow above its average growth rate in the future (i.e. the corresponding element of μ) then that element of Z(t) will be below its permanent or trend level.

In this paper we consider an extension of the Beveridge-Nelson decomposition to the case where the data generating process for Z(t) is nonlinear. In fact, the extension we suggest is quite straightforward, and follows from calculating the permanent and temporary components of Z(t) as defined in (1) and (2) but with explicit account taken of the nonlinear data generating process for Z(t). The justification for treating the component of Z(t) defined in (1), in the nonlinear case, as the permanent component of Z(t) is exactly as in the linear case. The long-horizon prediction less any component which has accumulated due to a purely deterministic trend component must be equal to the level which the series is expected to reach after all transitory dynamics have worked themselves out.

Of course, for the Beveridge-Nelson decomposition to be applicable, the long-horizon prediction must be stable in the sense that condition (3) is satisfied. In the linear case, the fact that Beveridge and Nelson and Stock and Watson consider only I(1) processes means that the condition (3) automatically holds. In considering the nonlinear case, we need to place some analogous restrictions on the limiting behaviour of the series. For the purposes of this paper, we shall say that a process whose long-horizon prediction is not a constant but for which the long-horizon prediction of the first differences of the process satisfies (3) is *integrated short memory in mean,* or ISMM.[4] In this paper we consider only the class of processes which are integrated short memory in mean. Clearly, the class of ISMM processes includes linear I(1) processes but also allows for a much more general range of processes, including nonlinear processes. With this restriction on the class of processes considered, the Beveridge-Nelson decomposition as defined in (1) and (2) may be generalized to the class of processes which are ISMM, including nonlinear ISMM processes.

While there are well-established methods for checking the order of integration of linear processes, the same is not true for nonlinear processes. A tractable, albeit informal method of checking whether or not a process is ISMM would, however, simply be to overplot the conditional forecasts of the first difference series, produced either parametrically or using a nonparametric estimator (Robinson, 1983), for a range of initial conditions, from one period ahead to a large number of steps ahead, into a ‘profile bundle’ (Gallant, Rossi and Tauchen, 1993). If the process is ISMM, then each forecast profile should tend to enter a well defined steady state, so that the bundle should converge on a single value – or ‘flatline’ - for conditional forecasts a sufficiently large number of periods ahead (Gallant, Rossi and Tauchen, 1993). In discussing this method, Gallant, Rossi and Tauchen (1993, p. 884) note: ‘this seems to be the only practical strategy for investigating issues of integration in a model fitted to a general nonlinear process.’

While the generalization of the Beveridge-Nelson decomposition to ISMM nonlinear processes seems straightforward, in practice computing the permanent-temporary decomposition is far from so, because of the inherent difficulty in computing the long-horizon predictions in the presence of nonlinearity. The difficulties inherent in producing multi-step forecasts for nonlinear processes have been well documented elsewhere, both in discussions of forecasting itself and in discussions relating to the construction of impulse response functions for nonlinear dynamic models (see e.g. Granger and Teräsvirta, 1993; Gallant, Rossi and Tauchen, 1993; Granger, 1999), so that our discussion of this issue here can be relatively brief. The general problem is that, unlike as in forecasting with linear models, future disturbances cannot be ignored because one is implicitly forming conditional expectations of *nonlinear* functions of white noise disturbances which will, in general, not be zero. Thus, deriving the analytic expression for the multi-step forecast involves multiple integration over the density of the stochastic error term and, in general this will be extremely difficult or non-tractable (see, for example, Granger and Teräsvirta, 1993, chapter 8, or Gallant, Rossi and Tauchen, 1993 for further discussion).

One way of circumventing this problem is by evaluating the integrals by Monte Carlo methods. Effectively, this is done by replacing the exact conditional expectations by a set of ‘Monte Carlo forecasts’. Suppose, for example, that Z(t) is a univariate ISMM process, such that (1-L)Z(t) is a nonlinear function $\Theta[.]$ of its own past values, plus an additive white noise disturbance:

$$(1-\mathrm{L})\mathrm{Z}(\mathrm{t}) = \Theta\left[(1-\mathrm{L})\mathrm{Z}(\mathrm{t}-1), (1-\mathrm{L})\mathrm{Z}(\mathrm{t}-2), \ldots\right] + e(\mathrm{t}) . \tag{5}$$

To form the multi-step forecasts by Monte Carlo integration, the Monte Carlo forecast, which we shall denote $\exists\left[.\middle|\Omega(\mathrm{t})\right]$, we proceed as follows. Equation (5) is simulated for k steps, starting with an initial condition of the known values of the lagged values of (1-L)Z(t), which are contained in the information set at time t, $\Omega(\mathrm{t})$, and let this be done a total of M replications. At each replication, k draws from an appropriate distribution for $e(t+i)$ for i=1,2,..k are required; denote this sequence of k draws for the m-th replication as $e_m$. Finally, denote the simulated value of (1-L)Z(t+k) in the m-th replication, which will be conditional on both $\Omega(\mathrm{t})$ and $e_m$, as $V_m\left[(1-\mathrm{L})\mathrm{Z}(\mathrm{t}+\mathrm{k})\middle|\Omega(\mathrm{t}),e_m\right]$. Then the Monte Carlo forecast may be denoted[5]:

$$\exists\left[(1-\mathrm{L})\mathrm{Z}(\mathrm{t}+\mathrm{k})\middle|\Omega(\mathrm{t})\right]=\mathrm{M}^{-1}\sum_{m=1}^{M}V_m\left[(1-\mathrm{L})\mathrm{Z}(\mathrm{t}+\mathrm{k})\middle|\Omega(\mathrm{t}),e_m\right]\ . \qquad (6)$$

By the law of large numbers, and under weak regularity conditions, the Monte Carlo forecast converges almost surely to the exact conditional expectation as the number of draws increases, so that for large M the exact and Monte Carlo forecasts should be identical (Granger and Teräsvirta, 1993 ; Gallant, Rossi and Tauchen, 1993).[6] If the process we are considering is ISMM and if we consider a sufficiently long forecast horizon, so that (3) holds to an arbitrarily close approximation for finite k (which is exactly what the profile bundle is designed to check), then the permanent component can be extracted as in (1) but with k set to this finite value.

If one is fairly confident about the underlying distribution of the stochastic error processes, then one can use fully parametric Monte Carlo simulation to construct the forecasts. If not, then one can use bootstrap methods based on sampling with replacement from the fitted residuals. Indeed, one might in some contexts wish to avoid fitting a parametric model to the series altogether and use a nonparametric estimator to construct the conditional forecasts themselves (Robinson, 1983; Gallant, Rossi and Tauchen, 1993).

The practical method we propose for deriving the Beveridge-Nelson permanent-temporary decomposition of a nonlinear ISMM process may therefore be summarized as follows. First, estimate a satisfactory nonlinear time series model for the process using standard nonlinear modeling procedures (Granger and Teräsvirta, 1993 ; Granger, 1999), or fit a nonparametric model. Second, use Monte Carlo methods applied to the estimated model to provide conditional forecast profiles for a range of initial conditions and use these to construct profile bundles to check that the process is ISMM. Third, use the long-horizon forecasts to construct the Beveridge-Nelson decomposition as in (1) and (2) but with the conditional expectations replaced by the Monte Carlo forecasts and with k set equal to a suitably large number.[7]

In constructing the generalized Beveridge-Nelson decomposition in this way, the issue naturally arises in any particular application as to how many replications should be used in the Monte Carlo integration and for how many periods ahead should the dynamic forecasts be constructed. These questions may be important if the underlying model is particularly complicated, for example if a high-dimensional nonlinear system is examined. The answer to both of these questions can be deduced, however, from examining the profile bundle, which is produced in the second step to check that the processes are ISMM. If the different profiles do not appear to converge to a single flat line profile after a sufficient number of steps ahead for each of the processes under consideration, it may be necessary to increase the number of replications. In the applications reported below, we also found the total number of replications needed in order to induce a flat line in the profile bundle of an ISMM process was reduced if we used the method of antithetic random variates by splitting the number of replications in half and re-using the same random numbers in the second half of the replications, but with their sign reversed (see Hendry, 1984; Davidson and MacKinnon, 1993). Once the minimum number of replications has been found such that the profile bundle converges to a single flat line for each process, then the number of steps taken for the profile bundle for all of the processes to 'flatline' is the minimum number of steps ahead that should be considered in constructing the generalized Beveridge-Nelson decomposition, and the number of replications required to produce this result is also the minimum number of replications that should be used in the decomposition. In the applied examples reported below, it was found that a relatively small total number of replications – around 2,000 – was adequate in this respect. If convergence to a single flat line does not appear to be possible for any number of replications or steps ahead, then the process under consideration is probably not ISMM.

Finally, before proceeding to a discussion of our univariate and multivariate applied examples, it is worth noting that the Beveridge-Nelson (1981) and Stock-Watson (1987) result that the permanent component of a linear I(1) process can be reduced to a linear random walk process also goes through for our generalization to the nonlinear case. To see this, begin by noting from (4) that $Z^{\pi}(t+1)$ can be written as

$$Z^{p}(t+1) = Z(t) + \lim_{k\to\infty}\sum_{j=2}^{k}\left(E\left[(1-L)Z(t+j)\middle|\Omega(t+1)\right]-\mu\right)+(1-L)Z_{t+1}\,. \tag{7}$$

Subtracting (4) from (7), we have

$$Z^{p}(t+1) - Z^{p}(t) = \lim_{k\to\infty}\sum_{j=1}^{k}\left(E\left[(1-L)Z(t+j)\middle|\Omega(t+1)\right]-E\left[(1-L)Z(t+j)\middle|\Omega(t)\right]\right)+\mu\,. \tag{8}$$

The innovation on the right side of (8) is the revision in expectations about the infinite sum of future first differences in the level of the Z series, plus the steady-state drift component. The revision in the expectation from period t to period t+1 as new information arrives in period t+1 must be orthogonal to information available at time t, i.e. Ω(t), and follows immediately from taking the expectation of first term on the right-hand side of (8) conditional on Ω(t) and applying the law of iterated expectations. Hence, (8) may be rewritten :

$$Z^p(t+1) = \mu + Z^p(t) + z(t+1), \qquad E\left[z(t+1) \mid \Omega(t)\right] = 0. \tag{9}$$

Thus, Beveridge and Nelson's (1981) original result that the permanent component is a random walk with drift, although derived originally for a univariate process with a linear ARIMA representation, and generalized to the case of multivariate I(1) processes by Stock and Watson (1987), continues to hold for nonlinear univariate and multivariate processes which are ISMM.

**3. A univariate application : nonlinearities in the US business cycle**

Beaudry and Koop (1993) analyze a nonlinear univariate model of US quarterly real GNP of the form :

$$\Phi(L)\Delta Y(t) = a + \{\Gamma(L) - 1\}CDR(t) + \Theta(L)e(t)\,, \tag{10}$$

where $\Phi(L), \Gamma(L)$ and $\Theta(L)$ are poynomials in the lag operator L of orders p, q and r respectively, $\Gamma(0) = 1$ and ε(t) is a white-noise process. Y(t) is the logarithm of real GNP, CDR(t) is a variable designed to measure the current depth of recession and is defined as the gap between the current level of output and the economy's historical maximum level :

$$CDR(t) \equiv \max\{Y(t-j)\}_{j\geq 0} - Y(t)\,. \tag{11}$$

Using data for the period 1947 :1-1989 :4 from the Citibase data tape, Beaudry and Koop (1993) find that this model fits the data well, with p=2, q= 0 and r=1 being their preferred specification.

The essence of the Beaudry-Koop analysis is that, because of the presence of significant capacity constraints during an expansionary phase and, conversely, excess capacity during a contractionary phase of the business cycle, real output shocks will impact asymmetrically according to whether the economy is in a contractionary phase or not and also to the depth of the current recession. In particular, when the current depth of recession is deep (and if, as Beaudry and Koop find, r=1), then positive innovations will tend to have a more persistent effect on output than negative innovations for $\Gamma(1) - 1 > 0$ , since a negative innovation will be at least partly compensated for by the CDR variable.

The Beaudry-Koop model of business cycle asymmetry has been criticized, for example by Elwood (1998), who fits asymmetric (threshold-disturbance) moving average and autoregressive models to real output growth and is unable to find strong evidence of asymmetry according to whether the innovation is negative or non-negative. In fact, Elwood finds that a standard linear MA(2) model provides the best fit for changes in real US GNP.

One way of comparing the two models is through their identification of business cycles and, as is common in this literature (e.g. Beveridge and Nelson, 1981), by comparing the business cycle turning points identified by the models with the National Bureau of Economic Research (NBER) business cycle reference points. To do this, however, requires the permanent and temporary or cyclical components of the real GNP series to be extracted using the estimated models. Since the Beaudry-Koop model is nonlinear, this is a natural application of our extension of the Beveridge-Nelson decomposition.

Re-estimating the Beaudry-Koop equation, using a slightly longer data set for quarterly real US GNP, for the period 1947 :1-1998 :2 (also from the Citibase data tape) we obtained :

$$\underset{\phantom{(0.001)}}{\Delta Y(t) =} \underset{(0.001)}{0.002} + \underset{(0.078)}{0.433}\ \Delta Y(t-1) + \underset{(0.079)}{0.182}\ \Delta Y(t-2) + \underset{(0.111)}{0.337}\ CDR(t-1), \qquad (12)$$

$R^2$=0.47, s=0.00985

(where $R^2$ =is the coefficient of determination, s is the estimated standard error of the regression residuals and figures in parentheses denote estimated standard errors), which closely replicates Beaudry and Koop's results as reported in their Table 1.

We then checked that the log real GNP series was generated by an integrated short memory in mean process by plotting the profile bundle for a range of initial values for $\Delta Y$ and CDR. The range of $\Delta Y$ over the sample period was between approximately –0.03 and +0.04 and the mean and median were both 0.008, while the range for CDR was between zero and +0.04 with a mean and median of 0.004 and 0.0 respectively. In order to examine initial conditions which were both 'typical' and 'extreme' for each of these, we considered all possible combinations of starting values for $\Delta Y$ of ±0.1, ±0.05, and 0.0 and starting values for CDR of 0.1, and 0.0 – a total of ten combinations (experiments with other starting values yielded qualitatively identical results).[8] As noted above, we also used the method of antithetic random variates by splitting the number of replications in half and re-using the same random numbers in the second half of the replications, but with their sign reversed (Hendry, 1984; Davidson and MacKinnon, 1993).

The resulting profile bundle is shown in Figure 1 for a forecast horizon of fifteen quarters. The fact that the profile bundle flatlines and each of the conditional forecasts of the change in real GNP settles down to the same value after about fourteen quarters is a strong indication that the process is ISMM. This profile bundle was constructed using 1,000 sets of antithetic random numbers, or a total of 2,000 replications.

We then used our estimated equation to effect the nonlinear Beveridge-Nelson decomposition by Monte Carlo integration. For comparison, we also computed the linear Beveridge-Nelson decompositions for linear AR(2) and MA(2) models, since the first of these is suggested by Beaudry and Koop as the best alternative to their nonlinear model, while an MA(2) specification is suggested as superior by Elwood (1998).

For the nonlinear model (16), we constructed the conditional forecasts necessary for the decomposition through Monte Carlo replications based on 1,000 sets of antithetic random numbers (2,000 replications in all) for each data point in the sample, and with the long horizon set at twenty, since the construction of the profile bundle suggested that these settings would be adequate. The

Beveridge-Nelson decompositions for the AR(1) and MA(2) models were carried out in the standard manner (Beveridge and Nelson, 1981).

All three cyclical components are shown in Figure 2, which also shows the NBER business cycle reference points. While these results are presented for purely illustrative purposes and it is not our purpose here to champion any one model, it is nevertheless interesting to note that the cyclical component extracted using the nonlinear Beaudry-Koop model appears to concur with the NBER reference points in the sense that the NBER points in nearly every case correctly coincide with a peak and a trough of the cyclical component, while no such pattern is apparent for the AR(2) or MA(2) Beveridge-Nelson cyclical components.

## 4. A multivariate application: nonlinearities in the term structure of interest rates

Enders and Granger (1998) estimate nonlinearities in the term structure of interest rates through an analysis of the interest differential, RD, between the quarterly federal funds rate, RS, and the interest rate on ten-year US government securities, RL, over the period 1958 :1-1994 :1, (RD=RS-RL) They start by looking for evidence of nonlinear cointegration between RS and RL by estimating a momentum threshold autoregressive (M-TAR) equation of the form

$$\Delta \mathrm{RD(t)} = a + r_1 \mathrm{I(t)[RD(t-1)} + t\,] + r_2[1 - \mathrm{I(t)][RD(t-1)} + t\,] + e\mathrm{(t)}, \tag{13}$$

where ε(t) is a white noise disturbance, τ is the size of the threshold, representing the equilibrium excess of long-term interest rates over short-term interest rates due to term premiums, and I(t) is the momentum Heaviside indicator defined as

$$\mathrm{I(t)} = \begin{cases} 1 \text{ if } \Delta \mathrm{RD(t-1)} \geq 0 \\ 0 \text{ if } \Delta \mathrm{RD(t-1)} < 0 \end{cases} . \tag{14}$$

Using the same data source as Enders and Granger (the International Monetary Fund’s *International Financial Statistics* CD-ROM) but extending the sample to 1958 :1-1999 :3, we re-estimated (13), applying a least squares grid search method to locate the appropriate threshold (Chan, 1993 ; Hansen, 1997). This yielded (with estimated standard errors in parentheses):

$$\Delta RD(t) = \underset{(0.092)}{0.275} - \underset{(0.041)}{0.064}\ I(t)[RD(t-1) + 2.65] - \underset{(0.051)}{0.290}\ [1 - I(t)][RD(t-1) + 2.65], \qquad (15)$$

which very closely replicates equation (11) in Enders and Granger (1998) both in terms of the estimated coefficients and in terms of the estimated threshold (Enders and Granger estimate a threshold of 2.64 percent compared to our estimate of 2.65 percent.) The F-statistic for the null hypothesis $H_0 : \rho_1 = \rho_2 = 0$ was 16.18, strongly exceeding the 1% critical value of around 7 calculated by Monte Carlo methods by Enders and Granger under the null hypothesis of a random walk. This therefore implies nonlinear cointegration between RS and RL. By a nonlinear form of the Granger Representation Theorem (Balke and Fomby, 1997 ; Granger, 1999), Enders and Granger then infer the existence of a nonlinear vector error correction representation for RS and RL which they proceed to estimate.

Define the asymmetric error correction terms:

$$Z^{+}(t-1) = I(t)[RS(t-1) - RL(t-1) + 2.65], \qquad (16)$$

$$Z^{-}(t-1) = [1 - I(t-1)][RS(t-1) - RL(t-1) + 2.65]. \qquad (17)$$

Given the definition of the Heaviside indicator function (14), the term defined in (16) will be zero when the last period change in the interest differential was negative and will otherwise be equal to the distance of the interest differential from the threshold of 2.65 percent. Similarly, the error correction term defined in (17) will be zero when the most recent change in the interest differential was positive and will otherwise be equal to the distance of the interest differential from the threshold. Enders and Granger estimate a nonlinear vector error correction model (NLVECM) for $\Delta RS$ and $\Delta RL$ consisting of current and lagged values of $\Delta RS$ and $\Delta RL$ and one lag of each of the asymmetric error correction terms $Z^+$ and $Z^-$. While we were able closely to replicate Enders and Granger's estimated asymmetric NLVECM, because we wished to use the estimated model to effect a nonlinear Beveridge-Nelson decomposition through multi-step forecasts, it was important that we obtained as efficient an estimate as possible. We therefore discarded terms with estimated coefficients that were insignificantly different from zero at the five percent level and estimated the NLVECM by multivariate least squares in order to gain efficiency. The resulting estimates were as follows:

$$\Delta RS(t) = \underset{(0.143)}{0.694}\ \Delta RL(t-1) - \underset{(0.119)}{0.427}\ \Delta RL(t-2) - \underset{(0.047)}{0.219}\ Z^{-}(t-1) \tag{18}$$

$$\Delta RL(t) = \underset{(0.077)}{0.227}\ \Delta RL(t-1) + \underset{(0.020)}{0.046}\ Z^{+}(t-1) \tag{19}$$

$$\hat{\Sigma} = \begin{bmatrix} 0.241 & 0.263 \\ 0.263 & 0.824 \end{bmatrix} \tag{20}$$

where (20) gives the estimated covariance matrix of the residuals and figures in parentheses are estimated standard errors.

We then checked that the processes, as modeled in (18)-(20), were ISMM by computing the profile bundles through Monte Carlo simulation of the estimated system, assuming a bivariate gaussian distribution for the error terms with a covariance matrix as given in (20). As in the univariate case discussed above, we used antithetic random variates. The profile bundle for the interest rate model, constructed using 1,000 sets of antithetic random variables or 2,000 replications in all, is shown in Figure 3, which shows that the short rate profile bundle flatlines after about thirty-four months, while the long rate profile bundle flatlines a little sooner, at about thirty months.

We then proceeded to carry out the nonlinear Beveridge-Nelson decomposition using Monte Carlo integration and assuming a bivariate gaussian distribution for the error terms with a covariance matrix as given in (24). We used 1,000 sets of antithetic random variables (2,000 Monte Carlo replications) for each data point and a long horizon of forty steps ahead, as the constructed the profile bundle indicted that this would be adequate.

The resulting permanent and temporary components of the short-term and long-term interest rates are given in Figure 4. The temporary component of the short rate is quite volatile, with a standard deviation of 1.40%. As one might perhaps expect, the temporary component of the long rate is much smaller, with a standard deviation of 0.34%.

## 5. Conclusion

In this paper, we have presented and implemented a tractable method for decomposing a class of *nonlinear* time series into the sum of permanent and temporary components. Following the conceptual definition of the permanent component of a time series with a univariate ARIMA representation introduced in Beveridge-Nelson (1981) and extended to the case of multivariate I(1) processes in Stock and Watson (1987), we defined the permanent component of a nonlinear ISMM time series as the as the long-horizon forecast of the *level* of that time series. In the nonlinear case, generating long-horizon forecasts is non-trivial and we suggested a tractable method of doing this using Monte Carlo integration. We also demonstrated that Beveridge and Nelson's (1981) original result that the permanent component is a random walk with drift, although derived originally for a univariate process with a linear ARIMA representation, continues to hold for univariate and multivariate nonlinear models which are ISMM. We applied our method to a univariate example of a nonlinear model of US output fluctuations and to a multivariate example involving the term structure of interest rates, and in both cases it was found to yield sensible results.[9]

**References**


Balke, Nathan S., and Thomas B. Fomby (1997), ‘Threshold Cointegration’, *International Economic Review*; 38(3), pp. 627-45.

Beaudry, Paul and Gary Koop (1993), ‘Do Recessions Permanently Change Output ?’ *Journal of Monetary Economics*, 31(2), pp. 149-164.

Beveridge, Stephen and Charles R. Nelson (1981), ‘A New Approach 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(2), pp. 151-74.

Blanchard, Olivier J. and Quah, Danny (1989), ‘The Dynamic Effects of Aggregate Supply and Demand Disturbances’, *American Economic Review*, 79, pp. 655-73

Chan, K.S. (1993), ‘Consistency and Limiting Distributions of the Least Squares Estimator of a Threshold Autoregressive Model’, *The Annals of Statistics*, 21, pp. 520-33.

Clarida, Richard H. and Jordi Gali (1994), ‘Sources of Real Exchange-Rate Fluctuations: How Important Are Nominal Shocks?’, *Carnegie-Rochester Conference Series on Public Policy*, 41, pp. 1-56.

Clark, Peter K. (1987) ‘The Cyclical Component of U.S. Economic Activity’, *Quarterly Journal of Economics*; 102(4), pp. 797-814.

Cochrane, John H. (1994), ‘Permanent and Transitory Components of GNP and Stock Prices’, *Quarterly Journal of Economics*; 109(1), pp. 241-65.

Davidson, Russell and James G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford and New York : Oxford University Press.

Elwood, S. Kirk (1998), ‘Is the Persistence of Shocks to Output Asymmetric?’ *Journal of Monetary Economics*, 41(2), pp. 411-26.

Enders, Walter, and Clive W.J. and Granger (1998), ‘Unit Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates’, *Journal of Business and Economic Statistics*, 16(3), pp. 304-11.

Gallant, A.Ronald, Peter E. Rossi and George Tauchen (1993), ‘Nonlinear Dynamic Structures’, *Econometrica*, 61(4), July 1993, pp.871-907.

Granger, Clive W.J. (1999), ‘An Overview of Nonlinear Macroeconometric Empirical Models’, Working Paper, Department of Economics, University of California, San Diego.

Granger, Clive W.J., and Timo Teräsvirta (1993), *Modeling Nonlinear Economic Relationships*, Oxford : Oxford University Press.

Hansen, Bruce E. (1997), ‘Inference in TAR Models,’ *Studies in Nonlinear Dynamics and Econometrics*, 21(1), pp. 1-14

Harvey, Andrew C. (1981), *Time Series Models*, Philip Allen, Hemel Hempstead.

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

Hendry, David F. (1984), 'Monte Carlo Experimentation in Econometrics,' chapter 16 in Zvi Griliches and Michael D. Intrilligator (eds.) *Handbook of Econometrics*, Volume II, Amsterdam: North-Holland.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson (1991), 'Stochastic Trends and Economic Fluctuations', *American Economic Review*, 81, pp. 819-40.

Robinson, Peter M. (1983), 'Nonparametric Estimators for Time Series', *Journal of Time Series Analysis*, 4, pp. 185-207.

Stock, James H., and Mark W. Watson (1987), 'Testing for Common Trends', *Journal of the American Statistical Association*, 83(404), pp. .

Watson, Mark W. (1986) 'Univariate Detrending Methods with Stochastic Trends', *Journal of Monetary Economics*, 18(1), pp. 49-75.

## Notes

[1] We are grateful to Clive Granger for comments on a previous version of this paper, although the usual disclaimer applies. Clarida is grateful for funding from the Columbia University Research Fund.

[2] Another popular method of decomposing a time series with a linear ARIMA representation into permanent and temporary components is the orthogonal unobservable components approach presented in Watson (1986), Clark (1987) and Harvey (1981, 1989). As with the Beveridge-Nelson approach, this method also yields a decomposition, which is the sum of a martingale and a temporary component with an ARMA representation. The innovations in these components are uncorrelated by construction. The unobservable components themselves must be estimated using linear state space methods that take advantage of the Kalman filter. The Beveridge-Nelson approach appears to be more tractable for non-linear applications because the unobservable components can be recovered directly from the long-horizon forecast of the process itself. A popular method for decomposing multivariate series is through long-run restrictions on the coefficients and covariance matrix of the multivariate Wold representation (e.g. Blanchard and Quah, 1989 ; King, Plosser, Stock and Watson, 1991). This method generally requires the imposition of economic theory-specific restrictions, however. A major advantage of the Beveridge-Nelson decomposition over these alternative methods is that the permanent component has an immediately intuitive appeal as the long-horizon prediction.

[3] See e.g. Clarida and Gali (1994) for an application.

[4] Our notion of a process being integrated short memory in mean is closely related to the notions of 'integrated in mean' and 'not integrated in mean', employed by Bollerslev and Engle (1993) and the definitions of long memory in mean and short memory in mean of Granger and Teräsvirta (1993). An important feature of our definition, however, is that the long-horizon prediction of the first difference process is a vector of constants, thereby ruling out, for example, limit cycles and chaotic behavior.

[5] For the one-step-ahead predictor for a model with an additive disturbance, the Monte Carlo forecast is in fact identical to the true conditional expectation:

$$\exists\left[(1-L)Z(t+1)\middle|\Omega(t)\right] = E\left[(1-L)Z(t+1)\middle|\Omega(t)\right] = \Theta\left[(1-L)Z(t),(1-L)Z(t-1),...\right].$$

Equation (6) will still hold for k=1, however, although it is not necessary the use Monte Carlo integration for one step ahead.

[6] For large M, moreover, the Monte Carlo forecast will also be independent of the particular draws of random numbers. We have therefore omitted $e_m$ as a conditioning variable on the left-hand side of (6).

[7] A variant would be to construct the forecasts using bootstrapping rather than Monte Carlo methods.

[8] In the research reported in both this and the next section, we assumed that the model disturbances were gaussian for the purposes of the Monte Carlo simulations. This appeared to be a reasonable assumption, however, since redoing the work using bootstrap methods on the fitted residuals generated qualitatively virtually identical results.

[9] It should be borne in mind that these applications were chosen for purely illustrative purposes.