WORKING PAPER
No. 10 February 1992

FORECASTING THE BUSINESS CYCLE NOT USING MINIMUM AUTOCORRELATION FACTORS

Karl-Gustaf Löfgren
Bo Ranneby
Sara Sjöstedt

Svensk resumé på sid 26.

UNIVERSITY OF UMEÅ
Department of Economics
Department of Mathematical Statistics
October 1991

ABSTRACT: FORECASTING THE BUSINESS CYCLE NOT USING MINIMUM AUTOCORRELATION FACTORS

by Karl–Gustaf Löfgren, Bo Ranneby, and Sara Sjöstedt

We introduce a forecasting technique based on multivariate ideas previously applied in remote sensing. The approach has the trivial, but nonetheless fundamental, purpose of splitting the information inherent in the time series into important and unimportant information. The important information is used for forecasting purposes, while the unimportant is thrown away. Although related to vector autoregression, giving asymptotically the same estimates, there are reasons to believe that the approach gives better precision of parameter estimates for finite samples, as well as more precise predictions.

UNIVERSITY OF UMEÅ
Department of Economics and
Department of Mathematical Statistics
January 1989
Revised October 1991

FORECASTING THE BUSINESS CYCLE NOT USING MINIMUM AUTO-CORRELATION FACTORS*

by

Karl–Gustaf Löfgren, Bo Ranneby and Sara Sjöstedt

1 INTRODUCTION

The view and status of business cycle analysis within the economic profession has changed considerably over time. The interest in cycles has, as Gordon (1985) points out, been almost cyclic. After the First World War, interest peaks in the 1930–40s and 1980s and has troughs in the 1920s and 1960s.

The predominant approach to dealing with business cycles among theoreticians and applied econometricians over the last 20 years or so, dates back to Frisch (1933) and Slutsky (1927), and rests on the assumption that economic fluctuations originate essentially from sources that are exogenous to the private sector of the economy. Business cycles are viewed as caused by an exogenously given, stationary stochastic process of economy–wide shocks. Modern proponents of the so called real business cycle theory include Kydland and Prescott (1982), Long and Plosser (1983), and King and Plosser (1984).

However, some earlier students of the business cycle did indeed try to identify internal mechanisms that can be responsible for observed variations in prices and quantities. Work belonging to this tradition includes Goodwin (1951), Harrod (1936), Hicks (1950), Kaldor (1940), and Samuelson (1939) just to mention a few. Recent attempts are found in, for example, Grandmont (1985) and Puu (1988) and differ from the earlier ones in that they include the study of the occurrence of

* The authors acknowledge helpful comments from Peter Englund and Anders Vredin at The Swedish Council of Economic Advisors, Stockholm, and Lars–Erik Öller, Department of Economics, University of Helsinki. Research assistance has been provided by Eva Rovainen, Department of Biometry, Swedish University of Agricultural Sciences, Umeå.

complex deterministic dynamics based on the mathematical theory of non-linear dynamical systems.

The present paper has little to add to the ongoing methodological controversy between these two schools, but belongs to a recent philosophy of forecasting the business cycle. In the late 40s and early 50s the technical development of econometrics was proceeding rapidly and there existed considerable optimism with respect to the possibilities of building workable large scale econometric models for forecasting purposes. This optimism depreciated considerably during the 70s, and disappeared due to the "Lucas critique" and Sims' critique[1] of the exclusion restrictions assumed in the specification of structural equations of large-scale models. The latter also introduced the small-scale reduced-form vector autoregression (VAR) models distinguished by their symmetry in treating all variables of interest as endogenous.

The present approach has the trivial, but nonetheless fundamental, purpose of splitting the information inherent in the time series into important and unimportant information.[2] The important information is used for forecasting purposes, while the unimportant ("noise") is thrown away. Although related to VAR, the approach stems from statistical multivariate ideas, which have been applied in remote sensing. Collected spectral signals are "aggregated" into a combined signal by weighting them in a manner such that the spatial autocorrelation[3] in the combined signal is maximized. In a time series setting, the analog is to maximize the autocorrelation over time. We can continue and construct more signals, also linear combinations of the spectral signals, obtained by maximizing the autocorrelation given restrictions on orthogonality with higher order combined signals. The new signals are reminiscent of principal components (PC), but a key difference is that, whereas in principle component analysis the weights are chosen to maximize the variance of the combined signal, here they are chosen to maximize the autocorrelation. In a time series setting, a linear combination of the original time series is constructed by maximizing its autocorrelation. It should contain important

---

1 See Lucas (1976) and Sims (1980).

2 It is closely related to ideas in a paper by Box and Tiao (1977), where a canonical analysis of multiple time series is conducted. The components of the transformed autoregressive process are ordered from least to most predictable. Similar ideas are also found in Sims (1981) where an autoregressive index model is discussed.

3 Under a restriction that the length of the vector of weights equals one.

information about the original time series and be valuable for forecasting purposes. Business cycle phenomena have been defined as "the recurrent fluctuations of output about trend and the co–movements among other aggregate time series[4]", and it would be strange if the joint signals that constitute the maximum correlation over time did not contain information about these co–movements.

The rest of the paper is structured as follows: in Section 3 we work out the statistical theory behind the MAFs, which is a special case of a more general analysis found in Switzer and Green (1984), Switzer (1985), and Conradsen, Kjær Nielsen, and Thyrsted (1986). In Section 4 we examine some important practical problems such as how to detrend the original series to create stationarity. It should be clear that the implicit assumptions about the growth process, which are made in any detrending or other transformations, are not necessarily innocuous. These practical problems are discussed in Section 5 in connection with a sample of quarterly data containing GNP in Sweden and its main components. Next we estimate the autocorrelation factors for this data set and the relevant regression equations to be used in the forecasts. Finally, in Section 6 we briefly compare our forecasting technique with the more widely used VAR–method; an exercise which generates questions for future research.

2 A STYLIZED EXAMPLE

To illustrate our ideas, we have borrowed a stylized macroeconomic model from Stock and Watson (1988). They use the model to illustrate the inappropriateness of inference methods based on stationarity when performing econometric analysis with cointegrated variables. We will use the model to illustrate some general aspects of forecasting using the suggested technique.

---

[4] See Prescott (1986). The classical definition is that of Burns and Mitchell (1946):

Business cycles are a type of fluctuation found in the aggregate activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occuring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own.

In this artificial economy, disposable income consists of two parts: a permanent and a transitory. The permanent component of disposable income is assumed to follow a random walk, while the transitory component is an independently and identically distributed random variable which is independent of the permanent component. Consumers are assumed to know their permanent income, and to be consuming precisely the permanent component of their disposable income.

Price changes are assumed to be random and unforecastable with mean zero. They do not confuse consumers in the sense that real consumption and disposable income are determined independently of the price level or its changes. The model has the following shape:

$$\begin{aligned} &\text{(i)} \quad && y_t = y^p_t + y^s_t \\ &\text{(ii)} \quad && y^p_t = y^p_{t-1} + u_t \\ &\text{(iii)} \quad && c_t = y^p_t \\ &\text{(iv)} \quad && p_t = p_{t-1} + v_t \end{aligned} \tag{1}$$

where y^p_t and y^s_t respectively denote the permanent and transitory components of disposable income (y_t). The innovations y^s_t, u_t, and v_t are assumed to be mutually independent and normally distributed with mean zero and unit variance. Consumption and the price level are denoted c_t and p_t, respectively.

The observables are y_t, c_t, and p_t, and the important information for forecasting purposes is obviously c_t and p_t, while $y_t - c_t = y^s_t$ is white noise and worthless for forecasting purposes.

We use the observables to create three new variables $z_1(t)$, $z_2(t)$, and $z_3(t)$ as linear combinations of the observables:

$$z_i(t) = a_{1i}y_t + a_{2i}c_t + a_{3i}p_t = (a_{1i}+a_{2i})c_t + a_{3i}p_t + a_{1i}y^s_t$$

$$i = 1, 2, 3 \tag{2}$$

and solve the maximization problems:

$$\underset{a_i \in M_i}{\mathrm{Max}} \ \mathrm{corr}(z_i(t), z_i(t-1)) \qquad i = 1, 2, 3 \tag{3}$$

where $M_i = \{a_i \mid \mathrm{corr}(z_i(t), z_j(t)) = 0, j < i\}$

and $a_i = [a_{1i}, a_{2i}, a_{3i}]$

In appendix A we show that:

$$z_1(t) = c_t$$

$$z_2(t) = p_t \tag{4}$$

$$z_3(t) = \frac{1}{\sqrt{2}}(y_t - c_t) = \frac{1}{\sqrt{2}} y_t^s$$

Following the terminology in Switzer and Green (1984), we will call $z_1(t)$, $z_2(t)$ and $z_3(t)$ the Min/Max Autocorrelation Factors (MAF). Four preliminary statements are crucial. Firstly, the important information is recovered in the first two MAFs $z_1(t)$ and $z_2(t)$. In the forecasting equation

$$y_t = \alpha_0 + \alpha_1 z_t(t-1) + \alpha_2 z_2(t-1)$$

we would expect to get the asymptotic estimates $\alpha_0 = E[y_t^s] = 0$, $\alpha_1 = 1$, and $\alpha_2 = 0$. Secondly, we will in general need more than one MAF for forecasting purposes. This fact is highlighted in the present context, since c_t and p_t are orthogonal. Thirdly, the MAF that minimizes the autocorrelation is white noise, and indeed worthless for forecasting purposes, since $\mathrm{corr}(z_3(t), z_3(t-1)) = 0$. Finally, although the technique to be introduced below requires that the time series are stationary, the idea is general. However, when the time series are non–stationary difficulties may arise, especially in the estimation and maximization of the correlations.

3 THE MIN/MAX AUTOCORRELATION FACTORS

We now turn to the derivation of the autocorrelation factors. As was pointed out in the introduction, the MAF–method has been used by Switzer and Green (1984), Switzer (1985), and Switzer and Ingebritsen (1986) in a remote sensing context. They transform the spectral signals of a satellite picture linearly to a new set of signals by maximizing the spatial correlation between information from adjacent pixels. These signals are used in attempts to isolate the "fundamental information" from noise.

Starting from a stationary multidimensional stochastic process:

$$x'(t) = [x_1(t),\ldots,x_p(t)] \tag{5}$$

a set of new variables is formed through linear combinations of the components of the original process, $z_i(t) = \gamma_i'x(t)$, $i = 1,\ldots,p$. The new variables (the MAFs) have the property that they are orthogonal and that the spatial correlations decrease with the index i. In other words, the spatial correlation is at a maximum for $z_1(t) = \gamma_1'x(t)$, and $z_1(t)$ is thus a variable with important information together with other high correlation MAFs, while the spatial correlation is at a minimum for $z_p(t) = \gamma_p'x(t)$, and $z_p(t)$ therefore represents "noise".

The derivations of the MAFs can be found e.g. in Switzer (1985) or Conradsen et al. (1986).

Since we are interested in forecasting time series where temporal, not spatial correlation is maximized, we will present a slightly adjusted derivation of the MAFs. We start by introducing a formal definition of the Min/Max Autocorrelation Factors. Let $x(t)$ be a weakly stationary p–dimensional stochastic process. For γ in R^p, and $\Delta > 0$ define:

$$\rho(\gamma, \Delta) = \mathrm{corr}(\gamma'x(0), \gamma'x(\Delta)) \tag{6}$$

For each fixed Δ, we can maximize (or minimize) ρ over γ in subspaces of R^p. In particular, given Δ, we can find vectors $\gamma_1,\ldots, \gamma_p$ where for each i, γ_i is chosen so that the correlation between $\gamma_i'x(t)$ and $\gamma_i'x(t+\Delta)$ is maximized under the restriction $\mathrm{corr}(\gamma_i'x(t),\ \gamma_k'x(t)) = 0$ for all $k < i$. The linear combinations $\gamma_i'x(t)$, $i = 1,\dots,p$ are the Min/Max Autocorrelation Factors.

Now let:

$$\mathrm{Var}(x(t)) = \Sigma_0$$

and (7)

$$\mathrm{Var}(x(t+\Delta) - x(t)) = \Sigma_\Delta.$$

From

$$\mathrm{Var}[\gamma'x(t+\Delta) - \gamma'x(t)] = \gamma'\Sigma_\Delta\gamma = 2\gamma'\Sigma_0\gamma - 2\mathrm{cov}(\gamma'x(t),\ \gamma'x(t+\Delta)) \tag{8}$$

follows that:

$$\mathrm{corr}(\gamma'x(t),\ \gamma'x(t+\Delta)) = 1 - \frac{1}{2}\frac{\gamma'\Sigma_\Delta\gamma}{\gamma'\Sigma_0\gamma}. \tag{9}$$

Maximizing (9) is equivalent to minimizing

$$R(\gamma) = \frac{\gamma'\Sigma_\Delta\gamma}{\gamma'\Sigma_0\gamma}. \tag{10}$$

However, $R(\gamma)$ is at its minimum if γ is chosen equal to the eigenvector corresponding to the smallest eigenvalue of Σ_Δ with respect to Σ_0, see e.g. Rao (1973, p. 74).

Let $\lambda_1 \leq \dots \leq \lambda_p$ be the eigenvalues, and $\gamma_1,\dots,\gamma_p$ the corresponding normed eigenvectors of Σ_Δ with respect to Σ_0, i.e. λ_i and γ_i satisfy:

$$\Sigma_\Delta\gamma_i = \lambda_i\Sigma_0\gamma_i \qquad i = 1,\dots,p \tag{11}$$

Now put:

$$z_i(t) = \gamma_i' x(t) \qquad i = 1,\ldots,p \tag{12}$$

Then $z(t) = [z_1(t),\ldots,z_p(t)]$ are the MAFs.

Theorem 1: The MAFs $z(t)$ have the following properties:

(i) $\quad \mathrm{corr}(z_i(t),\, z_j(t)) = 0 \qquad i \neq j$

(ii) $\quad \mathrm{corr}(z_i(t),\, z_i(t+\Delta)) = 1 - \frac{1}{2}\lambda_i$

(iii) $\quad \mathrm{corr}(z_1(t),\, z_1(t+\Delta)) = \sup\limits_{\gamma} \mathrm{corr}(\gamma' x(t),\, \gamma' x(t+\Delta))$

$$\mathrm{corr}(z_p(t),\, z_p(t+\Delta)) = \inf_{\gamma} \mathrm{corr}(\gamma' x(t),\, \gamma' x(t+\Delta))$$

$$\mathrm{corr}(z_i(t),\, z_i(t+\Delta)) = \sup_{\gamma \in M_i} \mathrm{corr}(\gamma' x(t),\, \gamma' x(t+\Delta))$$

where $M_i = \{\gamma \,|\, \mathrm{corr}(\gamma' x(t),\, z_j(t)) = 0,\ j < i\}$

Proof: (i) The eigenvectors γ_i and γ_j, $i \neq j$ satisfy[5]:

$$\gamma_i' \Sigma_0 \gamma_j = \gamma_i' \Sigma_\Delta \gamma_j = 0$$

Thus,

$$\mathrm{cov}(\gamma_i' x(t),\, \gamma_j' x(t)) = \gamma_i' \Sigma_0 \gamma_j = 0$$

(ii) and (iii) are direct consequences of (9) and (11).

---

5 See Rao (1973) p 74.

It is also easy to prove:

THEOREM 2: The MAFs $z(t) = [z_1(t),...,z_p(t)]$ are invariant with respect to linear transformations of the original time series.

PROOF: Let:

$$w(t) = Tx(t) \tag{13}$$

where T is a nonsingular square matrix. Now:

$$Var(w(t)) = T\, \Sigma_0\, T' \tag{14}$$

and

$$Var(w(t+\Delta) - w(t)) = T\, \Sigma_\Delta\, T' \tag{15}$$

Since the eigenvalues of $T\, \Sigma_\Delta\, T'$ with respect to $T\, \Sigma_0\, T'$ are identical to the corresponding eigenvalues of Σ_Δ with respect to Σ_0 the proof follows.

One way of deriving the MAFs is suggested below.

From Theorem 2 we know that the MAFs are invariant with respect to linear transformations. Hence, to enable us to use standard eigenvalue procedures, it is wise to transform the original time series in such a manner that the covariance matrix equals the identity matrix, i.e.:

1 Find a linear transformation of the original time series $x(t)$, say $w(t) = Tx(t)$, such that $Var[w(t)] = I_{pxp}$. For example $w(t)$ can be the principal components based on the covariance matrix Σ_0, where each principal component is divided by its standard deviation, the square root of the corresponding eigenvalue, to obtain the identity matrix.

2 For a predetermined Δ, form the difference $w(t) - w(t+\Delta)$ and compute $\Sigma^*_\Delta = Var[w(t) - w(t+\Delta)]$.

3 Find the normed eigenvectors $U = (u_1,...,u_p)$ and the eigenvalues $\lambda_1 \leq ,...,$ $\leq \lambda_p$, of Σ^*_Δ.

4 Compute $U'w(t) = U'Tx(t) = z(t)$, the MAFs of $x(t)$.

4 THE FORECASTING EQUATIONS

The ultimate purpose of the above exercise is to use the MAFs for forecasting the components of the original time series. Since they maximize the autocorrelation over time of linear combinations of the original time series, they should contain information about future values of the series. Our forecasting instrument would then naturally be a mapping:

$$x_i(t+\delta) = \alpha_0 + \sum_{i=1}^{k} \alpha_i z_i(t) \qquad i = 1,...,p \quad k < p \tag{16}$$

where it is only worthwhile to include a potent sample of the MAFs, since $z_p(t)$, which minimizes the correlation over time, will very likely be (white) noise.

For

$$z(t) = \Omega x(t), \tag{17}$$

where Ω is a pxp matrix with the MAF–weights as elements, we have that:

$$x(t+\delta) = \Omega^{-1} z(t+\delta) \approx Bz(t) + A \tag{18}$$

where the matrix B and the vector A are chosen to minimize the square of the prediction error.

In practice, it will often be necessary to detrend the $x(t)$ variables to create stationarity before the MAFs are estimated. Moreover, the MAFs at the upper end of the scale will mainly contain noise, and would therefore be superfluous in a forecasting model. Finally, a nonlinear relation between $x(t+\delta)$ and $z(t)$ may also do better than the linear forecasting equation (18). However, below we will only consider linear forecasting equations, which are estimated by ordinary least squares with the higher order MAFs as regressors. It is worth noting that the MAFs in a sense represent "ideal" regressors, since they are mutually orthogonal.

From the discussion above it follows that the original time series have to be stationary. An array of different methods like moving averages, yearly averages, differentiation, and regression analysis tries to accomplish that.

The method applied below is regression analysis in which we decompose the original time series x(t) into two components, one component picking up a deterministic trend and seasonal variation, $\mu(t)$, and one consisting of a stationary time series, e(t):

$$x(t) = \mu(t) + e(t) . \tag{19}$$

To accomplish this, we estimate $\mu(t)$ as a polynomial with dummy variables picking up the seasonal variation. The particular regression model may have the following shape for a linear trend and seasonality of period four:

$$\mu(t) = \beta_0 + \beta_1 t + \sum_{i=2}^{4} \beta_i D_i . \tag{20}$$

When the coefficients have been estimated using OLS, we form the residuals:

$$e(t) = x(t) - \hat{\mu}(t) \tag{21}$$

which should be stationary provided that the residual variance is homoscedastic.

A small simulation experiment performed in Sjöstedt (1988) shows that the method works reasonably well, even if the variances are time dependent, provided that they are of the same magnitude, say $\sigma(t) = \sigma\sqrt{t}$, as in the Stock and Watson example introduced above. If not, there are problems.

The MAF–weights corresponding to the stationary residuals e(t) are estimated using the algorithm introduced above (p. 10), and we form the MAFs $z(t) = \Omega e(t)$. We then run regression equations like those indicated by equation (16):

$$\hat{e}(t+\delta) = \hat{\alpha}_0 + \sum_{i=1}^{k} \hat{\alpha}_i z_i(t) \qquad k < p \tag{22}$$

Our forecasting equation is then given by:

$$\hat{x}(t+\delta) = \hat{\mu}(t+\delta) + \hat{e}(t+\delta) \tag{23}$$

5 AN APPLICATION TO SWEDISH DATA

We will report below estimations of MAFs and forecasting equations on Swedish quarterly GNP data from the first quarter 1970 through the fourth quarter in 1987. The data include Y = GNP, I = gross investments, G = public consumption, C = private consumption, M = imports, and X = exports.[6] All series are deflated into 1980 prices.

These deflated series have been detrended and seasonally adjusted by the regression technique described above. We used a first order polynomial equation over time for the trend and three dummy variables to pick up the season (equation (20)). The results are listed in Table 1:

[6] See appendix B.

Table 1 The trend and season in Swedish quarterly GNP–data.

$$x(t) = \beta_0 + \beta_1 t + \sum_{i=2}^{4} \beta_i D_i$$

	β_0	β_1	β_2	β_3	β_4	$R^2(\mu)$ %
Y	119 288 ***	564 ***	–12 831 ***	–12 247 ***	–24 307 ***	96.8
I	28 996 ***	323 ***	–7 093 ***	–2 908 ***	–5 736 ***	82.3
G	30 047 ***	231 ***	–1 130 ***	–562	–8 939 ***	98.0
C	64 134 ***	203 ***	–7 899 ***	–5 614 ***	–8 544 ***	86.8
M	34 487 ***	218 ***	–3 634 ***	–2 772 **	–4 125 ***	79.1
X	28 183 ***	367 ***	–3 366 ***	1 864 *	–5 165 ***	92.4

*** = significant at the 99.9 percent probability level.[7]
** = significant at the 99.0 percent probability level.
* = significant at the 95.0 percent probability level.

The suggested regression equation fits the data well and there is little residual variation left to explain.

The MAF–weights (the matrix of eigenvectors U'T (p. 10)) corresponding to the residuals of the above equations, as well as the corresponding eigenvalues, for Δ = 1 are listed in Table 2:

7 The residual variance is probably underestimated, since we have serial correlation in the time series. Hence, the significance levels are very likely overestimated.

Table 2 MAF–weights for $\Delta = 1$ derived from the residuals e(t) (Swedish quarterly GNP–data 1970:1 – 1987:4).

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
e_Y	–0.03	0.25	–0.42	0.18	–0.37	–0.01
e_I	–0.24	0.31	–0.52	–0.10	0.78	–0.27
e_G	0.78	–0.74	–0.10	–0.80	0.44	–0.89
e_C	–0.53	–0.41	0.48	0.35	0.11	–0.15
e_M	–0.21	–0.13	0.00	–0.44	–0.18	0.19
e_X	0.09	0.33	0.56	–0.04	0.16	–0.28
λ	0.23	0.36	0.96	1.19	1.54	2.07
$\rho_{\Delta=1}$	0.88	0.82	0.52	0.40	0.23	–0.04

For example, from Table 2 we have the first MAF,

$$z_1(t) = \gamma_1'e(t) = -0.03e_Y(t) - 0.24e_I(t) + \dots + 0.09e_X(t)$$

with $\mathrm{corr}(z_1(t), z_1(t+1)) = 0.88$.

Heavy demand components such as public expenditure and private consumption dominate the MAFs with the greatest correlation over time. A little surprising is the fact that the influences from the export and import residuals are weaker than one would expect in a small open economy like the Swedish one.

Table 3 MAF–weights for $\Delta = 4$ derived from the residuals e(t) (Swedish quarterly GNP–data 1970:1 – 1987:4).

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
e_Y	–0.01	–0.31	–0.04	–0.45	–0.11	–0.07
e_I	0.10	–0.22	–0.46	0.53	–0.41	–0.75
e_G	–0.99	0.61	–0.76	0.65	–0.50	0.25
e_C	–0.11	–0.66	0.35	0.15	0.51	–0.15
e_M	0.05	0.19	0.09	0.01	–0.38	0.56
e_X	0.01	–0.10	–0.29	0.26	0.40	0.17
λ	0.28	0.60	0.90	1.66	2.10	2.37
$\rho_{\Delta=4}$	0.86	0.70	0.55	0.17	–0.05	–0.19

The same weight pattern as at the quarterly level emerges. Weights corresponding to residuals from demand components which constitute a large share of GNP dominate the MAFs with the strongest correlation over time. Imports and exports remain unimportant.

In Figure 1 we graph the MAFs corresponding to the eigenvectors γ_1 (1a), γ_2 (1b), γ_5 (1c), and γ_6 (1d) in Table 2. Worth noting is that, since $-1 \leq \rho \leq 1$, a MAF, other than the lowest ranked, can constitute the noise.

Figure 1 a The first MAF, $z_1(t)$, with $\Delta = 1$ achieved from the residuals of the GNP data.

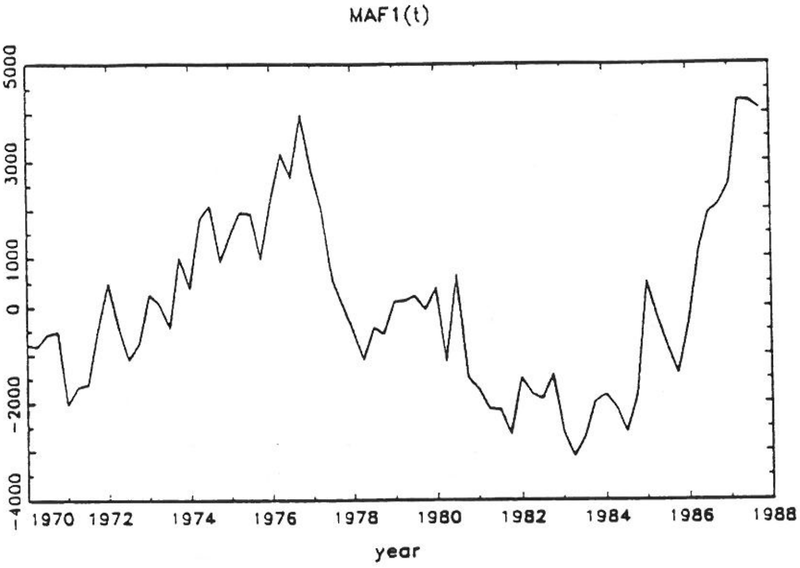


Figure 1 b The second MAF, $z_2(t)$, with $\Delta = 1$ achieved from the residuals of the GNP data.

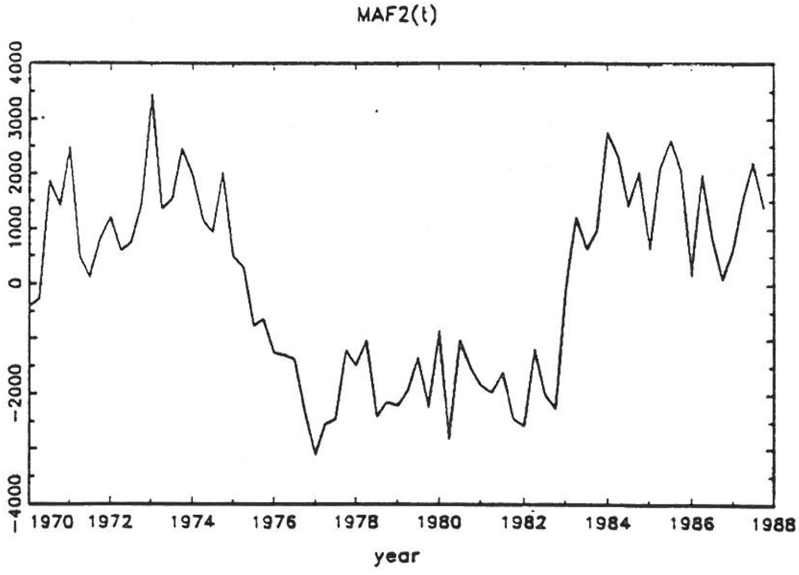


Figure 1 c The fifth MAF, $z_5(t)$, with $\Delta = 1$ achieved from the residuals of the GNP data.

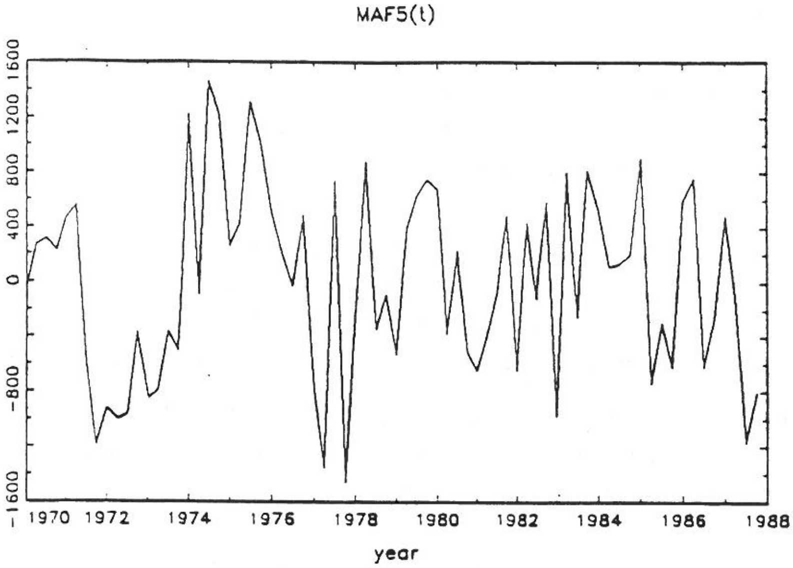
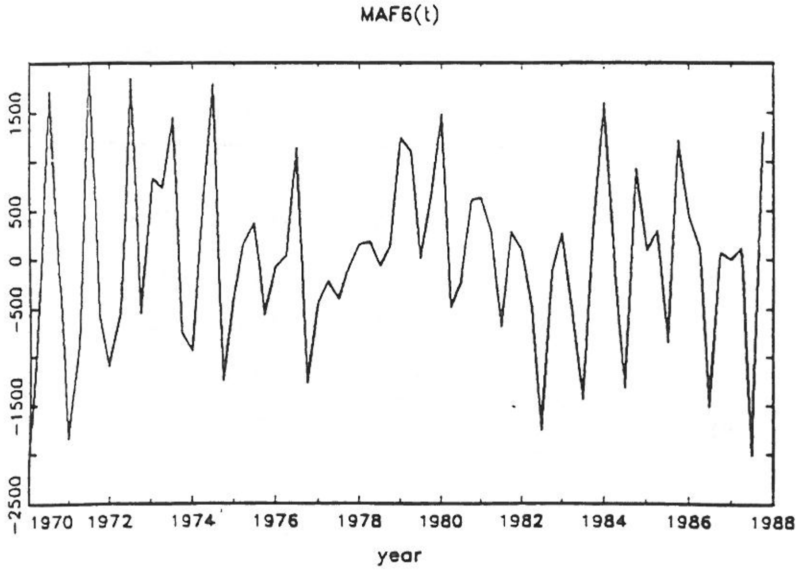


Figure 1 d The sixth MAF, $z_6(t)$, with $\Delta = 1$ achieved from the residuals of the GNP data.



Recall that the MAFs are created by applying the weights to the residuals achieved from removing their respective trends. Thus Figures 1a and 1b might be interpreted as "pictures" of the business cycle.

We now turn to the question of forecasting the time series. When forecasting δ steps ahead, it is natural to use MAFs with maximum correlation for $\Delta = \delta$, but other combinations are also possible, although not dealt with in this paper.

For a one step ahead forecast, we start by introducing a set of forecasting equations where the MAF correlations have been maximized for $\Delta = 1$, and where we have chosen to explain the residuals one step ahead by the three MAFs which show the greatest correlation over time $z_1(t)$, $z_2(t)$, and $z_3(t)$ (Table 4a). Note that the first two MAFs are highly significant, supporting the idea that the most important information is in the first MAFs. Remember that the MAFs are orthogonal and therefore only contribute new information when added to the model.

Table 4b presents the forecasting equations 4 quarters ahead, ($\delta = 4$), where the MAF correlations have been maximized for $\Delta = 4$.

Table 4a   Forecasting the residuals by the equations
$e(t+1) = \alpha_0 + \alpha_1 z_1(t) + \alpha_2 z_2(t) + \alpha_3 z_3(t)$, $\Delta = 1$.

	α_0	α_1	α_2	α_3	$R^2(e)\%$ $\Delta = 1$
$e_Y(t+1)$	117.0	−0.621 **	0.447 **	−0.122	25.8
$e_I(t+1)$	50.5	−0.507 ***	0.302 ***	−0.134	61.7
$e_G(t+1)$	0.57	0.128 **	−0.285 ***	−0.054	38.2
$e_C(t+1)$	78.2	−0.870 ***	−0.398 ***	0.123	64.0
$e_M(t+1)$	61.0	−1.090 ***	0.241 *	0.253	62.2
$e_X(t+1)$	45.2	−0.128	0.841 ***	0.596 **	48.3

Table 4b Forecasting the residuals by the equations
$e(t+4) = \alpha_0 + \alpha_1 z_1(t) + \alpha_2 z_2(t)\ \alpha_3 z_3(t),\ \Delta = 4.$

	α_0	α_1	α_2	α_3	$R^2(e)\%\ \Delta = 4$
$e_Y(t+4)$	165.0	0.248	−0.775 ***	−1.113 ***	46.8
$e_I(t+4)$	126.0	0.709 ***	−0.360 ***	−0.342 **	49.0
$e_G(t+4)$	−46.8	−0.782 ***	−0.055	−0.018	75.4
$e_C(t+4)$	273.0	−0.053	−0.986 ***	0.148	62.4
$e_M(t+4)$	218.0	1.011 ***	−0.863 ***	−0.337	44.0
$e_X(t+4)$	20.2	0.833 ***	0.012	−1.026 ***	35.2

The relative squared prediction error of forecasting equation (23) can be written in the following manner:[8]

$$1 - R^2(x) = [1 - R^2(e)]\,[1 - R^2(\mu)]$$

The last two columns of Table 5 contain the precision in forecasting equation (23) for the forecast generated by the trend and seasonal variation, plus the residual equations in Tables 4a and 4b respectively.

8 See appendix C.

Table 5 The precision in forecasting equation (23) for the forecast generated by the trend and seasonal variation plus the residual equations in Tables 4a and 4b respectively.

	$R^2(\mu)$	$R^2(x)$ $\Delta = \delta = 1$	$R^2(x)$ $\Delta = \delta = 4$
Y	96.8	97.6	98.2
I	82.3	93.2	91.0
G	98.0	98.8	99.5
C	86.8	95.2	95.0
M	79.1	92.1	88.3
X	92.4	96.1	95.1

A four steps ahead *within sample* forecast is sometimes better than a corresponding one step ahead forecast. This can be due to remaining seasonality in the residuals.

Some interesting questions are the *out of sample* performance and comparisons with other methods. We will consider two other methods. The first approach will forecast the time series using only the deterministic trend and seasonal variation and ignoring the residual information. The second approach also uses the deterministic trend and seasonal variation, but suggests a vector autoregressive model (VAR) when forecasting the residuals.

A general VAR–model can be written:

$$e(t) = C(L)e(t) + v(t)$$

where $C(L)$ are lag polynomials and $v(t)$ is a vector of white noise residuals.

We will consider the following special case of a VAR–model:

$$e(t+1) = Ce(t) + v(t)$$

where C is estimated by least squares.

Hence, we have the following three models:

(I) $x(t+1) = \mu(t+1)$

(II) $x(t+1) = \mu(t+1) + e_{MAF}(t+1)$

where

$$e_{MAF}(t+1) = \sum_{i=1}^{3} \alpha_i z_i(t)$$

(III) $x(t+1) = \mu(t+1) + e_{VAR}(t+1)$

where

$$e_{VAR}(t+1) = \alpha_1 e_Y(t) + \alpha_2 e_I(t) + \ldots + \alpha_6 e_X(t)$$

Note that $\mu(t+1)$ (corresponding to equation (20)) is the same in all three models, and is estimated by least squares based on the t first data points. The constant term α_0 is not present in models (II) and (III) for the residual forecasting equations. A reason for this is that in Tables 4a and 4b none of the constant terms are significantly different from zero. This could be expected since residuals are, by definition, constructed to have mean zero.

We estimate models (I), (II), and (III) from 1970:1 – 1977:3 and generate a one step ahead forecast (1977:4). Next we update the model using data including 1977:4 and again generate a one step ahead forecast (1978:1). This procedure is repeated until the observation from 1987:3 is included.

In order to compare the three different models we estimate their respective root mean square error:

$$RMSE = \left[\frac{1}{T-j} \sum_{t=j+1}^{T} (\hat{x}(t+1) - x(t+1))^2\right]^{1/2}$$

The results are presented in Table 6 below:

Table 6 The root mean square error (RMSE) for models (I), (II) and (III).

	(I)	(II) MAF	(III) VAR
Y	3 132	2 962	3 059
I	1 645	907	931
G	1 209	996	1 001
C	2 843	1 775	1 796
M	3 084	1 958	2 050
X	2 756	2 179	2 220

Table 6 shows us that there is still a great deal of information to be explained in the residuals. We can also see that MAF performs slightly better than VAR (although this might not be the "best" VAR or MAF model). This could be explained by the MAF technique trying to sort out the "signal" and throw away the "noise" before predicting, while VAR does not sort out the noise component.

6 CONCLUDING REMARKS

There are, of course, many remaining questions about how the MAF approach is related to vector autoregression and to cointegrated systems, which allow individual time series to be integrated, e.g. to have a linear, possibly stochastic trend, but require certain linear combinations of the series to be stationary. Here we will briefly deal with a few of them.

Consider the following special case of a VAR–model:

$$x(t) = Cx(t-1) + \nu(t) \tag{24}$$

where C is estimated by least squares. The same model would in our MAF approach be estimated in two steps. We would start by estimating the vectors γ_i in the equation:

$$z_i(t) = \gamma_i' x(t) \qquad\qquad i = 1, 2, \ldots, p \tag{25}$$

by maximizing the correlation over time. We would then run the regressions:

$$x(t+1) = \beta z(t) + \epsilon(t+1) = B(\gamma)x(t) + \epsilon(t+1) \tag{26}$$

and come up with a least square estimate of the vector β.

Clearly, we would obtain the same asymptotic estimates in both approaches, i.e. $B(\gamma) = C$.

Since more information is used in the MAF procedure, we would expect it to yield more precise estimates than VAR for finite samples.[9]

This conjecture is supported by the following example, which essentially is the Stock–Watson example mentioned in the introduction (we have omitted the price equation):

(i) $\quad y_t = y_t^p + y_t^s$

(ii) $\quad y_t^p = y_{t-1}^p + u_t$

(iii) $\quad c_t = y_t^p$

where $u_1 = 0$, $y_t^s = \theta_t$ with $E(\theta_t) = 0$, and $\theta_1, \theta_2, \ldots$ are identically independently distributed. Assume that we have three observations (c_1, y_1), (c_2, y_2), and (c_3, y_3). Similar to the proof in appendix A it can be shown that $z_1(t) = c_t$ and $z_2(t) = \frac{1}{\sqrt{2}}\theta_t$. We are forecasting c_{t+1} assuming that the constant term is zero. The MAF model to be estimated is:

$$c_{t+1} = \beta z_1(t) = \beta c_t$$

9 The argument is similar to the one used by Johansen (1988) to advocate a maximum likelihood estimator of the space of cointegration vectors instead of regression estimates.

and the least square estimate of β yields:

$$\hat{\beta}_{MAF} = 1 + \frac{u_3}{u_2}$$

The corresponding VAR–model is:

$$c_{t+1} = \alpha_1 c_t + \alpha_2 y_t$$

which can be rewritten as:

$$c_{t+1} = \beta c_t + \alpha \theta_t$$

The OLS estimates of β and α are:

$$\hat{\beta}_{VAR} = 1 + \frac{u_3}{u_2} - \frac{\theta_2}{\theta_1}$$

$$\hat{\alpha} = \frac{u_2}{\theta_1}$$

As the observant reader has realized $\hat{\beta}_{VAR}$ and $\hat{\beta}_{MAF}$ have no finite variance when θ_t and u_t are normally distributed with mean zero, since the ratio between two such variables is Cauchy–distributed. Technically this problem can be solved by not allowing values in an arbitrarily small symmetric interval around zero. We can then write:

$$\mathrm{Var}[\hat{\beta}_{VAR}] = \mathrm{Var}[\hat{\beta}_{MAF}] + \mathrm{Var}[\frac{\theta_2}{\theta_1}]$$

In a prediction of c_4 the squares of the predictions errors are, respectively:

$$E[\hat{c}_4(MAF) - c_4]^2 = \mathrm{Var}[u_3 + \frac{u_3^2}{u_2}] + \mathrm{Var}[u_4]$$

$$E[\hat{c}_4(VAR) - c_4]^2 = E[\hat{c}_4(MAF) - c_4]^2 + Var[-\frac{\theta_2}{\theta_1}(u_2 + u_3) + \frac{u_2}{\theta_1}\theta_3]$$

The last term equals:

$$3Var[u_2 \frac{\theta_3}{\theta_1}] = 3Var[u_2] \cdot E[\theta_3^2]\ E[\frac{1}{\theta_1^2}] \geq 3Var[u_2] \cdot \frac{E[\theta_3^2]}{E[\theta_1^2]} = 3Var[u_2]$$

where Jensen's inequality has been used. A simulation study (Sjöstedt (1991)) for larger samples ($n \geq 5$) justifies that the prediction error for VAR continues to be larger than for MAF. The study also supports that the errors reach the same limit as $n \to \infty$.

Worth mentioning is that we have run regressions with different numbers of MAFs, in order to study the optimal number of MAFs to include in the forecasting equations. We have also tried differentiation to make the time series stationary and have been able to reduce the prediction error further. This and related questions are to be discussed in a forthcoming paper.

Svensk sammanfattning

I uppsatsen introduceras en ny prognosmetodik som är baserad på en multivariat teknik som tillämpats i s k fjärranalys. Grundidén är helt enkelt att dela upp informationen i tidsserierna i två delar; en del som är intressant för prognosändamål och en del som inte är användbar i prognossyfte. Den senare delen rensas ut från prognosinstrumentet, då det brus som den annars skulle tillföra gör prognoserna mindre precisa.

I konjunktursammanhang är det en bra gissning att den information som är intressant för prognosändamål ligger inbäddad i ett mått på korrelationen över tiden. Vi bildar därför linjärkombinationer av en grupp av konjunkturvariabler – här komponenterna i bruttonationalprodukten – och väljer vikter så att korrelationen över tiden maximeras. När vi gjort detta en gång upprepas samma procedur med den skillnaden att vikterna nu väljs ortogonalt mot det första viktsystemet. Vi kan fortsätta att på detta sätt generera oberoende vikter lika många steg som antalet konjunkturvariabler. De linjärkombinationer av de ursprungliga tidsserierna som de härledda viktsystemen ger kallar vi Min/Max autokorrelationsfaktorer (MAF).

De MAF som skapas av viktsystem som ger låg korrelation över tiden rensar vi bort ifrån prognosinstrumentet, vilket i vårt exempel består av regressionsekvationer, där BNP–komponenter utgör de beroende variablerna, och de oberoende variablerna är MAF med hög korrelation över tiden. Teoretiskt förutsätter härledningen av MAF att de ingående tidsserierna är stationära och därför har vi rensat tidsserierna från trend innan korrelationen över tiden maximeras. I prognosinstrumentet finns trenden återlagd variabelvis.

Uppsatsen innehåller såväl teoretiska resultat kring egenskaperna hos MAF som en exemplifiering av användbarheten av prognostekniken. Vi jämför också, utom samplet, vår prognosidé med en variant av s k vektorautoregression, vilken svarar mot ett prognosinstrument där alla MAF finns med i regressionsekvationen. Det visar sig att vår prognosmetod resulterar i ett mindre prognosfel mätt som "Root Mean Square Error".

Appendix A

For the Stock and Watson example discussed in the Introduction the following holds:

$$\max_a \operatorname{corr}(z_t, z_{t-1}) = \max_a \frac{(a_1+a_2)^2\operatorname{cov}(c_t, c_{t-1}) + a_3^2\operatorname{cov}(p_t,p_{t-1})}{\sqrt{B_t B_{t-1}}} \tag{A.1}$$

subject to $\|a\| = 1$

is maximized by $a_2 = 1$ and $a_1 = a_3 = 0$. Here $B_t = V(z_t) = (a_1 + a_2)^2V(c_t) + a_3^2V(p_t) + a_1^2V(y_t^s)$.

PROOF: To be able to find unique MAFs, we assume that $p_0 = 0$, $c_0 = u_0$ and $p_t = c_t = 0$ when $t < 0$. This yields $V(c_t) = t+1$, $V(p_t) = t$, $\operatorname{cov}(c_t, c_{t-1}) = t$ and $\operatorname{cov}(p_t, p_{t-1}) = t-1$.

Assume that (a_1+a_2) and a_3 are optimally chosen. It is then obviously true that $\operatorname{corr}(\cdot)$ is a decreasing function of a_1^2. Hence, $a_1 = 0$ at maximum, implying that $a_3^2 = 1 - a_2^2$.

Now let

$$g(a_2^2) = \frac{a_2^2\operatorname{cov}(c_t, c_{t-1}) + (1-a_2^2)\operatorname{cov}(p_t, p_{t-1})}{\sqrt{a_2^2V(c_t) + (1-a_2^2)V(p_t)}\;\sqrt{a_2^2V(c_{t-1}) + (1-a_2^2)V(p_{t-1})}}$$

$$= \frac{a_2^2 t + (1-a_2^2)(t-1)}{\sqrt{a_2^2(t+1) + (1-a_2^2)t}\;\sqrt{a_2^2 t + (1-a_2^2)(t-1)}} =$$

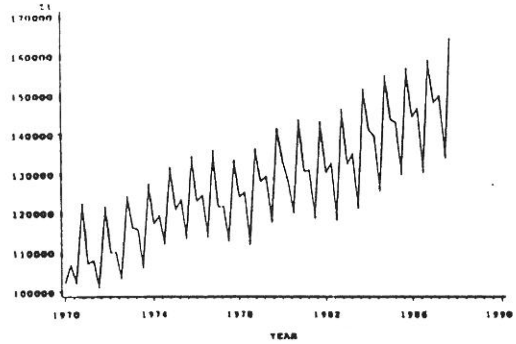
$$= \sqrt{\frac{a_2^2 + t - 1}{a_2^2 + t}} = \sqrt{1 - \frac{1}{a_2^2 + t}}.$$

Hence, $g(a_2^2)$ is maximized for $a_2^2 = 1$. Since $a_1 = 0$ and $a_3^2 = 1 - a_1^2 - a_2^2 = 0$, this proves the claim.

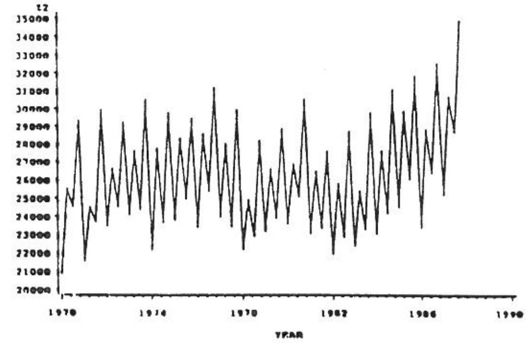
Remark: The statement in (A.1) is true regardless of the specification of the processes c_t and p_t as long as (i) and (iii) in relation (1) hold, y_t^s is white noise and we have orthogonality between y_t^s, p_t and c_t and $corr(p_t, p_{t-1}) < corr(c_t, c_{t-1})$. If $corr(p_t, p_{t-1}) > corr(c_t, c_{t-1})$, the first MAF is p_t instead of c_t.

Appendix B

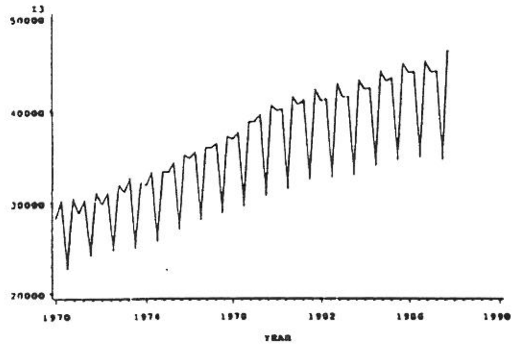
GNP



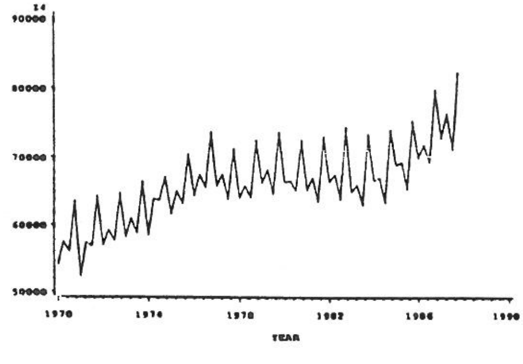
Investments



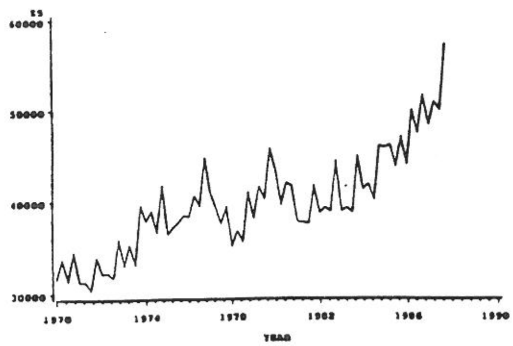
Government consumption



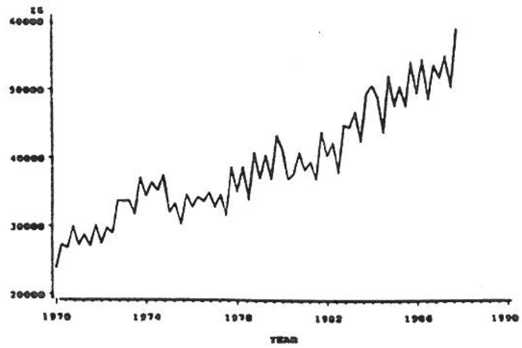
Private consumption



Import



Export



Appendix C

Proof that $(1-R^2(x)) = (1-R^2(\mu))(1-R^2(e))$

From the definition of $R^2(\mu)$, it follows that:

$$1 - R^2(\mu) = \frac{\Sigma(x(t) - \hat{\mu}(t))^2}{\Sigma(x(t) - \bar{x})^2} = \frac{\Sigma e(t)^2}{V}$$

The corresponding expression from the regression $\hat{e}(t+1) = f(\cdot)$ is given by:

$$1 - R^2(e) = \frac{\Sigma(e(t+1) - \hat{e}(t+1))^2}{\Sigma e(t+1)^2},$$

since $\bar{e} = \frac{1}{T}\,\Sigma e(t) = 0$

from the regression technique which was used to determine $\hat{\mu}(t)$.

The predictor $\hat{x}(t+1)$ is obtained as:

$\hat{x}(t+1) = \hat{\mu}(t+1) + \hat{e}(t+1)$

with the relative squared prediction error:

$$1 - R^2(x) = \frac{\Sigma(\hat{x}(t+1) - x(t+1))^2}{V}$$

Using the definition of $R^2(\mu)$, and that $\hat{x}(t+1) - x(t+1) = \hat{e}(t+1) - e(t+1)$ proves the claim.

REFERENCES

Box, G.E.P., and Tiao, G.C. (1977). A Canonical Analysis of Multiple Time Series. **Biometrika** 64, 355–365.

Burns, A.F., and Mitchell, W.C. (1946). **Measuring Business Cycles**. New York: National Bureau of Economic Research.

Conradsen, K., Kjær Nielsen, B., and Thyrstedt, T. (1986). **A Comparison of Min/Max Autocorrelation Factor Analysis and Ordinary Factor Analysis.** IMSOR, Technical University of Denmark, Lyngby (mimeographed).

Frisch, R. (1933). Propagation Problems and Impulse Problems in Dynamic Economics. In: **Essays in Honour of Gustaf Cassel**. London: George Allen.

Goodwin, R. (1951). The Non–linear Accelerator and the Persistence of Business Cycles. **Econometrica** 19, 1–17.

Gordon, R.J. (1985). Continuity and Change in Theory, Behavior, and Methodology. In: **The American Business Cycle, Continuity and Change**, R.J. Gordon (ed.). Chicago: NBER.

Grandmont, J.M. (1985). On Endogeneous Competitive Business Cycles. **Econometrica** 53, 995–1045.

Harrod, R.F. (1936). **The Trade Cycle**. Oxford: Oxford University Press.

Hicks, J.R. (1950). **A Contribution to the Theory of the Trade Cycle**. Oxford: Oxford University Press.

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

Kaldor, N. (1940). A Model of the Trade Cycle, **Economic Journal** 50, 78–92.

King, R.G. and Plosser, C.I. (1984). Money Credit and Prices in a Real Business Cycle. **American Economic Review** 74, 363–380.

Kydland, F. and Prescott E.C. (1982). Time to Build and Aggregate Fluctuations. **Econometrica** 50, 1345–1370.

Long, J.B. and Plosser, C.I. (1983). Real Business Cycles. **Journal of Political Economy** 91, 39–69.

Lucas, R.E. Jr. (1976). Econometric Policy Evaluation: A Critique. **Carnegie–Rochester Conference Series on Public Policy**, 19–46..

Prescott, E. (1986). Theory Ahead of Business Cycle Measurement. **Federal Reserve Bank of Minneapolis Quarterly Review** (Fall).

Puu, T. (1987). Complex Dynamics in Continuous Models of the Business Cycle. In: **Economic Evolution and Structural Adjustment**. Lecture Notes in Economics and Mathematical Systems 293:227–259. Berlin: Springer– Verlag, Batten, D., Casti, J. and Johansson, B. (eds).

Rao, C.R. (1973). **Linear Statistical Inference and its Applications.** Wiley, New York. Second Edition.

Samuelson, P.A. (1939). A Synthesis of the Principle of Acceleration and the Multiplier. **Journal of Political Economy** 47, 786–797.

Sims, C.A. (1980). Macroeconomics and Reality. **Econometrica** 48, 1–48.

Sims, C.A. (1981). An Autoregressive Index Model for the U.S., 1948– 1975. In: **Large–Scale Macro–Econometric Models.** Kmenta, J. and Ramsey, J.B. (eds.). Amsterdam: North–Holland.

Sjöstedt, S. (1988). Multivariat analys av ekonomiska tidsserier med hjälp av min/max autokorrelationsfaktorer. Department of Mathematical Statistics, University of Umeå (mimeo).

Sjöstedt, S. (1991). Forecasting Multivariate Time Series – A Case Study on Swedish National Accounts. Department of Mathematical Statistics, University of Umeå (mimeo).

Slutsky, E. (1927). **The Summation of Random Causes as the Source of Cyclic Processes** 3, No. 1. Moscow: Conjuncture Institute. In Russian with English Summary.

Stock, J.H. and Watson, M.W. (1988). Variable Trends in Economic Time Series. **Journal of Economic Perspective** 2, 147–174.

Switzer, P. (1985). Min/Max Autocorrelation Factors for Multivariate Spatial Imagery. In: **Computer Science and Statistics: The Interface,** L. Billard (ed.). Amsterdam: Elsevier Science Publishers B.V.

Switzer, P. and Green A.A. (1984). Min/Max Autocorrelation Factors for Multivariate Spatial Imagery. **Technical Report** No 6. Department of Statistics, Stanford University, 14 pp.

Switzer, P. and Ingebritsen, S.E. (1986). Ordering of Time–Difference Data From Multispectral Imagery. **Remote Sensing of Environment** 20, 85–94.

Prevoius titles in this serie:

No. 1 Current Account and Business Cycles: Stylized Facts for Sweden by Anders Warne and Anders Vredin. December 1989.

No. 2 Change in Technical Structure of the Swedish Economy by Göran Östblom. December 1989.

No. 3 Mamtax. A Dynamic CGE Model for Tax Reform Simulations by Paul Söderlind. December 1989.

No. 4 The Supply Side of the Econometric Model of the NIER by Alfred Kanis and Aleksander Markowski. November 1990.

No. 5 The Financial Sector in the SNEPQ Model by Lennart Berg. February 1991.

No. 6 Consumer Attitudes, Buying Intentions and Consumption Expenditures. An Analysis of the Swedish Household Survey Data by Anders Ågren & Bo Jonsson. April 1991.

No. 7 A Quarterly Consumption Funtion for Sweden 1979-1989 by Lennart Berg & Reinhold Bergström. October 1991.

No. 8 Good Business Cycle Forecasts - A Must for Stabilization Policies by Lars-Erik Öller. February 1992.

No. 9 Forecasting Car Expenditures Using Household Survey Data by Bo Jonsson and Anders Ågren. February 1992.