WORKING PAPER
No. 11 February 1992

CURRENT QUARTER FORECASTS OF SWEDISH GNP USING MONTHLY VARIABLES

Stefan Gerlach

Svensk resumé på sid 22.

January 1992
Second Draft

CURRENT QUARTER FORECASTS OF SWEDISH GNP USING MONTHLY VARIABLES

Stefan Gerlach

Brandeis University

This paper is a part of a research project on business cycles in Sweden, sponsored by The Swedish Economic Council and The National Institute of Economic Research (NIER). Helpful conversations with Alf Carling, Bo Dahlerus, Irma Rosenberg, Hans Sterte, Sven-Erik Sveder and, in particular, Lars-Erik Öller are gratefully acknowledged.

1. Introduction

Since economic data typically are reported with important time lags (and with measurement errors), economic policy decisions and short-term forecasts must be made subject to substantial uncertainty about the current state of the economy. This raises the important issue of how estimates can be made of current economic activity. In this paper we show that accurate "flash estimates" of quarterly GNP can be provided using simple econometric techniques that are implemented in standard econometric software packages. It should be stressed that the emphasis here is on ease of use, not on econometric sophistication.

The paper is structured as follows. In the second section we explain the methodology used, describe the data, and discuss the treatment of seasonality. Section 3 contains our estimates. We first estimate a univariate model for GNP, which serves as a benchmark. The model is estimated for data spanning 1974:2-1987:4, and is then used to provide out-of-sample flash estimates of GNP growth for the period 1988:1-1990:3. The actual forecast error for this model is roughly one percent. We next estimate a second model in which we introduce the growth of industrial production between the first month of the previous quarter and the first month of the current quarter, and show that the resulting out-of-sample forecast errors are about 22 percent smaller than when only information on GNP growth was used. This result suggests that the simple method used here is practically useful. Finally, we incorporate monthly sales and electric energy production in a third model, and show that although all three

variables are statistically highly significant in the econometric model, the actual out-of-sample forecasting performance worsens. The paper ends with some brief concluding comments.

2. Methodology

Consider a macroeconomic forecaster who is trying to estimate real GNP for the current quarter. Suppose that the forecaster has access to real GNP data for the preceding quarter, the log of which we denote y_{t-1}, and to some monthly data from the current quarter for some variable, the log of which we denote x_t.[1] For instance, industrial production in the first month of the quarter may be known before the end of the quarter. Since GNP and industrial production are both non-stationary, we use the first differences of y_t and x_t in the empirical analysis.[2] Furthermore, since the time series display substantial seasonality, we use seasonally differenced data and fourth-order AR and MA terms in the regressions reported below.

Consider the following simple model for GNP growth, proposed by Rathjens and Robins (1989)[3]

---

[1] Naturally, if there is data on several variables available, x_t may be interpreted as a vector.

[2] For a monthly series available in month i (i=1,2,3) of each quarter, the growth rate for quarter t is calculated as the percent growth from month i in quarter t-1, to month i in quarter t. Thus, for industrial production, which is known for the first month in each quarter, we have that $\Delta x_1 = \ln(x_{Jan}/x_{Oct})$, $\Delta x_2 = \ln(x_{April}/x_{Jan})$, $\Delta x_3 = \ln(x_{July}/x_{April})$, and $\Delta x_4=\ln(x_{Oct}/x_{July})$.

[3] A more general version of (1a) is obtained by including lagged Δy_t and Δx_t terms, that is, $A(L)\Delta y_t=C(L)\Delta x_t+\varepsilon_t$, where A(L) and C(L) are polynomials in the lag operator.

(1a) $$\Delta y_t = \gamma + c\Delta x_t + \varepsilon_t$$

(1b) $$\alpha(L)\varepsilon_t = \beta(L)u_t$$

where $\alpha(L)=1-\alpha_1 L^1-\ldots-\alpha_q L^q$ and $\beta(L)=1+\beta_1 L^1+\ldots+\beta_v L^v$. Eq. (1a) states that the rate of growth of GNP in the current quarter is related to the rate of growth of the monthly variable and a random error. Eq. (1b) governs the behavior over time of the error term, and states that it follows an ARMA-process.[4]

Given the model in (1a) and (1b), the current quarter forecast of Δy_t, denoted $\Delta\hat{y}_t$, is given by

(2) $$\Delta\hat{y}_t = \hat{\gamma} + \hat{c}\Delta x_t + \hat{\varepsilon}_t$$

where the expected value of ε_t can be computed from estimates of (1b).

Before proceeding, it should be stressed that the model is easily estimable using any standard econometric software package that can estimate linear equations with ARMA errors.

3. Some Empirical Results

In this section we attempt to forecast the growth rate of GNP in Sweden in the current quarter, using lagged quarterly growth rates of GNP, and monthly data on industrial production, electric energy production, and retail sales. These variables were chosen

[4]Estimation of models of this form is discussed in Box and Jenkins (1976) and Judge et al., (1985, ch. 8).

since NIER has previously used them to assess business cycle developments.[5]

To clarify our strategy, consider the following real life example: a forecaster would like to construct, using data available in the middle of March, a flash estimate of GNP growth during the first quarter. The researcher has access to four pieces of information: (i) GNP in the fourth quarter of the preceding year (y_{t-1}); (ii) the level of industrial production in January ($x_{1,t}$, which is reported with a 6 week lag); (iii) the level of retail sales in January ($x_{2,t}$, which also is reported with a 6 week lag); and (iv) the level of electric energy production in February (x_{3t}, which is reported with a 2 week lag).[6] Since all series are non-stationary, quarterly growth rates (computed using first differences of the logs of the series) are used in the empirical work.

3.1. Preliminaries

In the lower part of Figures 1-4 we have plotted the four data series mentioned above. As can be seen, in all cases there is very pronounced seasonality. GNP, for instance, contracts about 15 percent between the second and third quarter, and expands about 20 percent between the third and fourth quarters. The seasonality completely dominates any business cycles movements in the different series. This is illustrated by Figure 5, in which

[5] The data are all "day-corrected," but are not seasonally adjusted.

[6] We used real GNP in 1985 prices.

the quarterly growth rate of GNP has been plotted against the quarterly growth of industrial production. As can be seen, the data is clustered in four distinct groups, one for each season, so that a regression of Δy_t on Δx_t would only pick up the seasonality.

The severe seasonality in the data implies that we need to allow for seasonality in our simple model. The seasonality can be handled by using seasonally differenced data, and by allowing for seasonality in the error term. To seasonally difference the series, we compute

(3) $$\Delta y_t^* = \Delta y_t - \Delta y_{t-4}$$

and similarly for the Δx_t series. The seasonally differenced series are plotted in the upper part of Figures 1-4. As can be seen, the movements in the series are now much reduced; the growth rates of quarterly GNP, for instance, fall mainly in the interval ±2.5 percent. Moreover, the scattergram in Figure 6 suggests that the seasonally differenced data contain much more economically meaningful information than the unadjusted data.

The second step that is necessary is to allow for seasonality in the ARMA model for the error term in (1a). One way to do this parsimoniously is to incorporate seasonal lag polynomials in (1b). Thus, the model is given by

(4a) $$\Delta y_t^* = \gamma + c\Delta x_t^* + \varepsilon_t$$

(4b) $$\alpha(L)\alpha_s(L)\varepsilon_t = \beta(L)\beta_s(L)u_t$$

where $\alpha_s(L^4)=1-\alpha_{s1}L^4-\ldots-\alpha_{sp}L^{4p}$ and $\beta_s(L^4)=1+\beta_{s1}L^4+\ldots+\beta_{sq}L^{4q}$ contain the seasonal AR and MA terms. The specification of $\alpha(L)$, $\alpha_s(L)$, $\beta(L)$ and $\beta_s(L)$ is an empirical question; for the forecasting problem considered below, it appears that the seasonal $\alpha_s(L)$ and $\beta_s(L)$ polynomials can be disregarded.

Results

Below we estimate a few models of the form (4a) and (4b), and consider their forecasting abilities. In order to evaluate our models in a realistic manner, we estimate the model using data for the period 1974:2-1987:4, and investigate their forecasting ability for the period 1988:1-1990:3.

Given the results from some tentative empirical work, we estimate three models below; the first of these is a univariate model for Δy_t; the second also incorporates information of industrial productivity growth; and the third incorporates all three monthly variables.

Model 1

In our first model, which is to be viewed as a benchmark, we attempt to forecast Δy_t^* using only its own past values. That is, we estimate

$$\text{(5a)} \qquad \Delta y_t^* = \gamma + \varepsilon_t$$

After investigating the autocorrelation and partial autocorrelation functions, and after estimating and evaluating

several different models, we selected the following ARMA model for the errors

$$(5b) \qquad (1 + \alpha_1 L + \alpha_4 L^4)\varepsilon_t = (1 - \beta_4 L^4)u_t$$

The model was estimated using data for the period 1974:2-1987:4. As the results in Table 1 indicate, the parameters are all significant using standard confidence levels, the residuals appear serially uncorrelated and pass a Jarque-Bera test for normality, as well as tests for ARCH and hetereoscedasticity. In short, the model appears to fit the data well.

Next, we use the model to provide out-of-sample flash estimates of GNP for the period 1988:1-1990:3. To evaluate the model's forecasting performance, we calculate the forecast errors and their standard errors. It should be noted that the forecast error variance at time τ (τ>t) depends on the Δx_τ relative to the the mean of Δx_t during the estimation interval.[7] (Thus, the confidence interval for the forecast error varies across observations.)

In Figure 7 we plot the forecast errors and a ±2 standard deviation wide confidence band. As the diagram shows, the model appears to forecast current quarter GNP growth very well, with a Root Mean Square (RMS) error of .0117. Since the data are in logs, the RMS can be interpreted as indicating that the average

---

[7]The variance of the forecast error in eq. (2) for time τ can be written (Pindyck and Rubinfeld, 1991, p. 186)

$$\sigma_f^2 = \sigma^2\left[1 + 1/T + (\Delta x_\tau - \Delta\bar{x})/[\sum_{t=1}^{T}(\Delta x_t - \Delta\bar{x})^2]\right]$$

forecast error is 1.17 percent. In Figure 8 we plot GNP, actual and as predicted by the model. As can be seen, seasonal variations are the primary factor driving GNP movements. These are, however, captured by the model. All in all, the two diagrams suggest that this very simple univariate model does a good job forecasting GNP. We next turn to the question whether the forecasts of current quarter GNP can be further improved by incorporating information about the behavior of the monthly variables during the beginning of the quarter.

Model 2

In the second model we incorporate the growth rate of industrial production between the first month of the last and the current quarter, Δx_{1t}, as an additional explanatory variable in the model. Since industrial production is available with a six week lag, this model can be used for forecasts of current quarter GNP sometime after the middle of the third month of the quarter.

The model can now be written

$$\text{(6a)} \qquad \Delta y_t^* = \gamma + c\Delta x_{1t}^* + \varepsilon_t$$

Following the same procedure as above, we selected the following model for the errors

$$\text{(6b)} \qquad (1 + \alpha_1 L)\varepsilon_t = (1 - \beta_2 L^2 - \beta_4 L^4)u_t$$

The results in Table 1 indicate that the growth rate of industrial production is statistically very strongly linked to the growth

rate of GNP. Since this model uses more information than the earlier model, we would expect it to provide superior flash estimates of GNP. Before we evaluate the forecasting performance of the model, note that the parameters are again significant, and that the model passes the battery of diagnostic tests used above.

The out-of-sample forecasting performance of the model can be evaluated by inspection of Figure 9.[8] The RMS of .0091 (which is 22 percent smaller than for Model 1) indicates that the mean forecast error is less than one percent. Furthermore, and as witnessed by the narrower confidence band, the theoretical forecast error of this model is smaller than the theoretical forecast error of the univariate model discussed above. Moreover, as indicated by Figure 10, the model seems to forecast the level of GNP quite nicely. Although there are good reasons to be careful in interpreting RMS from different models, this result suggests that the incorporation of monthly data may well improve our ability to forecast GNP.

Model 3

Finally, we include in the model the growth rate of all three monthly variables that we observe; the growth of industrial production, Δx_{1t}, the growth rate of electric energy production, Δx_{2t}, and the growth rate of sales, Δx_{3t}. Since the sales data are available with a six week lag, we use data for the first month of the quarter. Data on electric energy production, however, is

---

[8]Note that the forecasts are unconditional, since Δx_t is known when the forecast of Δy_t is made.

available with a two week lag, and we therefore use data from the second month of the quarter.

The model can now be written

(7a) $$\Delta y_t^* = \gamma + c_1 \Delta x_{1t}^* + c_2 \Delta x_{2t}^* + c_3 \Delta x_{3t}^* + \varepsilon_t$$

Using the same procedure as before, we settled for the following model

(7b) $$(1 + \alpha_4 L^4)\varepsilon_t = (1 - \beta_1 L - \beta_8 L^8) u_t$$

Table 1 indicates that the all three monthly variables are highly significant. This suggests that the forecast performance of this model is likely to be better than the forecasting performance of the two earlier models.[9] The results in Figure 11 suggest, however, that this is not the case. As the RMS of .0159 indicates, the out-of-sample forecast error is about 32 percent larger than the forecast error for the univariate model. Furthermore, for two of the eleven forecasts, the prediction error is larger than two standard deviations. Thus, the use of all three monthly data series appears to have worsened the forecasting performance of the model.

The increase in the RMS as additional variables are included in the model is, however, not necessarily surprising. First, the actual forecast errors, and thus the RMS, represent particular realizations of random variables. In order to draw any firm

[9] This hypothesis is supported by the fact that the confidence band is smaller than the confidence bands for the two earlier models.

conclusions about a models' forecasting performance from the RMS, it is necessary to know whether the RMS:s of different models are statistically different. Given the RMS:s are calculated on the basis of only 11 forecasts, it is unlikely that we would find statistically significant differences in the RMS:s. This consideration suggests that the slight increase in the RMS for Model 3 should not lead us to immediately reject the model. Instead, the model's forecasting performance needs to be evaluated over a longer time period.

Second, in constructing the out-of-sample forecasts we have implicitly assumed that there is no structural break between the estimation and forecasting period. This assumption may be unwarranted.

Third and as noted above, forecast errors are partially due to random errors in the estimated regression parameters.[10] As we add additional variables, we lose degrees-of-freedom, which is likely to increase the variance of the estimated coefficients and thereby worsen the model's forecasting performance.

---

[10] The expected forecast error (for $\tau>t$) for eq. (2) can be written as (Pindyck and Rubinfeld, 1991, p. 185)

$$E(\Delta y_\tau-\Delta \hat{y}_\tau) = E(\gamma-\hat{\gamma}) + E(c-\hat{c})\Delta x_\tau + \varepsilon_\tau$$

so that the variance of the forecast error, σ_f^2, can be written

$$\sigma_f^2 = \mathrm{Var}(\hat{\gamma}) + \Delta x_\tau^2 \mathrm{Var}(\hat{c}) + 2\Delta x_\tau \mathrm{Cov}(\hat{\gamma},\hat{c}) + \sigma^2$$

Thus, σ_f^2 depends both on the variance of ε_τ, and the precision with which γ and c have been estimated.

4. Conclusions

In this paper we have reviewed an easily implementable econometric strategy for forecasting current quarter GNP in Sweden. The strategy is based on estimating simple linear regressions with ARMA errors, which are estimable using standard econometric packages. We showed that the models, despite their simplicity, yield accurate forecasts of GNP. Furthermore, the theoretical forecast error variances, and an evaluation of actual out-of-sample forecast errors, suggested that the incorporation of monthly data improves the models' forecasting ability. The differences in actual forecasts are quite small, which suggests that the models' performance needs to be evaluated over a longer time period before final conclusions can be drawn.

The fact that the monthly variables are so strongly related to quarterly GNP growth suggests that it may be possible to forecast monthly GNP, which is an unobserved variable, using data on the three monthly time series used here.[11] This should prove a fruitful topic for future research.

---

[11] Harvey (1990, ch. 6) discusses how Kalman filtering techniques can be used for this purpose.

REFERENCES

G. E. P. Box and G. M. Jenkins, 1976, Time Series Analysis: Forecasting and Control, Holden-Day.

A. C. Harvey, 1989, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lütkepohl and T.-C. Lee, 1985, The Theory and Practice of Econometrics, second ed., Wiley.

R. Pindyck and D. Rubinfeld, 1991, Econometric Models and Economic Forecasts, third ed., McGraw-Hill.

P. Rathjens and R. Robins, 1989, "Improved Current Quarter Forecasts," Working paper.

TABLE 1

MODEL 1

$$\Delta y_t^* = \gamma + \varepsilon_t$$

$$(1 + \alpha_1 L + \alpha_4 L^4)\varepsilon_t = (1 - \beta_4 L^4)u_t$$

<u>Estimated Coefficients (t-Statistics)</u>

γ:	0.00	(0.00)	α_1:	0.44	(4.29)
α_4:	0.31	(2.51)	β_4:	-0.33	(1.98)

R^2	0.48	Adj. R^2	0.45	SEE	0.015	Loglik.	155.97

<u>Marginal Significance Levels for Diagnostic Tests</u>

Breusch-Godfrey, 4 lags	0.30	Jarque-Bera	0.97
Box-Pierce, 4 lags	0.82	Arch(4)	0.57
Ljung-Box, 4 lags	0.80	White	NA.

MODEL 2

$$\Delta y_t^* = \gamma + c_1 \Delta x_{1t}^* + \varepsilon_t$$

$$(1 + \alpha_1 L)\varepsilon_t = (1 - \beta_2 L^2 - \beta_4 L^4)u_t$$

<u>Estimated Coefficients (t-Statistics)</u>

γ:	0.00	(0.03)	c_1:	0.11	(3.65)
α_1:	0.56	(5.71)	β_2:	-0.31	(2.54)
			β_4:	-0.69	(6.55)

R^2	0.62	Adj. R^2	0.59	SEE	0.013	Loglik.	165.15

<u>Marginal Significance Levels for Diagnostic Tests</u>

Breusch-Godfrey, 4 lags	0.25	Jarque-Bera	0.45
Box-Pierce, 4 lags	0.96	Arch(4)	0.85
Ljung-Box, 4 lags	0.95	White	0.77

TABLE 1 (Cont.)

MODEL 3

$$\Delta y^*_t = \gamma + c_1 \Delta x^*_{1t} + c_2 \Delta x^*_{2t} + c_3 \Delta x^*_{3t} + \varepsilon_t$$

$$(1 + \alpha_4 L^4)\varepsilon_t = (1 - \beta_1 L - \beta_8 L^8) u_t$$

Estimated Coefficients (t-Statistics)

γ:	0.00	(0.11)	c_1:	0.09	(3.70)
c_2:	0.07	(4.78)	c_3:	0.15	(3.44)
α_4:	0.69	(5.82)	β_1:	-0.49	(3.81)
			β_8:	-0.37	(2.37)

R^2	0.71	Adj. R^2	0.67	SEE	0.011	Loglik.	172.15

Marginal Significance Levels for Diagnostic Tests

Breusch-Godfrey, 4 lags	0.80	Jarque-Bera	0.50
Box-Pierce, 4 lags	0.96	Arch(4)	0.72
Ljung-Box, 4 lags	0.95	White	0.48

FIGURE 1: QUARTERLY GNP GROWTH

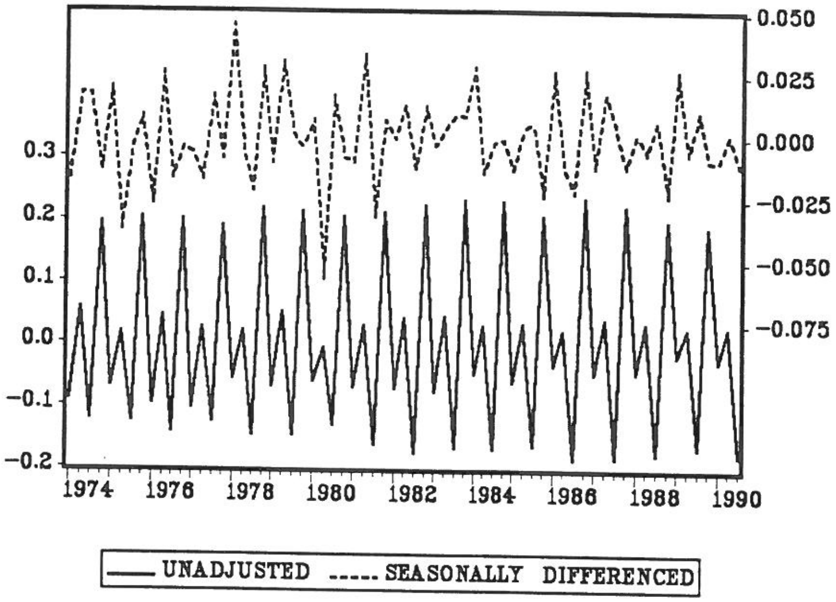


FIGURE 2: QUARTERLY GROWTH OF INDUSTRIAL PROD.

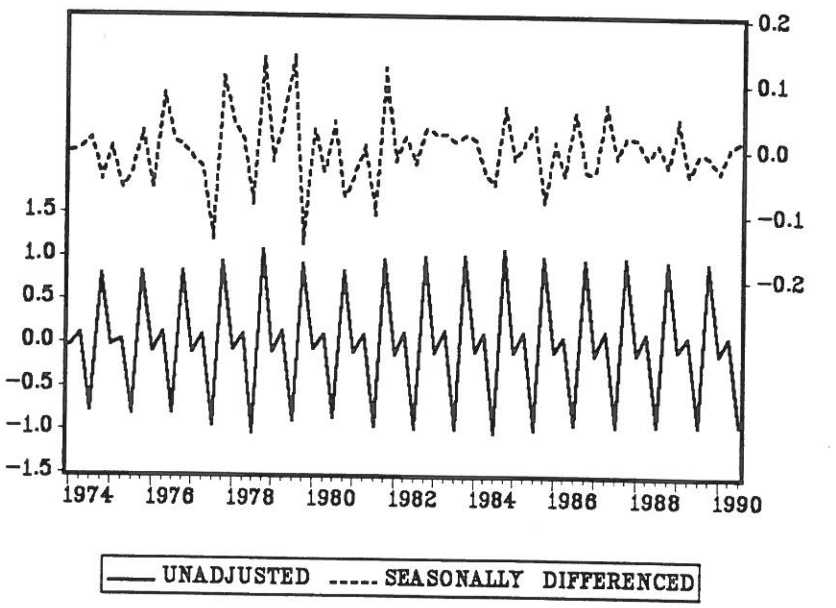


FIGURE 3: QUARTERLY GROWTH OF ELEC. ENERGY PROD.

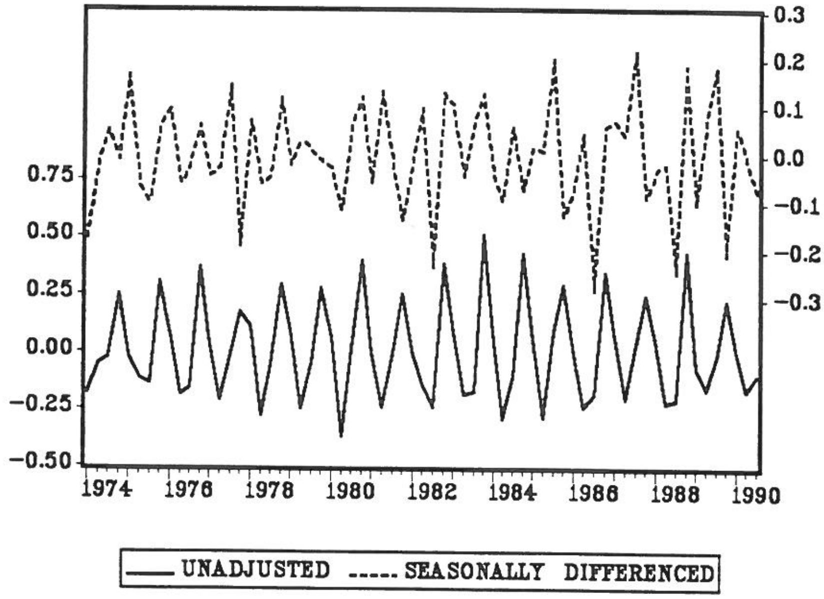


FIGURE 4: QUARTERLY GROWTH OF SALES

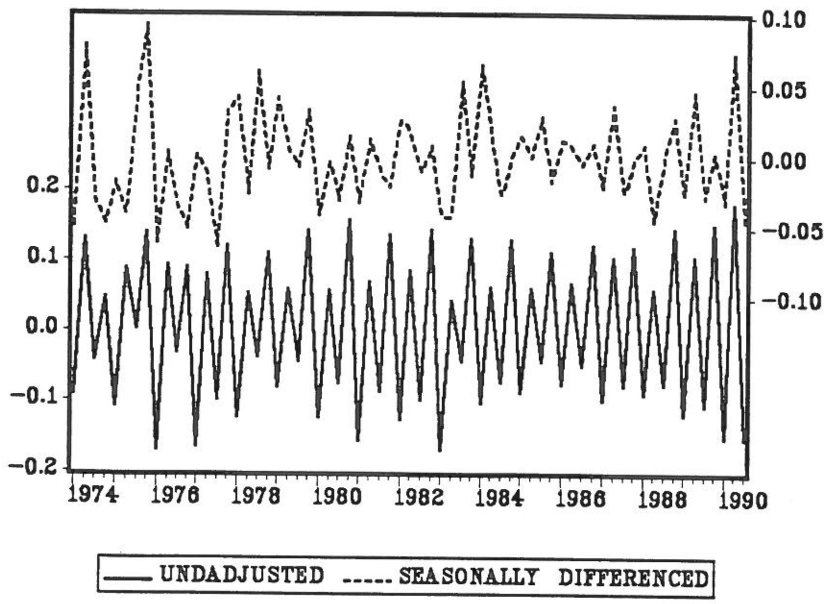


FIGURE 5: UNADJUSTED GNP AND IND. PROD.

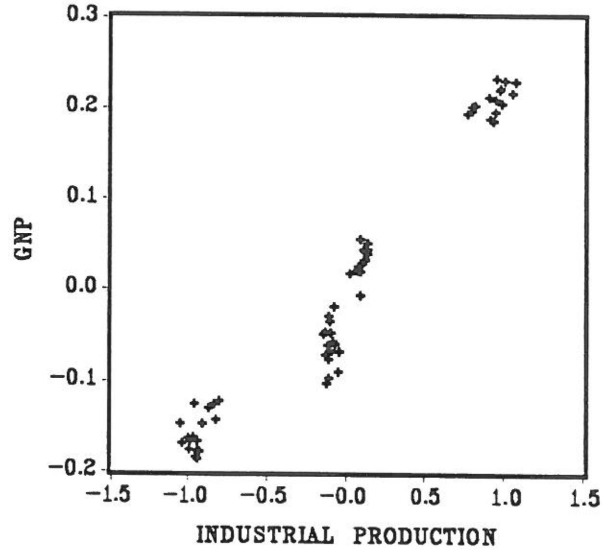


FIGURE 6: SEASONALLY DIFF. GNP AND IND. PROD.

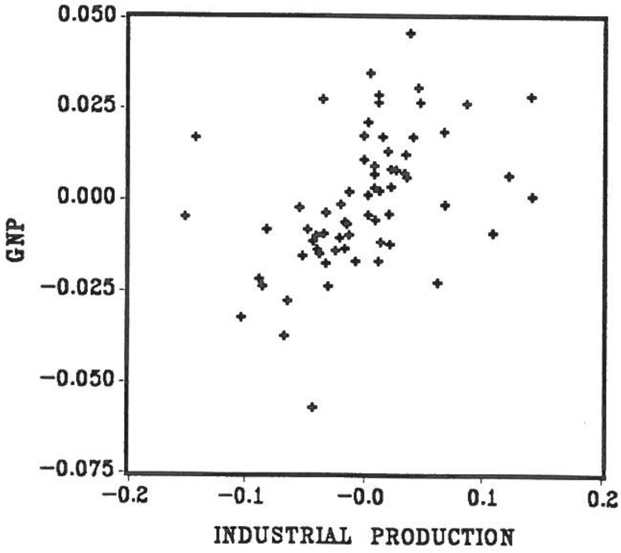


FIGURE 7: MODEL 1

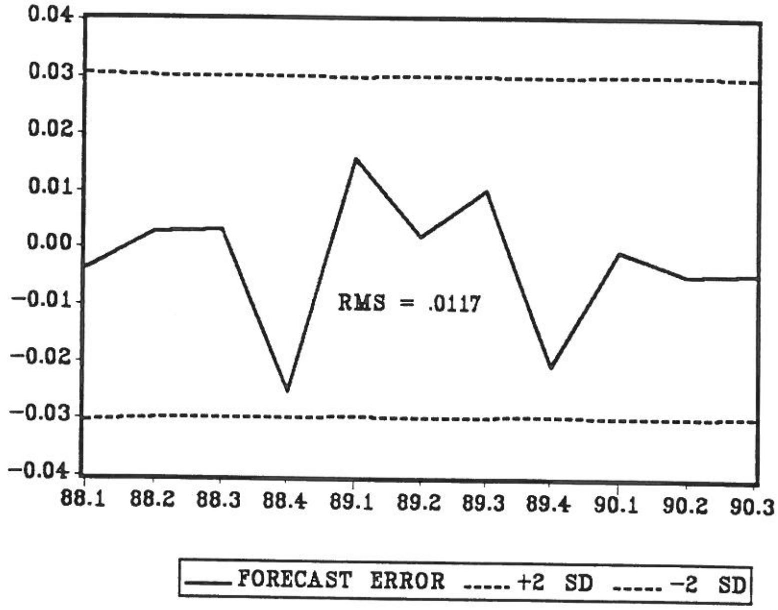


FIGURE 8: MODEL 1

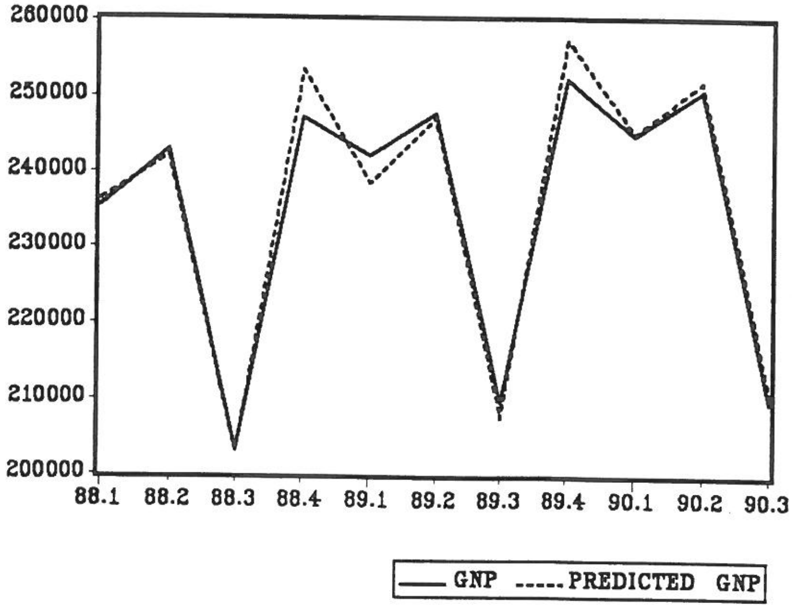


FIGURE 9: MODEL 2

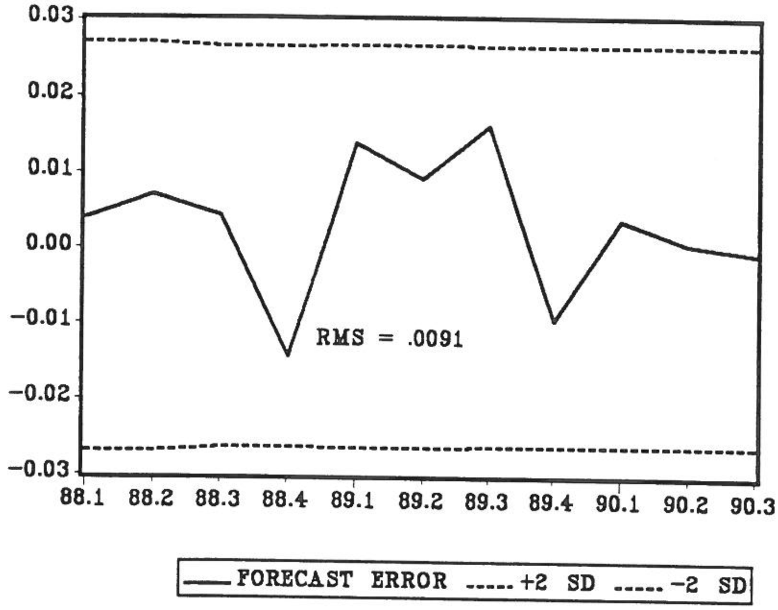


FIGURE 10: MODEL 2

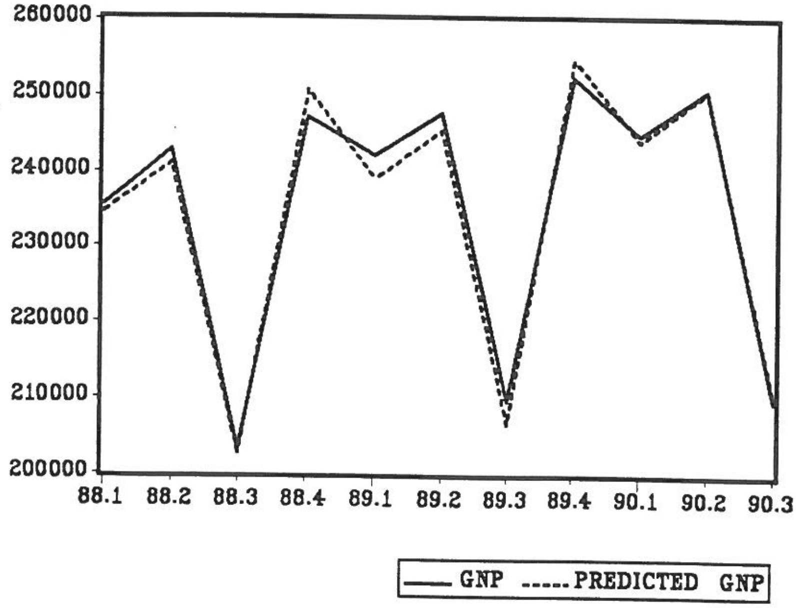


FIGURE 11: MODEL 3

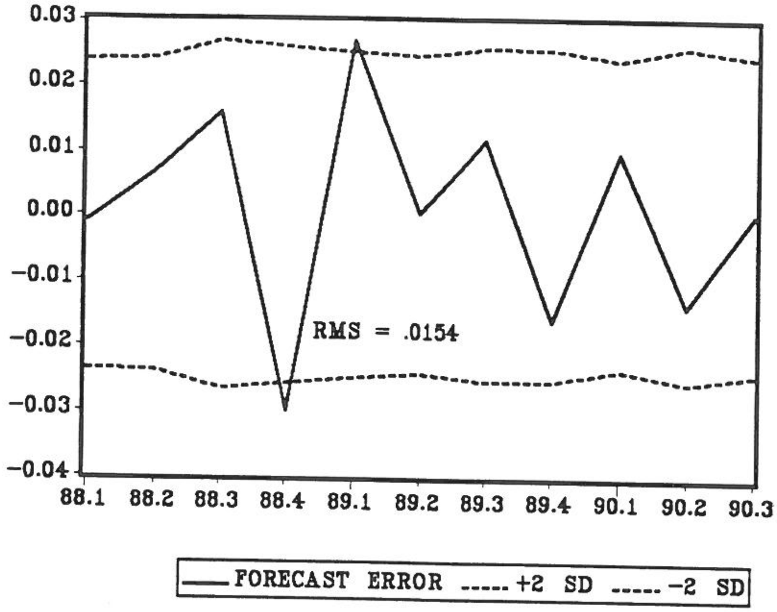
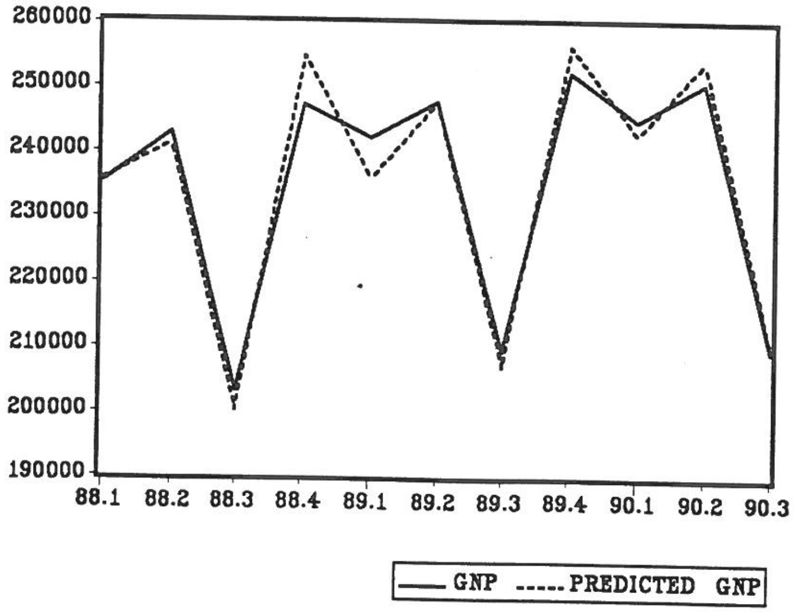


FIGURE 12: MODEL 3



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.

No. 10 Abstract: Forecasting the Business Cycle Not Using Minimum Autocorrelation Factors by Karl-Gustaf Löfgren, Bo Ranneby & Sara Sjöstedt. February 1992.