CURRENT QUARTER FORECASTS OF SWEDISH GNP USING MONTHLY VARIABLES

Stefan Gerlach

Svensk resumé på sid 22.
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 \( \Delta x_1 = \ln(x_{\text{Jan}} / x_{\text{Oct}}) \), \( \Delta x_2 = \ln(x_{\text{April}} / x_{\text{Jan}}) \), \( \Delta x_3 = \ln(x_{\text{July}} / x_{\text{April}}) \), and \( \Delta x_4 = \ln(x_{\text{Oct}} / x_{\text{July}}) \).

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

\[ \alpha(L) \epsilon_t = \beta(L) u_t \]

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 \( \Delta y_t \), denoted \( \hat{\Delta y}_t \), is given by

\[ \Delta \hat{y}_t = \hat{\gamma} + \hat{c} \Delta x_t + \hat{\epsilon}_t \]

where the expected value of \( \epsilon_t \) 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.\footnote{The data are all "day-corrected," but are not seasonally adjusted.}

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 (\(x_{3,t}\), which is reported with a 2 week lag).\footnote{We used real GNP in 1985 prices.} 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
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 $\Delta y_t$ on $\Delta x_t$ 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

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

and similarly for the $\Delta x_t$ 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

$$(4a) \quad \Delta y^*_t = \gamma + c\Delta x^*_t + \varepsilon_t$$

$$(4b) \quad \alpha(L)\alpha_s(L)e_t = \beta(L)\beta_s(L)u_t$$
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 $\Delta y_t$; 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 $\Delta y_t^*$ using only its own past values. That is, we estimate

\begin{equation}
(5a) \quad \Delta y_t^* = \gamma + \varepsilon_t
\end{equation}

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

\[(5b) \quad (1 + \alpha_1 L + \alpha_4 L^4) \varepsilon_t = (1 - \beta_4 L^4) u_t\]

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 heteroscedasticity. 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 \(t\) \((t>t)\) depends on the \(\Delta x_t\) relative to the mean of \(\Delta x_t\) 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

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

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

\begin{equation}
\Delta y_t^* = \gamma + c\Delta x_{1t} + \epsilon_t
\end{equation}

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

\begin{equation}
(1 + \alpha_1 L)\epsilon_t = (1 - \beta_2 L^2 - \beta_4 L^4)u_t
\end{equation}

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. 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, $\Delta x_{1t}$, the growth rate of electric energy production, $\Delta x_{2t}$, and the growth rate of sales, $\Delta x_{3t}$. 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

---

8Note that the forecasts are unconditional, since $\Delta y_t$ is known when the forecast of $\Delta y_t$ 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

\[(7a) \quad \Delta y_t^* = \gamma + c_1 \Delta x_{1t}^* + c_2 \Delta x_{2t}^* + c_3 \Delta x_{3t}^* + \epsilon_t\]

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

\[(7b) \quad (1 + \alpha_4 L^4)c_t = (1 - \beta_1 L - \beta_8 L^8)u_t\]

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. 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. 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)

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

so that the variance of the forecast error, $\sigma_f^2$, can be written

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

Thus, $\sigma_f^2$ depends both on the variance of $\epsilon_{\tau}$, and the precision with which $\gamma$ 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


TABLE 1

MODEL 1

\[ \Delta y_t^* = \gamma + \varepsilon_t \]
\[ (1 + \alpha_1 L + \alpha_4 L^4) \varepsilon_t = (1 - \beta_4 L^4) u_t \]

Estimated Coefficients (t-Statistics)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>0.00</td>
<td>(0.00)</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>0.31</td>
<td>(2.51)</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>-0.33</td>
<td>(1.98)</td>
</tr>
</tbody>
</table>

\( R^2 \) | 0.48 |
\( \text{Adj. } R^2 \) | 0.45 |
\( \text{SEE} \) | 0.015 |
\( \text{Loglik.} \) | 155.97 |

Marginal Significance Levels for Diagnostic Tests

- 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

\[ \Delta y_t^* = \gamma + c_1 \Delta x_{it} + \varepsilon_t \]
\[ (1 + \alpha_1 L) \varepsilon_t = (1 - \beta_2 L^2 - \beta_4 L^4) u_t \]

Estimated Coefficients (t-Statistics)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>0.00</td>
<td>(0.03)</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.56</td>
<td>(5.71)</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.31</td>
<td>(2.54)</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>-0.69</td>
<td>(6.55)</td>
</tr>
</tbody>
</table>

\( R^2 \) | 0.62 |
\( \text{Adj. } R^2 \) | 0.59 |
\( \text{SEE} \) | 0.013 |
\( \text{Loglik.} \) | 165.15 |

Marginal Significance Levels for Diagnostic Tests

- 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

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

$$(1 + \alpha_4 L^4) \varepsilon_t = (1 - \beta_1 L - \beta_2 L^2) u_t$$

<table>
<thead>
<tr>
<th>Estimated Coefficients (t-Statistics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$: 0.00 (0.11)</td>
</tr>
<tr>
<td>$c_2$: 0.07 (4.78)</td>
</tr>
<tr>
<td>$\alpha_4$: 0.69 (5.82)</td>
</tr>
<tr>
<td>$\beta_2$: -0.37 (2.37)</td>
</tr>
</tbody>
</table>

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

Marginal Significance Levels for Diagnostic Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch-Godfrey, 4 lags</td>
<td>0.80</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>0.50</td>
</tr>
<tr>
<td>Box-Pierce, 4 lags</td>
<td>0.96</td>
</tr>
<tr>
<td>Arch(4)</td>
<td>0.72</td>
</tr>
<tr>
<td>Ljung-Box, 4 lags</td>
<td>0.95</td>
</tr>
<tr>
<td>White</td>
<td>0.48</td>
</tr>
</tbody>
</table>
FIGURE 1: QUARTERLY GNP GROWTH

--- UNADJUSTED ----- SEASONALLY DIFFERENCED

FIGURE 2: QUARTERLY GROWTH OF INDUSTRIAL PROD.

--- UNADJUSTED ----- SEASONALLY DIFFERENCED
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

RMS = .0117

FIGURE 8: MODEL 1

--- GNP ------ PREDICTED GNP
FIGURE 9: MODEL 2

RMS = .0091

FIGURE 10: MODEL 2

--- GNP ---- PREDICTED GNP
FIGURE 11: MODEL 3

--- FORECAST ERROR ---- +2 SD ----- -2 SD

RMS = .0154

FIGURE 12: MODEL 3

--- GNP ----- PREDICTED GNP
Previous titles in this serie:


