Research and Development – Methodology reports from Statistics Sweden 2006:1

Quantifying the quality of macroeconomic variables

Lars-Erik Öller
Alex Teterukovsky

Statistiska centralbyrån   Statistics Sweden

The series entitled "**Research and Development – Methodology Reports from Statistics Sweden**" presents results from research activities within Statistics Sweden. The focus of the series is on development of methods and techniques for statistics production. Contributions from all departments of Statistics Sweden are published and papers can deal with a wide variety of methodological issues.

Research and Development – Methodology reports from Statistics Sweden 2006:1

Quantifying the quality of macroeconomic variables

Lars-Erik Öller
Alex Teterukovsky

Statistiska centralbyrån
2006

Research and Development – Methodology reports from Statistics Sweden 2006:1

Quantifying the quality of macroeconomic variables

Statistics Sweden
2006

Producer Statistics Sweden, Research and Development Department
SE-701 89 ÖREBRO
+ 46 19 17 60 00

Enquiries Thomas Laitila, + 46 19 17 62 18
thomas.laitila@scb.se

Cover: Ateljén SCB

ISSN 1653-7149

URN:NBN:SE:SCB-2006-X103OP0601_pdf

This publication is only published electronically on Statistics Sweden's website www.scb.se.

Preface

This report is the first issue of the new series entitled "Research and Development – Methodology Reports from Statistics Sweden". This new series contains a collection of work from Statistics Sweden combined in a single publication, replacing the earlier series "R&D Reports". Contributions from all the departments of Statistics Sweden concerning research and development of methodology for statistics production are included. Papers based on empirical results as well as formal theoretical derivations are of interest. The intended distribution is among researchers within and outside Statistics Sweden, and users of statistics. Hopefully this series can add to an increased exchange of knowledge of methods for official statistics production.

This first issue contains a report on a theoretically derived measure with applications to empirical data. Sound decision making utilizes information retrieved from relevant statistical data. A prerequisite for good decision making is high data quality. If data is of poor quality rational decision making can not be expected. One area where this quality aspect is of great importance is within the production of macroeconomic data. In this report the authors contribute with a proposal for a measure of quality of macroeconomic variables. It is shown to be realistically feasible and informative, and is a potentially useful tool for implementation within production processes.

Statistics Sweden, May 2006

Folke Carlsson

Gunnel Bengtsson

Disclaimer

Contents

Preface 3

Abstract 7

1 Introduction 9

2 A quality measure 13

2.1 Signal-to-noise ratio 13

2.2 The producer's information and the user's 15

2.3 Information Window 16

3 Three examples 19

4 Information theory approach 23

4.1 Theil's information measures 23

4.2 A new entropy-based measure 25

4.3 Testing the empirical distributions 29

5 Optimal properties of information measures 31

6 Conclusions 35

Acknowledgements 37

References 39

Abstract

Methods to quantify the quality of a macroeconomic statistical time series are presented. The measures are based on a combination of how predictable the series is and how much its statistics needs to be revised. An "information window" based on signal-to-noise ratios (SNR) provides a snapshot of the quality. A formulation of information in terms of entropy is considered. Our nonparametric approach is an attractive alternative to the parametric entropy estimation suggested by Theil (1966, 1967), since it allows for testing whether a forecast or a preliminary value is informative. Concavity and monotonous convergence of information accrual are discussed. Finally, we show how the suggested measures signal that either a given forecast or the macroeconomic variable itself is of dubious quality.

Keywords: statistical quality, forecast errors, revisions, information measures, entropy.

1 Introduction

Macroeconomics should become more useful if models on how the economy works and data tracing its performance improve. The philosophical abstractions could then be tested against reality in econometric models using empirical counterparts of theoretical variables. A better understanding could also be expected to lead to better forecasts. At least, this has been the dream of economists ever since the 18th century. But do we know today how well statistical time series really reflect what they are supposed to, in other words can we measure the user quality of e.g. the System of National Accounts (SNA)?

Even today, most experts would answer in the negative. Instead, data quality is generally expressed verbally in terms of *contents, accuracy, timeliness, comparability, availability* and *clarity*. It is difficult to turn these properties into numerical measures, and even harder to aggregate them into one global figure in a generally acceptable way. Here an attempt is made to express the quality as scalar measures. But instead of starting from the verbal criteria above, an implicit shortcut is chosen: measure how early and how accurately the value of a macroeconomic variable can be assessed with some certainty.

Contrary to the general view there are studies that endeavor to measure quality properties of macroeconomic variables. One is to calculate how accurate the forecasts are. If a policy maker has but a very faint idea about the value of a relevant variable next year, how is s/he to make a decision today that takes effect in that year? Research has focused on assessing the forecast accuracy of macro-economic variables (see e.g. Öller and Barot (2000)), especially since the *International Institute of Forecasters* was founded 25 years ago. A general discussion on macroeconomic forecast accuracy is Fildes and Stekler (2002). Forecast information contents is discussed in e.g. Öller (1985), Diebold and Kilian (2001) and Oke and Öller (1999).

Another numerical approach to data quality is to look at *revisions*. Assuming honest and diligent statisticians and an open revision policy, the number and amount of revisions reflect the reliability of a macroeconomic variable. If its value has to be fundamentally changed up to two years after the event, how trustworthy are then the first (and most important) figures published, and how relevant is the last ex post figure? Can anybody be expected to forecast a variable that keeps changing its value? A recent study on revisions of Swedish National Accounts expenditure variables, Öller and Hansson (2004)[1], reveals that revisions may be quite considerable. Following a line of research initiated by Mankiw and Shapiro (1986), Swanson and van Dijk (2006) study the *rationality* of revisions, another measure of ex post quality.

A third way to measure a quality aspect of an economic variable (in levels) is to analyze the values of successively published figures for common unit roots. The figures should have the same common trend, and this trend should be close to the trend of the final figures. Interesting results can be found in Siklos (1996) and Patterson and Heravi (2002) for US data, Patterson (2002) for UK data, and Öller and Hansson (2004) for Swedish data. This approach will not be further pursued here.

"The process of estimating GNP starts with forecasts made many years before a quarter has begun and continues for years after it has ended". These words by McNees (1989) are the starting point of the quality measures presented in this study[2]. An answer is given to the question: "How accurately can a statistical variable be estimated within a reasonable period of time?" If little is known about a variable *ex post* and even less *ex ante*, the series is of doubtful practical use. Here, quality is understood precisely as usefulness, an attribute that encompasses all the verbal quality desiderata listed above. Useful data help a person to understand and to act accordingly. The opposite means that the data leave the user in ignorance. We consider the ratio between forecast errors/revisions and the variability of the time series. It should be noted that both the numerator and the denominator in the ratio suffer from the fact that the correct outcome is unknown. Here, the final figure is taken to represent that value, but a hint is given for when this may or may not be a tenable assumption.

An early technique of using forecast errors and revisions to assess the quality of macroeconomic variables is presented in Theil (1966, 1967). There an information theory approach is chosen to measure the entropy reduction on the arrival of a new forecast or preliminary

---

[1] Studies of US National Accounts revisions include Morgenstern (1963), Stekler (1967, 1987), Young (1974, 1995), Mork (1987), De Leeuw (1990) and Fixler and Grimm (2002).

[2] For a stimulating discussion of the connection between "…data collectors, data users, econometric theorists and mathematical economists" see Hendry (2000, Ch. 1).

outcome. The entropy is (parametrically) defined in terms of the normal distribution. We shall present a nonparametric alternative to Theil's method, which allows for testing whether a forecast or a preliminary value is informative.

A 19th century economist had little numerical knowledge of the economy, simply because there were no SNA statistics, although many of the variables were defined as philosophical abstractions[3]. Efforts to record and model the activity of the economy should have improved our knowledge, but how informed are we today? Another question that can be answered by measures presented here is whether, in terms of quality, one variable is superior to another. We use the terms "quality" and "information" somewhat interchangeably meaning that high quality is synonymous to large information content.

In the next section our first quality measure is defined and explained. In Section 3 the measure is highlighted in three examples. Section 4 discusses entropy-based measures and introduces a test of information content. Other quality aspects are discussed in Section 5, and Section 6 concludes.

[3] Many economists worked on building a mathematical and empirical structure for economics similar to that of physics. Statistical data would have to substitute for experimental. SNA was eventually introduced internationally half a century ago, an accomplishment for which Sir Richard Stone was awarded the 1984 Prize in Economic Sciences in Memory of Alfred Nobel.

2 A quality measure

2.1 Signal-to-noise ratio

Good luck may sometimes result in a wild guess being correct, but without knowledge of the process it is not possible to consistently beat some naïve forecasting rule. In this sense forecast accuracy reflects how well the process is understood. The economic data are generated within a framework based on economic theory and the statistics are compiled accordingly. If the theory is inadequate, measurements become meaningless, and if the produced measurements are of poor quality, what may have been understood theoretically is not reflected in the data, and the empirical model collapses. Both deficiencies would lead to bad forecasts, but probably also to a measurement process that has to be repeated over and over again, with no guarantee of convergence. This process could also lend itself to a Bayesian interpretation, where theory, experience and historical data would form the prior, which is updated when new data accrue.

These arguments support the use of forecast errors and revisions to measure the quality of a statistical variable. The construction of such a measure is the next task.

Assume given a time series:

$$y_{lt}\,,\ l = 1,\ldots,L\ ;\ t = 1,\ldots,\ T,$$

where the measurements y_{lt} are successive estimates of an unknown variable η_t, where *t* stands for time (month, quarter or year). For each *t* the measurement is made *L* times. Following Croushore and Stark (2001), we call *l* a vintage. The values of the first vintages are forecasts, the subsequent ones are outcomes. The observation y_{1t} therefore is the first forecast of the value for period *t*. There is an observation $y_{\lambda t}$, $1 < \lambda < L$, called the preliminary figure, which is the first estimate (outcome) based on data from period *t*. The last estimate of period *t*, y_{Lt}, is called the final figure.

The estimation process is stopped at vintage *L*. Under the assumption that the vintages converge toward η_t, in empirical studies y_{Lt} substitutes for η_t. The accuracy of vintage *l* can be estimated by the Mean Squared Error (MSE):

$$MSE_l = \frac{1}{T}\sum_{t=1}^{T}\left(y_{lt} - y_{Lt}\right)^2 \, , \; l = 1,\ldots,L \, . \tag{1}$$

These values are then standardized by the variance of the final figures s_L^2 :

$$\overline{MSE}_l = \frac{MSE_l}{s_L^2} \, , \; l = 1,\ldots,L \, . \tag{2}$$

The estimate s_L^2 is assumed to be close to the true, but unknown variance of η_t . As in Theil's U^2, see Theil (1966, Ch. 2), a value of the standardized MSE_l at or above unity indicates a worthless forecast or preliminary figure, because the arithmetic average of historical values would be at least as accurate. In other words, for each t, y_{lt} could then be regarded as a white noise process for which the expected value is the optimal univariate forecast at any time.

We will study the revision history of a variable over time, as measured by

$$\overline{MSE}_1, \overline{MSE}_2, \ldots, \overline{MSE}_L \, , \tag{3}$$

where now by definition $\overline{MSE}_L = 0$ according to (1) and (2). A natural desideratum of convergence of vintage estimates will then be satisfied trivially. Moreover, one would wish the convergence to be monotonous, and fast in the beginning of the process. A non-monotonous convergence means that estimation errors become larger between two successive vintages. This could happen if the most recent data accumulating in the process are bad to the point of being misleading, so that a forecast or a preliminary figure is revised in the wrong direction. Fast convergence only at a late stage of the estimation process raises suspicion concerning the reliability of the final figure. If not even y_{Lt} were close to the unknown η_t, the present analysis breaks down, but then, trivially, we can state that nothing is known about the variable in question.

Let us denote the Signal-to-Noise-Ratio[4] (SNR) as I_l and define it as follows:

$$I_l = 1 - \frac{MSE_l}{s_L^2}, \ l = 1,...,L . \qquad (4)$$

This statistic estimates how much of the variance (uncertainty) of the variable has been accounted for by vintage *l*. It has an analogue in the degree of determination R^2 in regression analysis.

Forecast errors and revisions may for some vintages be so large as to exceed the variance of the variable itself, making the corresponding I_l smaller than zero. This would be a hint for either completely changing the ways the forecasts/outcomes are produced, or for refraining from producing the vintage altogether. The same message can be said to be delivered by $\overline{MSE}_l = 1$. In the sequel unity is substituted for $\overline{MSE}_l > 1$, so that for all vintages $0 \le I_l \le 1$, as is the case with R^2. The higher the I_l, the less uncertainty (the more information) there is in vintage *l*.

2.2 The producer's information and the user's

We assume that the information increase from vintage to vintage is approximately linear for the producer of forecasts or statistics and zero for the user. That is, the user's knowledge of the variable increases stepwise at publication times. The information increase at the time of the very first vintage is treated separately, assuming that both the user and the producer are totally ignorant (possessing the same zero information) about the variable up to the first vintage. This assumption is plausible since we (the evaluators) have no data prior to the first publication, and therefore could not reason differently.

Having established that, we proceed to graphical and numerical illustrations of the information measures.

[4] This term is best known in signal theory. Suppose the signal comes from a distant location in the Universe and is corrupted by noise on its way to the Earth. The closer an object is the less noise. Here, the outcome η_t, or actually, its empirical counterpart y_{Lt}, is distant in time, which as space, corrupts the signal.

2.3 Information Window

Figure 1 shows how the producers's information about η_t changes as time passes and more data accumulate. The information increases linearly from vintage to vintage.

Denote the white area under the curve as the integrated SNR measure II. The value of II is calculated using numerical integration:

$$II = \frac{1}{2}\sum_{l=1}^{L-1}(I_l + I_{l+1})\tau(l,l+1)$$

where $\tau(l,l+1)$ is the length of the time interval between vintages l and $l+1$. Without loss of generality we can normalize these intervals to sum to unity. Then both the vertical and the horizontal axes of the diagram in Figure 1 are of unit length, and the overall measure of quality, II will be between $\tau(l,l+1)/2$ and 1. Here the lower limit is reached when total ignorance prevails until the next to the final figure is released.

From the user's viewpoint the diagram looks differently (Figure 2). The integrated SNR measure II^u (where superscript u stands for “user”) would be between 0 and 1 and calculated as

$$II^u = \sum_{l=1}^{L-1}(I_l + I_{l+1})\tau(l,l+1)$$

It is easy to visualize the two extreme cases - the totally black square (which can happen only in the user's case) and the totally white one if the first forecast and all successive estimates would be identical to the final figure. We call the squares in Figures 1 and 2 Information Windows. Henceforth, we are going to use the terms “Information Window” and “integrated SNR measure” as in Figure 1, i.e. from the producer's point of view.

Figure 1
Producer's information

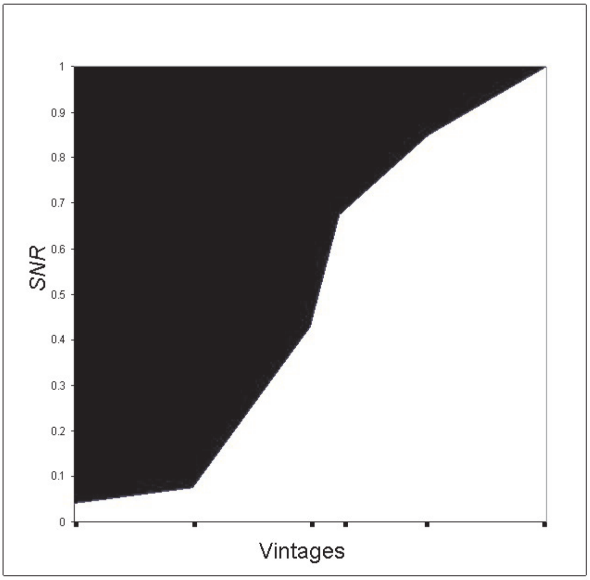
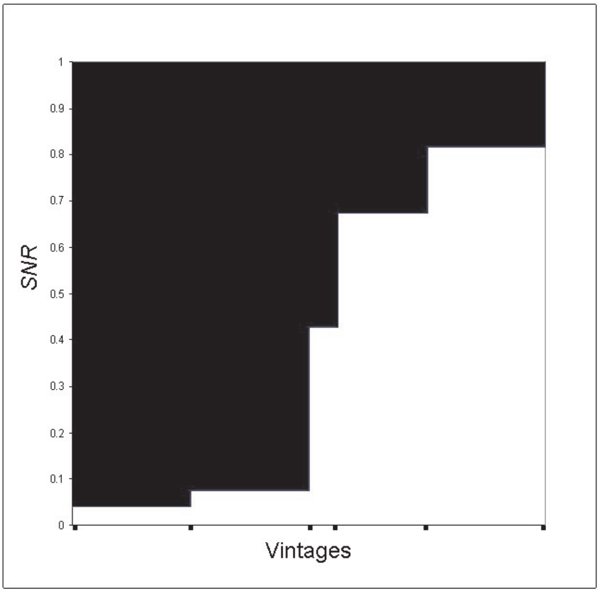


Figure 2
User's information



3 Three examples

In order to illustrate the new concepts, three variables are chosen from the annual Swedish SNA: Private Consumption, Central Government Consumption and Exports. They are measured as annual growth rates, assumed to be stationary, and cover the period 1980-2002. The forecasts are made by the National Institute of Economic Research, Sweden, in December[5] years t - 2, t - 1 and t. The subsequent outcome figures, published by Statistics Sweden, are from March (t + ¼) and December year t + 1 and December t + 2. Hence the number of vintages $L = 6$, and intervals are not equidistant. During these four years a user should be able to form a good idea of the numerical value of the variable in question. If not, it is doubtful whether there is any point in defining such a variable and compiling statistics on it, not to mention forecasting.

Figures 3-5 present the Information Windows of the three variables. The lighter the window, the better the information. As was explained before, zero has been substituted for negative SNRs. The curves start in t - 2 when the first forecast for year t becomes available and end in December t + 2. The right border of the grey triangular area marks when the vintages become statistically significantly different from a naïve estimate. The corresponding test is explained in 4.3. The black area can be seen as a curtain, and the grey as a veil. The integrated information measures *II*, i.e. the areas underneath the black curtains, are given in the captions and also in Table 1.

Note that for all variables the curves start at zero in t - 2, at the time of the first forecast. This means that there was no point in forecasting the variables two years ahead; the mean would have done the same, or even a better job. However, note that the first vintage partially consists of imputed data. In Figure 3 the curve subsequently rises fast, with a distinct kink where the preliminary figure succeeds the last forecast.

[5] Before 1986, only forecasts made in September are available. Two-year-ahead forecasts start in 1990. Earlier values have been imputed by sampling from the normal distribution.

Figure 3
Private Consumption. SNR measure *II* = 0.69.

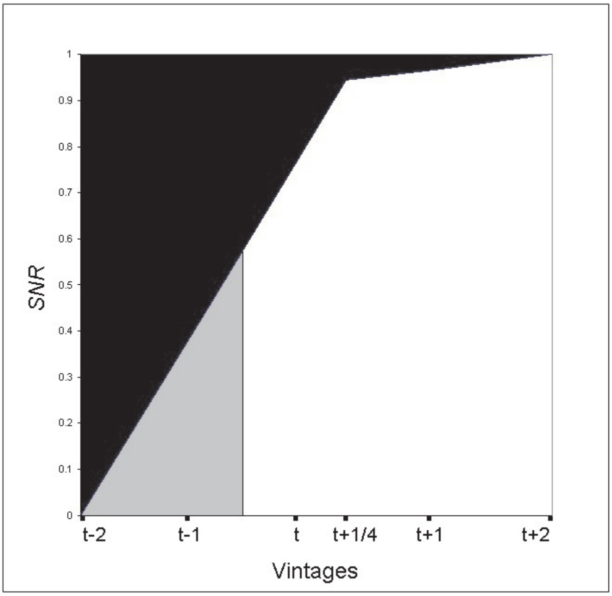


Figure 4
Central Government Consumption. SNR measure *II* = 0.31

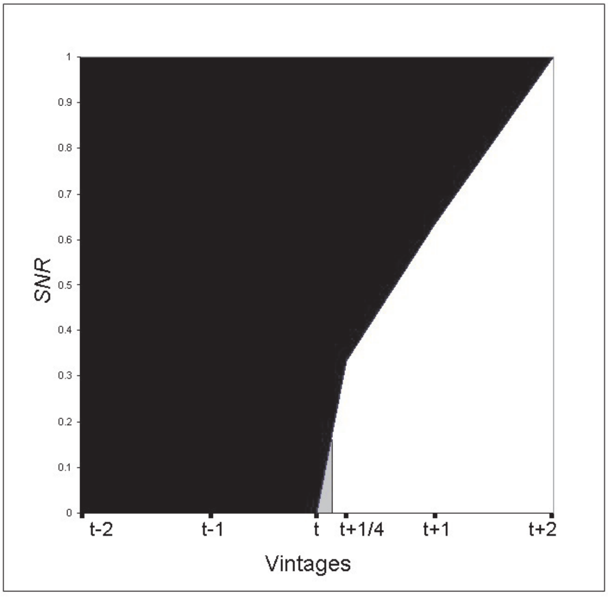


Figure 5
Exports. SNR measure *II* = 0.72

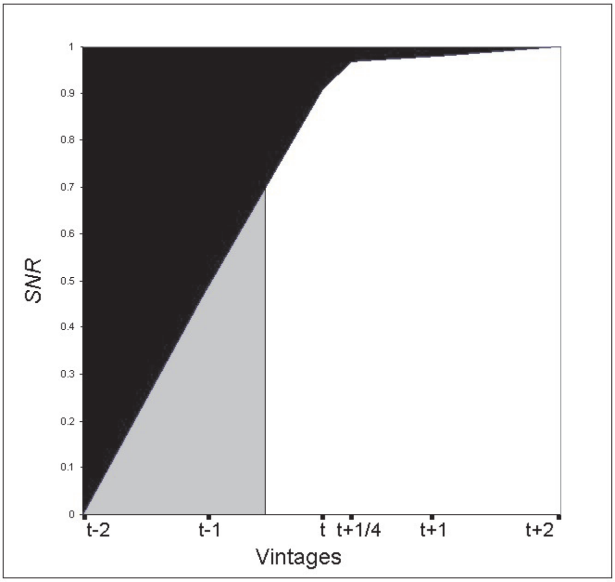


Figure 4 looks very different. Central Government Consumption has a SNR equal to (or below) zero for all forecasts. Remarkably it stays at zero till December year *t* when statistical data are available for ¾ of the year! A probable reason to this undesirable behavior is that the preliminary (quarterly) data are so bad as to mislead the forecaster to revise in the wrong direction. Equally well s/he could have stuck to the first (naïve) forecast from *t* - 2 all the way to year *t*. In fact looking at the early outcomes we see that they require substantial revisions. This comes close to making the entire statistical variable almost meaningless. If so large revisions are needed all the way to *t* + 2, how do we know that the need for revision stops there?

The most informative of the three variables is Exports (Figure 5). The reason could be that this variable is relatively easy to define as outbound goods and services passing the national border. The statistical measurements are also quite accurate, especially since Eurostat introduced the Intrastat system, checking outbound streams against the corresponding incoming streams of goods.

Table 1
Comparing variables

Measures	Priv. Consumption	Centr. Gov. Consumption	Exports
II	0.69	0.31	0.72
Theil's A^2	2.8	21.4	9.8
IH	0.45	0.31	0.48

4 Information theory approach

4.1 Theil's information measures

An early attempt at measuring the information content of macroeconomic variables, which, however, has not become standard in the literature, is Henri Theil's information measures in Theil (1966, 1967). His setup was the same as ours with forecasts followed by outcomes, and measurements based on forecast errors and revisions. He suggests a parametric decomposition of the forecast variance into three multiplicative components: variable, time and vintage:

$$\sigma_{itl}^2 = A_i^2 B_t^2 C_l^2 \tag{5}$$

Here σ_{itl}^2 is the variance of the i^{th} variable for the t^{th} year at vintage l. A_i measures the inaccuracy corresponding to the i^{th} variable, B_t - the inaccuracy corresponding to the t^{th} year (forecast errors and revisions decreasing or increasing over time) and C_l - the inaccuracy of vintage l. The components are estimated iteratively from the data with starting values chosen to reflect our prior expectations, e.g. of decreasing uncertainty through vintages. For details we refer to Theil (1966, Ch. 7).

A well-known information measure is entropy (originating from thermodynamics and statistical mechanics). Theil found this concept useful for describing the reliability of forecasts and preliminary estimates of macroeconomic variables. Entropy quantifies the degree of uncertainty or ignorance in the data.

If a variable takes values on the real axis according to a continuous probability density distribution $f(\cdot)$, then its entropy H_f can be defined as[6]

$$H_f = -\int f(x) \ln f(x) dx$$

[6] This is Shannon's definition. For other mathematical forms, see e.g. Esteban and Morales (1995).

Among all distributions with an infinite range of variation and with a given variance, the normal distribution is the one with the largest entropy. For any normal distribution with variance σ^2 the entropy is

$$H_N = \tfrac{1}{2}\ln(2\pi e\sigma^2) \qquad (6)$$

where *e* is Neper's number. We see that for a normal distribution the entropy is a linear function of the logarithm of the variance.

Theil assumes that the forecast errors/revisions are normally distributed. He gives the decrease in uncertainty (entropy) about the final value y_{itL} of the i^{th} variable when a new vintage $y_{it,l+1}$ arrives as:

$$H(y_{itL} \mid y_{itl}) - H(y_{itL} \mid y_{it,l+1}) = \ln\frac{C_l}{C_{l+1}} \qquad (7)$$

which follows from (5) and (6). This decrease in entropy is also called the information gain, which at the last vintage is equal to infinity if we assume that the last vintage is the truth. We see that the information gains are independent of both *t* and the variable *i*, and represent the average vintage information gains across variables and years. Therefore, (7) is a multivariate measure.

We followed Theil and calculated the A^2, B^2 and C^2 components for our data. The estimates of A^2's are presented in Table 1. The A^2 effect can be considered an alternative to our *II*. We see that the relative order of the variables according to A^2 is somewhat different compared to *II*.

Table 2
Theil's information gains (per month)

Vintages	Priv. Consumption	Centr. Gov. Consumption	Exports	Across Variables
t – 1	0.03	-0.01	0.02	0.02
t	-0.01	0.11	-0.02	0.08
t + ¼	0.32	0.38	0.29	0.37
t + 1	0.08	0.07	0.13	0.10
t + 2	∞	∞	∞	∞

One can treat each variable separately and decompose the total error variance in two components - B_t^2 accounting for time and C_l^2 - for vintage. This procedure essentially incorporates the variance of the variable into the variances of vintages and years. Table 2 presents gains according to (7) for each variable and across variables. Since the intervals are not equidistant, in order to make the information gains comparable we present the values per month, i.e. divided by the number of months between two publications. The information gain for the first vintage is impossible to calculate per month. The last information gain is equal to infinity since we assume that the last vintage is error-free (i.e. has zero variance). Note that I_l by definition circumvents these complications. From Table 2 we see that by far the largest information gain is achieved for the first preliminary figures, vintage $t + ¼$.

In order to compare the different variables, Theil pools them which requires identical publication times and time spans. The more variables one compares (in particular, internationally), the less likely it is that this holds in reality.

Theil's method is based on restrictive assumptions on normality and proportionality of variances. The *II* measure does not imply any such assumptions and is more straightforward to implement and interpret. The advantages of Theil's measures are that the entropy is a well-known information measure and that one gets numerical expressions for different sources of uncertainty, also for a group of macroeconomic variables. On the other hand, as with many parametric approaches, these measures are most effective when data are abundant, but become vulnerable in real-life situations with short time-series. Another disadvantage is that Theil's measures are not bounded, as is *II*. Furthermore, just as *II*, they are not readily testable (e.g. against a naïve alternative), because the vintages are dependent. Otherwise, e.g. an F test of successive mean square errors against the variance of the final figures would have been a natural choice, showing where the information becomes significant. Below we suggest a solution to these problems.

4.2 A new entropy-based measure

As was mentioned, we assume that prior to the first forecast all we know is the historical outcomes. Consider the normal distribution with mean and variance calculated from the historical outcomes (taken at $t + 2$). Note that this is the variance we used to construct our I_l in (4). Subtract the mean to obtain the error distribution $N(0, s_L^2)$, which we denote $\phi(y_L)$. It represents the absolute minimum of information and we call it the “total ignorance” distribution. The normal distribution is the one with largest entropy among all infinite-range distributions with a given variance, and the variance of the historical outcomes is a natural choice of a benchmark.

Divide the real axis into K intervals so that the probability of an observation distributed according to $\phi(y_L)$ falling into any of the intervals is $1/K$. Now, for each variable and each vintage l, consider the following discrete entropy:

$$H_l = -\sum_{k=1}^{K} f_{lk} \ln f_{lk} \tag{8}$$

where f_{lk} is the observed relative frequency of forecast errors/revisions for vintage l falling into interval k.

The entropy in (8) achieves its maximum ln$K$ when all $f_{lk} = 1/K$, $k = 1, 2,\ldots,K$. This maximum corresponds to the "total ignorance" distribution, which one obtains by constantly making the naïve historical mean forecasts. The minimum of (8) is zero and is achieved when all observed data fall into any single interval. Consequently, as with Theil's measures, our entropy regards variance as the single source of uncertainty. The choice between measures I_l and H_l depends on how one regards bias. If one has a clear picture of bias then it does not contribute to the uncertainty! If not, then studying both I_l and H_l may help. For example, a high value of H_l together with a low value of I_l would indicate a small variance in the forecast errors but a high bias.

As for K, an odd number is recommended in order to center one interval at zero. K should be kept moderate. Our data spanned over 23 years and we chose $K = 5$, which is the first odd number larger than the values suggested (for the chi-square test of fit) by the "minimum five in each interval" rule of thumb and the rule $K = 0.9\times23^{0.4}$ (see e.g. Koehler and Gan (1990) and Boero, Smith and Wallis (2005)).

With very few observations, H_l becomes sensitive not only to K but also to single observations; I_l is more stable in this respect. Moreover, I_l reacts to outliers, whereas for H_l the following can happen. From (8) we note that, if all errors but one fall into the same interval, the measure pays no attention to how distant the interval containing that single error is. With a large number of observations, H_l becomes less sensitive but its extreme values (0 and 1) are almost never attained. This can give a false impression of information content when there is most likely none (as for the first vintages in Figures 4 and 7), or vice versa.

Figures 6-8 show the entropy-based information windows corresponding to those in Section 3. Again, we integrate the area under the curve to get the integrated entropy-based measure *IH* (similar to *II* in Section 4) that also are within the unit interval (see Table 1).

It should be mentioned that H_l is in fact linearly related to the Kullback-Leibler distance between the observed distribution of errors and the uniform distribution on the equiprobable intervals of the benchmark distribution $N(0, s_L^2)$.

Figure 6
Private Consumption. Entropy measure *IH* = 0.45

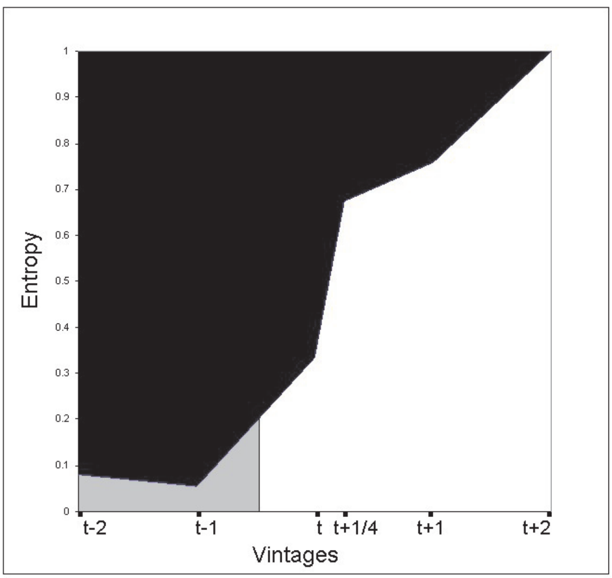


Figure 7
Central Government Consumption. Entropy measure *IH* = 0.31

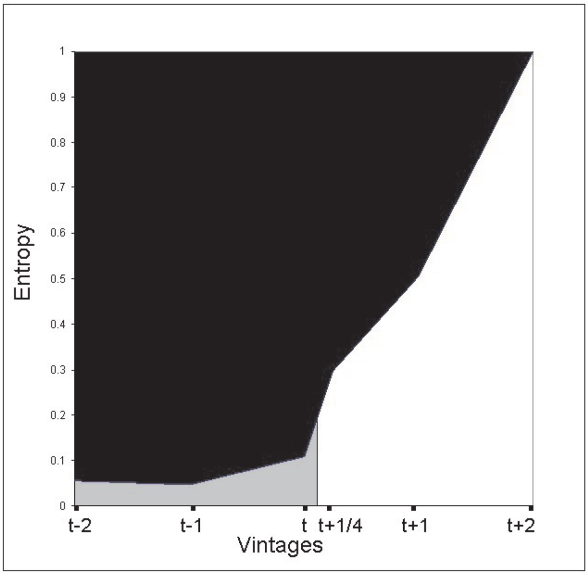


Figure 8
Exports. Entropy measure *IH* = 0.48

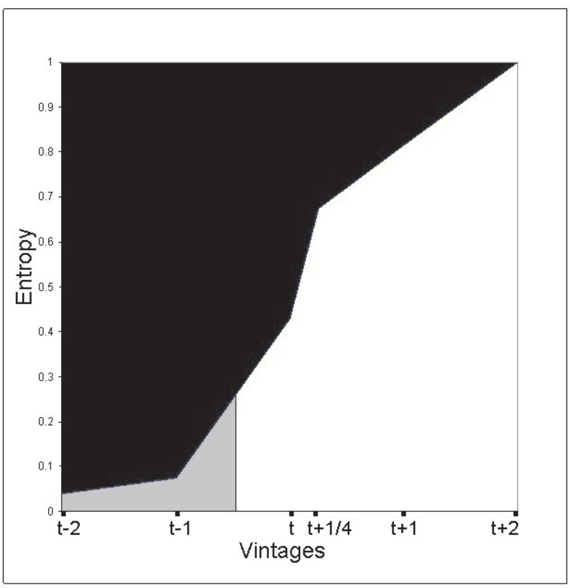


Table 3
Anderson-Darling test p-values (* marks significance at the 1% level)

Vintages	Priv. Consumption	Centr. Gov. Consumption	Exports
t – 2	0.250	0.250	0.140
t – 1	0.250	0.250	0.231
t	0.008*	0.064	0.004*
t + ¼	0.002	0.009*	0.001
t + 1	0.001	0.006	0.001

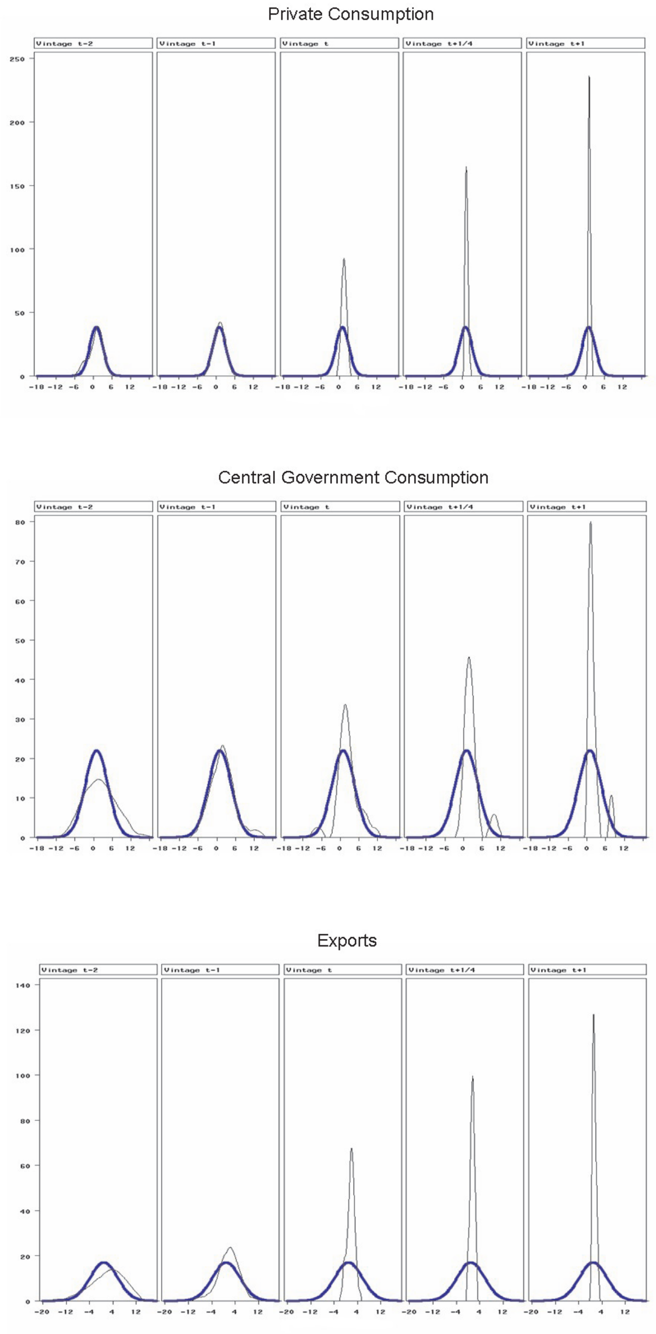
4.3 Testing the empirical distributions

The next step is testing the data against the "total ignorance" hypothesis. Table 3 presents the p-values of the nonparametric Anderson-Darling test for all variables and vintages. The right end of the grey triangular area in each of the Figures 3-5 and Figures 6-8 marks the midpoint between the last vintage which is not significant and the first significant one.

One may also choose the Mariano-Diebold or Granger-Newbold tests, which can be applied on both forecasts and outcomes. The choice of a particular test should be based on size and power considerations.

In Figure 9 we have plotted both the smoothed (by a quadratic kernel) histograms of forecast/revision empirical distributions for each variable and each vintage of the data and the density function of $\phi(y_L)$. The graphs show how fast the information increases from vintage to vintage.

Figure 9
Smoothed vintage distributions vs. ”total ignorance” (bold) distribution



5 Optimal properties of information measures

Let us consider the Information Window in Figure 1. It is indeed desirable that the information is monotonously increasing across vintages. Moreover, since η_t is not perceived as deterministic (which would correspond to a totally light window), one has to make do with the following optimality rule: "as much information as early as possible". The optimality rule can be put in a strict form. One has to cope with the costs of producing the information and will be able to improve the vintages only as long as the costs do not exceed what the user is prepared to pay. Hence we have a classical dynamic optimization problem of equating marginal cost to marginal information gain. This is the first order condition. The second order condition is that the marginal information gain should be diminishing, assuming that increasing the efforts will produce less and less improvement. This would result in concave curves. The steepness of the curve would be determined by the functions of the first order condition.

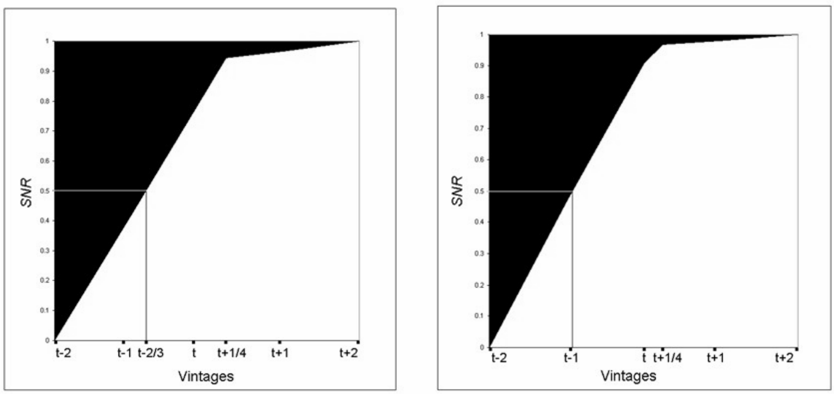
Concavity can be tested nonparametrically as suggested by Abrevaya and Jiang (2005). The test is based on the number of three-tuples that can be formed by the data points, which are classified according to concavity or convexity. Unfortunately, our data comprise only six vintages, which is not enough for a meaningful application of the test.

We can improve our understanding of the forecasting/revision process and tackle monotonicity and concavity by answering the following two questions:

1) How early does information irrevocably exceed a certain information level?
2) When has the forecasting/revision process accumulated a certain share of all information there will eventually be about the variable?

The questions relate to the problem of comparing two or more variables, especially if they have more or less the same *II* or *IH* values. A case in point would be Private Consumption (*II* = 0.69; *IH* = 0.45) and Exports (*II* = 0.72; *IH* = 0.48). One could then choose, for instance, point 0.5 on the vertical axis in the Information Window

Figure 10
Private consumption (left) vs. Exports (right). Vintages when II = 0.5 is achieved



and project it to the vintage axis (Figure 10). Then for Private Consumption, $I_l = 0.5$ yields $l = t - 2/3$ and for Exports $l = t - 1$. Hence, also in this respect the quality (timeliness) of Exports is better than that of Private Consumption.

The answer to the second question suggests an analysis of the area under the curve in the Information Window. Recall that the normalization in 2.3 resulted in a unity window so that the first vintage can be said to arrive at moment 0, and the last vintage at moment 1. Then in case of linear increase of information from zero in the first vintage to unity in the last one, a vertical gravity axis of the area under the curve (dividing the area by half) would cross the horizontal axis at $1/\sqrt{2} = 0.71$ (Figure 11a). If the first vintage were considerably different from zero and, say, equal to Q, the threshold value V could easily be found from the following equation (Figure 11b):

$$2(1-Q)V^2 + 4QV - (1+Q) = 0$$

The vertical gravity axis of the area under the curve to the left of the threshold signifies that the process gathers information earlier, i.e. concavity (Figure 12a), whereas a gravity axis to the right of the threshold indicates the opposite, i.e. convexity (Figure 12b).

Figure 11 Gravity axis of the area under the information curve. Linear growth from a) 0 and b) Q

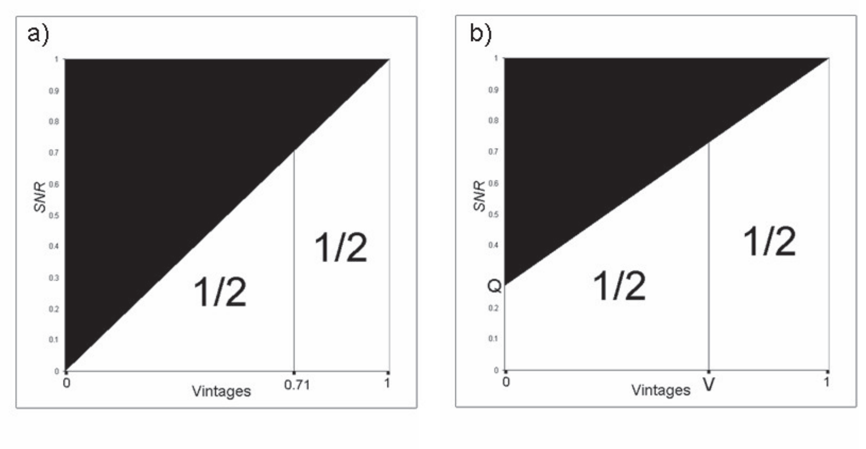
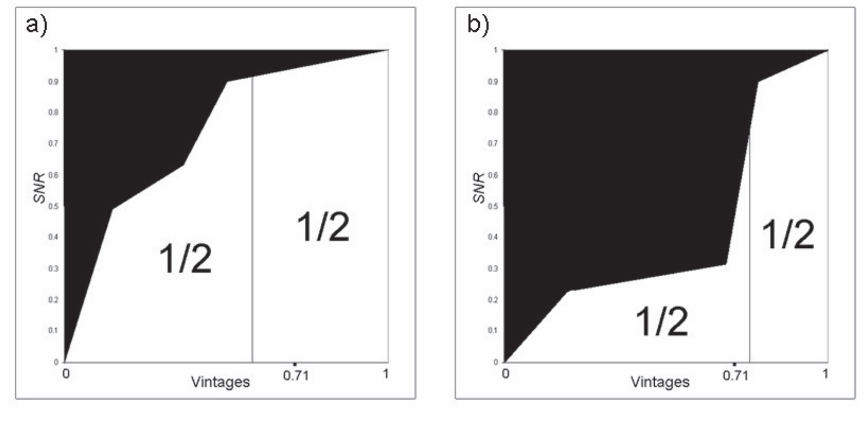


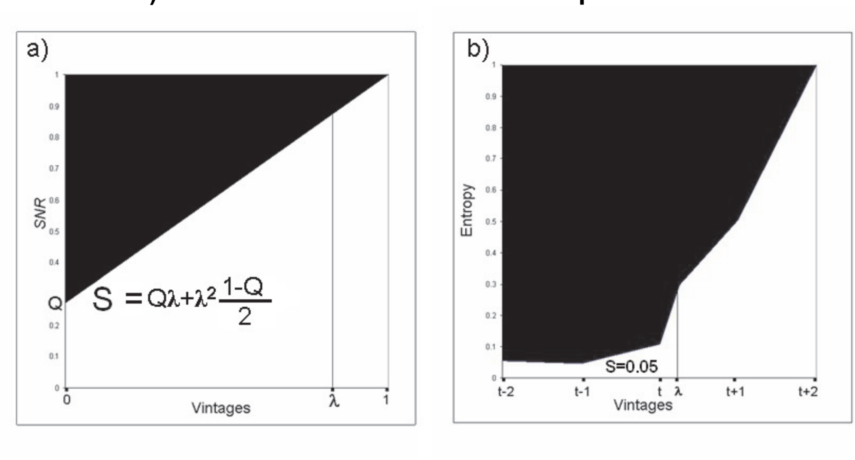
Figure 12: a) Concave information growth. b) Convex information growth



Finally, consider the vintage λ (between 0 and 1), which is the first preliminary figure (Figure 13a). In case of linear information growth from Q at moment 0 to 1 at moment 1, the area S under the curve up to λ is equal to

$$S = Q\lambda + \lambda^2 \frac{1-Q}{2}$$

Figure 13
Area under the information curve up to vintage λ a) in case of linear growth and b) for Central Government Consumption



An area smaller than S would signify that the forecasting process has been economically inefficient. Let us take H_l for the Central Government Consumption (Figure 13b). From (9) we obtain (for Q = 0.05 and λ = 0.5625) S = 0.18. The area under the curve up to vintage λ in Figure 13b is equal to 0.05, i.e. smaller than 0.18, which confirms the inefficiency of the forecasting process for this variable.

6 Conclusions

A number of measures have been discussed, some old and some new, that reflect the quality of macroeconomic variables. A graphical device, the Information Window, shows in a simple way how information about a variable develops across vintages and how much information there is. The more light from the window, the better is the quality of the variable. The window can be used for deciding if it is meaningful to forecast a variable a certain amount of steps ahead, or to statistically record it at all. Both forecast errors and revisions were included in the measure, but with a minor change of definitions, the measures could be applied to either of the two.

We saw that according to the Information Windows in Section 3, none of the variables were informative at t - 2. Why then make such forecasts? Two years is not a long time horizon and hence for planning purposes concrete figures are needed. But then one should just rely on the long-term average growth, thereby saving the considerable costs of a serious forecasting process.

The Information Window is based on well-known concepts of the SNR or entropy, and takes the form of a unit square. This simple graph can be added as meta data when publishing forecasts and outcomes.

Defining information in terms of probability distributions allows for testing at what stage in the estimation process the information becomes significant, as compared to a naïve alternative.

What happens if the statistical data producer never revises a preliminary figure? The first outcome then becomes the final one, and therefore the measure seems to quantify only the quality of forecasts. But since the first figure is often shaky, its uncertainty is passed on to the forecasters and adds to the forecasters' own uncertainty. More generally, even if there are revisions, the forecasters sometimes aim at preliminary figures. In this case a part of the forecast error similarly can be attributed to data inaccuracy (see Öller and Barot (2000)). Our information measures are still applicable.

One should keep in mind the recurring changes in definitions of the macroeconomic variables. In order to cope with this problem, it is important to take as final value a vintage that is not too distant from t.

Our measures reflect the quality as observed over a time interval. An innovation leading to a substantial quality improvement, will ultimately be detected by recurrent application of the introduced methods. Thus they quantify a quality improvement, possibly already intuitively recognized by the public at large.

Acknowledgements

This paper was presented at the workshop "The Future of Economic Forecasting", December 16-17, 2005 in Leipzig, Germany. The authors are grateful for valuable comments from the participants. Special thanks go to Michael Clements and Markus Jäntti. All shortcomings are ours.

References

Abrevaya, J. & Jiang, W. (2005). A nonparametric approach to measuring and testing curvature. *Journal of Business and Economic Statistics*, 23, 1-19.

Boero, G., Smith, J. & Wallis, K. (2004) The sensitivity of chi-squared goodness-of-fit tests to the partitioning of data. *Econometric Reviews,* 23, 341-370.

Croushore, D. & Stark, T. (2001). A real-time data-set for macroeconomists. *Journal of Econometrics*, 195, 11-130.

De Leeuw, F. (1990). The reliability of U.S. Gross National Product. *Journal of Business and Economic Statistics*, 8, 191-203.

Diebold, F. & Kilian, L. (2001). Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics,* 16, 657-669.

Esteban, M. & Morales, B. (1995). A summary on entropy statistics. *Kybernetika*, 31, 337-346.

Fildes, R. & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, 24, 435-468.

Fixler, D. & Grimm, B. (2002). Reliability of GDP and related NIPA estimates. *Survey of Current Business*, January, 9-27.

Hendry, D. F. (2000). *Econometrics, Alchemy and Science*. Oxford University Press.

Koehler, K. & Gan, F. (1990). Chi-squared goodness-of-fit tests: cell selection and power. *Communications in Statistics - Simulation and computation*, 19, 1265-1278.

Mankiw, N. & Shapiro, M. (1986). News or noise: an analysis of GNP revisions. *Survey of Current Business*, 66, 20-25.

McNees, S. (1989). Forecasts and actuals: the trade-offs between timeliness and accuracy. *International Journal of Forecasting*, 5, 409-416.

Morgenstern, O. (1963). *On the Accuracy of Economic Observations*. Second edition. Princeton University Press.

Mork, K. (1987). Ain't behavin': Forecast errors and measurement errors in early GNP estimates. *Journal of Business and Economic Statistics*, 5, 165-17.

Oke, T. & Öller, L.-E. (1999). Testing for short memory in a VARMA process. *Journal of Forecasting*, 18, 477-487.

Öller, L.-E. (1985). How far can changes in business activity be forecasted? *International Journal of Forecasting*, 1, 135-141.

Öller, L.-E. (2005). *An ignorance measure of macroeconomic variables*. (*Background Facts on Economic Statistics 2005:1).* Statistics Sweden.

Öller, L.-E. & Barot, B. (2000). The accuracy of European growth and inflation forecasts. *International Journal of Forecasting*, 18, 293-315.

Quantifying the quality of macroeconomic variables

This report proposes simple numerical quality measures for macroeconomic statistical variables. The measures are based on forecast errors and revisions, and build on a similar measure introduced in Öller (2005), on a forecast accuracy study, Öller and Barot (2000), on a study of revisions, Öller and Hansson (2004), and on earlier work by Theil (1966, 1967). The measures provide numerical devices to assess how reliable and informative a statistical variable is.

The report can also be seen as a continuation of the work on statistical quality published in Statistics Sweden (2001), as well as being a complement to a current project on statistical quality.

ISSN 1653-7149

Publikationstjänsten:
E-post: publ@scb.se, tfn 019-17 68 00, fax 019-17 64 44. Postadress: 701 89 Örebro.

Information och bibliotek: E-post: information@scb.se, tfn 08-506 948 01, fax 08-506 948 99,
Försäljning över disk, besöksadress: Biblioteket, Karlavägen 100, Stockholm.

Publication Services:
E-mail: publ@scb.se, phone +46 19 17 68 00, fax +46 19 17 64 44, Address: SE-701 89 Örebro.

Information and Library: E-mail: information@scb.se, phone +46 8 506 948 01, fax +46 506 948 99.
Over-the-counter sales on our premises: Statistics Sweden, Library, Karlavägen 100, Stockholm, Sweden.