An Introduction to Forecasting*

Hüseyin Akkaya, Thomas Fleischhacker and Paul Legerer

March 24, 2004

Abstract

This paper summarizes the first chapter of Forecasting Economic Time Series by Clements and Hendry. After a brief history of forecasting, a framework for forecasting is presented together with alternative methods. Finally we present an example based on an artificial data sample.

1 Background

Clements/Hendry [1] address the problems facing a theory of macroeconomic forecasting based on empirical econometric models.

The forecasting models are assumed to be correctly specified. In models “hypothetical” states must capture appropriate aspects of the real world. Nevertheless many analyzes if economic forecasting have been based on the assumptions of a constant, time-invariant, data generating process (DGP), that is stationary (non-integrated), and coincides with the econometric model used.

DGP rules out the evolution of the economy brought about by intermittent structural changes and regime shifts. Such shifts call into question the standard results on optimal forecasting. When the process has undergone a regime shift, forecasts based on past information need not be unbiased despite being the previous conditional expectation.

It is assumed that the consequences for forecasting of the DGP being an integrated-cointegrated system rather than a stationary vector process. The

---

*Paper submitted in partial fulfilment of the requirements of the course Ökonometrische Prognose by Prof R. M. Kunst, University of Vienna, summer term 2004. The authors thank Mathias Kühnen for his technical support.
properties of forecast-error variances for integrated series can differ notably from those of stationary variables, so the behavior of forecast-error variances depends on the choice of data transform examined, namely integrated levels, stationary differences or cointegrated combination.

Dropping the third assumption coinciding of forecasting with the econometric model leads to a discussion of the nature of empirical econometric models and of possible tensions between choosing models for forecasting, as against goodness of fit or policy analysis.

2 A History of the Theory of Economic Forecasting

Morgenstern\footnote{A list of references of the authors mentioned can be found in Clements/Hendry.} made the first comprehensive treatise for the methodology of economic forecasting in the late 1920’s. He argued against the possibility of economic and business forecasting in principle. His first argument is that economic forecasting could not rely in probability reasoning because economic data are neither homogeneous nor independently distributed 'as required for statistical properties'. Samples or the size which satisfied the assumptions would be too small to be usable.

Person held a similar way in 1925. The economists at the time had been making increasing use of statistical methods and statistical methods and statistical inference to test economic theories against the data. Standard errors and multiple correlation coefficients were sometimes given as gauge of the validity of the relationship.

Morgenstern raised also the problem of forecasts would be invalidated by agents reactions to them (Lucas Critique, in the form of self-fulfilling expectations, and was an issue which concerned many business -cycle analysts of the period). Morgenstern foresaw dangers in the use of forecasting for stabilization and social control.

Marget who interpreted him, gives a different conclusion than him, since most economic forecasting from statistical depended on extrapolating previous patterns, rather than on probability-based forecasting techniques (but extrapolating is useful to derive the relevant pattern). Also he argued that if a causal explanation were possible, even if not exact or entirely regular,
forecasting should be feasible. This presupposes that causation implies predictability then a potential counterexample would be the theory of evolution. He insisted on regularity between causes and their effects.

Morgenstern’s problem of bandwagon feedback effects is denied by Marget, assumed as either specific to economics or that are precluded by accurate forecasting.

Finally Marget argued for using forecasting to test economic theories.

Tinbergen developed in 1939 firstly forecasting tests for the models (in response to Keynes, as a criticism).

Haavelmo (1944, accepted widely in US by the end of 40s) treated forecasting as an application of his methodology (as a probability statement about the location of a sample point not yet observed). Given the joint data density function, there were no additional problems specifically associated with prediction. His approach:

Suppose \((x_1, \ldots, x_T)\) observable values on random variable \(X\), from which to predict the future values \((x_{T+1}, \ldots, x_{T+H})\). \(D_X(\cdot)\) is the joint probability of the observed and future \(x_s\) (assumed to be known). We denote the probability distribution of the future \(x_s\) conditional on past \(x_s\) by:

\[
D_{X_2|X_1}(x_{T+1}, \ldots, x_{T+H} | x_1, \ldots, x_T)
\]

and the probability distribution of the observed \(x_s\) as \(D_{X_1}(\cdot)\), and factorizing it, we get:

\[
D_X(x_1, \ldots, x_{T+H}) = D_{X_2|X_1}(x_{T+1}, \ldots, x_{T+H} | x_1, \ldots, x_T) \times D_{X_1}(x_1, \ldots, x_T)
\]

For a realization of the observed \(x_s\), denoted by the vector \(x_T^1 \in \mathbb{R}^T\), we can calculate from \(D_{X_2|X_1}(\cdot)\) the probability that the future \(x_s\) will lie in any set \(E_2 \subseteq \mathbb{R}^H\). Conversely, for a given probability we can calculate regions of prediction. Thus the problem of prediction is merely a problem of probability calculus.

In practice \(D_{X_2|X_1}(\cdot)\) must be estimated on \(x_T^1\), which requires the basic assumption that; The probability law \(D_X(\cdot)\) of the \(T + H\) variables \((x_1, \ldots, x_{T+H})\) is of such a type that the specification of \(D_{X_1}(\cdot)\) implies the complete specification of \(D_X(\cdot)\) and, therefore, of \(D_{X_2|X_1}(\cdot)\).

The first macroeconomic model is built by Tinbergen in 1930. His first
model had a reaction from Keynes. Keynes position was that if the data analysis failed to confirm the theory, then blame the data and statistical methods employed. The influence of Tinbergen ensured that official Dutch economic forecasting was more model-oriented than anywhere else in the world. Thus Dutch approach is based on a formal model. Marris in 1954 compares the Dutch and British approaches to forecasting, and concludes that British makes use of informal models. Marris also fears the mechanistic adherence to models in the generation of forecasts when the economic system changes. In chapter 8 of Clements/Hendry, a general framework for the analysis of adjustments to model-based forecast, which is based on relationships between DGP, the estimated econometric model, the mechanics of the forecasting technique, data accuracy, and any information about future events held at the beginning of the forecast period, is established (related with this problem). The key was that the DGP is not constant over the time and in practice, the econometric model and the DGP will not coincide.

Klein (1950, built his model for the Cowles commission) explicitly recognize that, when the goal is forecasting, econometric modelling practice may differ from when explanation or description are the aim, also he construct a mathematical model consisting of structural relationships in the economy arguing for forecasting it is not necessary to know the structure of the economy (one may focus on a reduced form). But also he will also believe that this will result worse (due to loss of information) unless the system is known exactly.


Granger and Newbold in 1973 showed some methods that are misleading and inadequate. Chong and Hendry developed a test for large scale macroeconomic models.

3 A Framework for Forecasting

Amongst the many attributes of the system to be forecasted Clements/Hendry identify 6 facets of the system:

- The nature of the DGP
A. NATURE OF THE DGP
i) Stationary DGP
ii) Co-integrated stationary DGP
iii) Evolutionary, non-stationary DGP

B. KNOWLEDGE LEVEL
i) Known DGP, known $\theta$
ii) Known DGP, unknown $\theta$
iii) Unknown DGP, unknown $\theta$

C. DIMENSIONALITY
OF THE SYSTEM
i) Scalar Process
ii) Closed vector process
iii) Open Vector Process

D. FORM OF ANALYSIS
i) Asymptotic analysis
ii) Finite sample results

E. FORECAST HORIZON
i) 1-step
ii) Multi-step

F. LINEARITY
OF THE SYSTEM
i) Linear
ii) Non-linear

Table 1: Forecasting framework

- The knowledge level about that DGP
- The dimensionality of the system under investigation
- The form of the analysis
- The forecast horizon
- The linearity of the system

Let $\theta \in \Theta \subseteq \mathbb{R}^k$ for DGP parameter vector, and $x_t$ as the vector of outcomes at $t$, the complete sample is from $t = 1$ to $t = T + H$ denoted by $X_{T+H}^T = (X_{T-1}^T, x_t, X_{T+1}^T, X_{T+H}^T)$, the DGP at $t$ is $D_x(x_t \mid X_{t-1}^T, X_0, \theta)$, where $X_0$ denotes the initial conditions.

Then a forecast is a function of past information, and for period $t + h$ is given by $\hat{x}_{t+h} = f_h(X_{t-1})$ where $f_h(\cdot)$ may depend on a prior estimate of $\theta$. This framework is shown in table 1.

The combination of the probable cases will give a number 216 cases. The state $\{i, i, i, i, i, i\}$ is the simplest case but unrealistic. More realism can be obtained at the cost of greater complexity, to $\{i, ii, i, i, i, i\}$, $\{ii, i, i, i, i, i\}$, $\{ii, iii, ii, ii\}$ or to $\{iii, iii, iii, ii, ii\}$ (where the last one is most realistic).
4 Alternative Forecasting Methods

Necessary conditions for a good forecast can be counted as:

- Regularities should be captured;
- these regularities should be informative for the future,
- and they should be included in the model.
- Non-regularities should be excluded.

The first two properties are valid for all forecast methods. The latter two depend on the method that is being used.

Alternatives methods of forecasting are:

4.1 Guessing

This approach can be ruled out, it is just depend on luck. There might be some expert systems, but the extensions of formal procedures to flexibly incorporate expert knowledge that the associated model might not otherwise be able to exploit.

4.2 Extrapolation

This method is fine as long as the perceived tendencies do indeed persisting, but likelihood of it in doubt. In addition forecasts are most useful when they predict changes in tendencies, and extrapolative methods cannot do so.

4.3 Leading Indicators

This is based on forecasting on current, leading, and lagging indicators, like composite leading economic indicator (CLI). This method is used extensively throughout the world. The idea is that market economies are characterized by business cycles which are sequences of expansions and contractions in economic activity. The aim is to distinguish the cycles by their turning points, to relate cycles to economic indicators which lead their emerging stages, and to use the indicators to predict future turning points. It emphasizes identifying business cycles and the recession and expansion stages of the cycle, and this method is relatively cheap and easy to use forecasting tool and
the least theoretical. But such methods are unreliable unless the reasons for the the link of the past to the future are clear (like production time as a rationale for indicator of new orders for production process). There are actually two problems of identification of business cycle analysis. The first is identification of the business cycle from the raw macro-aggregates on GDP, unemployment etc. The second is translating movements in the CLIs into predictions about turning points.

According to Lahiri and Moore (in 1991), it is unwise to rely on CLIs as the primary forecasting tool, proposed for more limited objectives, such as predicting phases of the business cycle (peaks, troughs) or very short-term horizons when updated measures of other variables may not be available. Most economic series are highly autocorrelated and intercorrelated so it is hard to get the regularities. This can be done only by only mixing the economic theory and data analysis.

4.4 Surveys

Sample surveys of plans, expectations, and anticipations by consumers and business can be used for forecasting, requiring much research for a success. Although there are advantages to pool information from different sources but the problem is incorporation of the relevant information into the forecasting procedure. It should not be assumed a substitute of econometric systems. To treat survey information as an additional reading on the state of the economy might be an alternative. Surveys may play a role to choose a departure point for forecast.

4.5 Time-series Models

Kalman, in 1960 and Box and Jenkins in 1976 proposed scalar version of time series. In this version of time series ARIMA is the dominant class of scalar time series models. Showing that any purely indeterministic stationary time series has an infinite moving-average (MA) representation can be proved by Wold decomposition and any infinite MA can be approximated by ARMA model. Many series might be non-stationary, by differencing (integrating-d) these series can be stationary. Then ARMA model becomes ARIMA($p,d,q$). Putting a time linear function will also eliminate the deterministic polynomial in time of degree ($d-1$).
Harvey has proposed some class of models known structural time series models in 1989, and these are modelled as the sum of a number of unobserved components which nevertheless have a direct interpretation as trend, seasonal and irregular components. From these we can derive a reduced form with a single disturbance term (called as Box-Jenkins ARIMA models).

The multivariate form of Box-Jenkins models are the autoregressive representation. These models are claimed success in USA, but econometric systems should be possible outperform these models (VARs omit equilibrium-correction feedbacks). VARs in levels often ignore integrated components, so could predict poorly.

4.6 Econometric Systems

These models combine the theoretical and empirical knowledge of how economies function, providing a progressive research strategy. Like time series, econometric systems are based on statistical models (by this way allowing deriving measure of forecast uncertainty and forecasting adequacy). In order to interpret point forecast (in the absence of guide), Chatfield proposes an interval forecasts (range within which predictions will fall within a certain probability, usually 95%) and assessing the ways in which these can be calculated. These interval forecast is introduced for univariate case in the Clements/Hendry, Forecasting Economic Time Series. Calculations are easy for scalar or VAR representations of stationary process (without parameter uncertainty). Forecast error variances are more difficult to calculate for large, non-linear macroeconometric models, so analogous expressions for prediction intervals are not available. Econometric models are the primary method used in the book. Most of the macroeconometric uses some methods depending on complementary rather than adversarial roles (like using intercept correction). It can be given some conditions of econometric systems to capture some certain features of economic systems, which make more reliable than the alternatives:

1. Well specification, so the model can capture the available information for forecast,

2. dominating the alternatives namely how well it accounts for the results obtained by rival explanations;
3. invariance to regime shifts and structural change;
4. accuracy and precisely estimation to minimize the impact of estimation uncertainty

5 An Artificial-Data Example

The last section presents and analyzes a short example\(^2\) based on an artificial data model. This should help to illustrate the basic ideas and concepts. The first step is the data generation. This is to mimic actual data collected. Then the parameters are estimated and tested against the actual data set. Finally, the forecast generates estimates for the future.

5.1 Data Generating Process

5.1.1 Process Representation

First a data set must be created. The data are named “consumption”, “income” and “saving”, denoted by \(C\), \(Y\) and \(S\) respectively, and were generated satisfying the following three equations:

\[
\begin{align*}
\Delta C_t &= 0.02 + 0.5 \Delta Y_t + 0.2 S_{t-1} + \epsilon_{1t} \\
\Delta Y_t &= 0.05 + 0.5 \Delta Y_{t-1} + \epsilon_{2t} \\
S_t &= Y_t - C_t
\end{align*}
\]

where \(\epsilon_{it} \sim \text{IN}[0, \sigma_{ii}]\) with \(E[\epsilon_{1t}\epsilon_{2t}] = 0 \ \forall t, s, \ \text{and} \ \sigma_{11} = (0.05)^2 \ \text{and} \ \sigma_{22} = (0.1)^2\). The notation \(\text{IN}[\mu, \sigma_{ii}]\) denotes an independently sampled, normal random variable with expectation of \(\mu\) and variance of \(\sigma_{ii}\).

This model defines an autoregressive process, since the data values in \(t\) are dependent on the previous values. This process is here expressed in its error correction form, which takes the form:\(^3\)

\[
\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \epsilon_t, \quad t = 1, \ldots, T
\]

\(^2\)The example is taken from Clements/Hendry pp. 18-32

\(^3\)This notation is taken from Johansen [3]
where $D_t$ is the deterministic term. Note that for the process defined in (1)-(3), $S_t$ from equation (3) must substituted into (1). The vector $X_t$ is then $(C_t, Y_t)'$ and $\Delta X_t$ becomes $(\Delta C_t, \Delta Y_t)'$. The dimension of $X_t$ is therefore 2, so the matrices $\Pi$, $\Gamma_i$ and $\Phi$ are $2 \times 2$.

$k$ refers to the number of lags of the process, which is the number of steps backwards one needs to perform in order to compute a new data point. So if $k = 2$ e.g., then $X_{t-2}$ and $X_{t-1}$ are required in order to generate $X_t$. This is indeed the case for the model we are considering. This will become obvious if we express the process in its level form, which takes the general form:

$$X_t = \Pi_1 X_{t-1} + \ldots + \Pi_k X_{t-k} + \Phi D_t + \epsilon_t, \quad t = 1, \ldots, T. \quad (5)$$

For the process we are considering, we replace $\Delta C_t$ and $\Delta Y_t$ with $C_t - C_{t-1}$ and $Y_t - Y_{t-1}$ in (1) and (2) respectively and solve for $C_t$ and $Y_t$:

$$C_t = 0.045 + 0.45Y_{t-1} - 0.25Y_{t-2} + 0.8Ct - 1 + \epsilon_{1t} + 0.5\epsilon_{2t} \quad (6)$$

$$Y_t = 0.05 + 1.5Y_{t-1} - 0.5Y_{t-2} + \epsilon_{2t} \quad (7)$$

Two things should be noted at this stage: First, the data have no direct relationship to actual series with these names, they are purely fictional. Secondly, the mechanism in the example is fully known, which is generally not the case in the real world.

Based on this mechanism, Clements/Hendry generate three time series with initial values of zero and replacing $\{\epsilon_{1t}\}$ and $\{\epsilon_{2t}\}$ with suitably scaled numbers from a random number generator. After discarding the first 50 observations to avoid a dependence on non-stochastic start-up conditions, the resulting time series are interpreted as quarterly data from 1950:1 to 1996:4.

### 5.1.2 Process Properties: Integration Order and Cointegration

Looking more closely at the model, we will be able to identify some properties. Equation (2) has the general form

$$y_t = \mu + \rho y_{t-1} + v_t \quad \forall t. \quad (8)$$

With $-1 \leq \rho \leq 1$, such a process is stationary. A property of such a
process is that the expected value is constant for all \( t \):

\[
E[y_t] = \frac{\mu}{(1 - \rho)}
\]  

(9)

It is easily shown that for \( \Delta Y_t \) the expected value \( E[\Delta Y_t] = 0.1 \). Note however that \( \Delta Y_t \) is a difference of order one, implying that \( Y_t \) is I(1). Similarly, we find that \( C_t \) is I(1). Substituting (2) into (1), we get the reduced-form consumption equation:

\[
\Delta C_t = 0.045 + 0.25\Delta Y_{t-1} + 0.2S_{t-1} + \epsilon_{1t} + 0.5\epsilon_{2t}
\]  

(10)

Knowing that \( Y_t = Y_{t-1} + \Delta Y_t \) and \( C_t = C_{t-1} + \Delta C_t \), (3) can be rewritten:

\[
S_t \equiv Y_t - C_t = (Y_{t-1} + \Delta Y_t) - (C_{t-1} + \Delta C_t)
\]

\[
S_t = S_{t-1} + \Delta Y_t - \Delta C_t
\]  

(11)

Substituting (2) and (10) into (11), we get:

\[
S_t = 0.005 + 0.25\Delta Y_{t-1} + 0.8S_{t-1} - \epsilon_{1t} + 0.5\epsilon_{2t}
\]  

(12)

In the long run, we can substitute \( \Delta Y_{t-1} \) by its expected value \( E[\Delta Y_t] \). The process \( S_t \) has the stationary form of the process displayed in (8), with \( \rho = 0.8 \leq 1 \) and \( E[S_t] = \frac{0.005 + 0.25E[\Delta Y_t]}{1 - \rho} \). This yields \( E[S_t] = 0.0375 \), or 15% annualized.

Since (12) is equivalent to (3), we have shown that the linear combination of \( Y_t \) and \( C_t \), both series being I(1), is itself I(0), i.e. stationary. This property is the cointegration property of \( Y_t \) and \( C_t \) and constitutes the relationships that hold in the long run: If two or more variables are cointegrated, they must obey an equilibrium relationship in the long run, although the may diverge substantially from equilibrium in the short run. In general, a process \( X_t \), integrated of order 1 is cointegrated with cointegration vector \( \beta \neq 0 \) if \( \beta'X_t \) can be made stationary.

Equation (2) and (12) define a 2-equation model that we can write in vector notation as in equation (5) with \( k = 1 \):

\[
X_t = D_t + \Pi X_{t-1} + \epsilon_t,
\]  

(13)
where $X_t = (S_t, \Delta Y_t)'$. Expressing $D_t$ and $\Pi$, we get:

$$X_t = \begin{pmatrix} 0.005 \\ 0.05 \end{pmatrix} + \begin{pmatrix} 0.8 & 0.25 \\ 0.0 & 0.50 \end{pmatrix} X_{t-1} + \epsilon_t. \quad (14)$$

Comparing (13) with the stationary process of (8), and realizing that all value of $\Pi$ are between $-1$ and $1$, we can intuitively see that the model is $I(0)$, i.e. stationary (which is in fact the case).

5.2 Parameter Estimation

This section deals with the estimation of the parameters. In our specific example the real values are known of course, since we used them in the data generating process. But we want to see the quality of the forecast, using estimates and assuming that we do not know the actual values. Three ways are briefly discussed: An unrestricted vector auto-regression (VAR), a restricted VAR and, assuming that the equations of the DGP are known (but not the values of the parameters), a structural model estimate.

5.2.1 Unrestricted Vector Auto-Regression

Considering the 2-dimensional autoregressive process $X_t$ with lag $k = 2$ defined by the equations:

$$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + D_t + \epsilon_t, \quad t = 1, \ldots, T, \quad (15)$$

with $X_t = (C_t, Y_t)'$, the aim is to find values for $\Pi_1$, $\Pi_2$ and $D_t$ so that the system produces data points that are as close as possible to the data set from the previous section. The estimation performed by Clements/Hendry produces the following result:

$$X_t = \begin{pmatrix} 0.92 & 0.25 \\ 0.25 & 1.22 \end{pmatrix} X_{t-1} + \begin{pmatrix} -0.17 & 0 \\ -0.31 & -0.2 \end{pmatrix} X_{t-2} + \begin{pmatrix} 0.07 \\ 0.05 \end{pmatrix} + \epsilon_t, \quad (16)$$

where the estimates for the variances of the error are $\sqrt{\hat{\omega}_{11}} = 0.064$ and $\sqrt{\hat{\omega}_{22}} = 0.097$.

To check the robustness of the model, misspecification tests should be performed. In particular, we have to make sure that the estimated residuals $\hat{\epsilon}_t$ of every variable ($C_t$, $S_t$ and $Y_t$) have no serial correlation and do not
deviate too much from Gaussian white noise.\footnote{For a brief discussion of misspecification tests see Johansen \cite{3}.}

The results of the fitting and the misspecification tests are shown in figure 1 for $C_t$ and in figure 2 and 3 for $Y_t$ and $S_t$ respectively.\footnote{All graphical data displays are taken from Clements/Hendry.}

Figure 1: Actual and fitted, residuals and graphical statistics for consumption

\subsection*{5.2.2 Restricted Vector Auto-Regression}

The second system estimation performed is a restricted auto-regression. This estimation builds on the cointegration property of the system to reflect the long-run correlation of the variables.

The starting point is again a 2-dimensional VAR-process with $k = 2$ lags, that we now express in its auto-correction form:

$$\Delta X_t = \Pi X_{t-1} + \Gamma \Delta X_{t-1} + D_t + \epsilon_t, \quad t = 1, \ldots, T.$$ \hfill (17)

From the cointegration property follows that there exists some kind of linear dependence between the series (in the long run), since a linear combination is stationary. The idea now is to include this property into the system...
Figure 2: Actual and fitted, residuals and graphical statistics for saving

Figure 3: Actual and fitted, residuals and graphical statistics for income

(17): This can be done in $\Pi$, because this is were the long run behavior lies. So instead of considering a $p \times p$-matrix $\Pi$, where $p$ is the dimension of
vector $X_t$, we set $\Pi = \alpha \times \beta'$, where $\alpha$ and $\beta$ are $p \times r$ matrices with $r < p$. The rank of $\Pi$ is reduced to reflect the linear combination. This leads to the reduced error correction model:

$$\Delta X_t = \alpha \times \beta' X_{t-1} + \Gamma \Delta X_{t-1} + D_t + \epsilon_t, \quad t = 1, \ldots, T.$$  (18)

$\beta$ is the cointegration vector. From the previous section we know that this vector is $(1, -1)$, as we found out that $S_t = Y_t - C_t$ is stationary. We also know that the reduced rank is 1, since we have 2 series $Y_t$ and $C_t$ being individually $I(1)$. However, this knowledge can not be expected if one does not know the complete DGP. The standard procedure is to determine the cointegration rank through statistical tests. Furthermore, not all solutions for $\alpha \times \beta'$ have the required properties. Table 2 displays the values for the two $\beta$-vectors that were found, whereas figure 4 plots the two levels combinations given by $\beta'_1 X_t$ and $\beta'_2 X_t$, and provides visual support for the proposition that the former is $I(0)$ while the latter is $I(1)$.

### 5.2.3 Structural Model Estimate

The last estimate performed by Clements/Henry is based on the assumption that equations defining the GDP are known, and only the parameters have to be estimated. The technique employed is referred as full information maximum likelihood-method (FIML). Full information means that all parameters of the model are estimated at once (whereas one equation at a time is estimated when performing VAR).\(^6\) The resulting equations of the estimated system are very close to the initial equations (1)-(2):

$$\Delta \hat{C}_t = 0.04 + 0.3 \hat{Y}_t + 0.24 \hat{S}_{t-1} + \epsilon_{1t} \quad (19)$$

---

\(^6\)See Davidson/MacKinnon [2] for a presentation of FIML-models
\[ \Delta \hat{Y}_t = 0.04 + 0.4\Delta \hat{Y}_{t-1} + \epsilon_{2t}, \quad (20) \]

with \( \sqrt{\hat{\sigma}_{11}} = 0.046 \) and \( \sqrt{\hat{\sigma}_{22}} = 0.098 \).

To check whether the model obtained is well specified, Clements/Hendry propose recursive estimation. Consistent results confirm that the parameters are constant over time.

### 5.3 Model Forecast

This section analyzes how accurate the forecasts are, using the estimates from the previous section. First, we look at a 1-step ahead forecast, then at a long run estimate.

#### 5.3.1 Forecasting 1-Step Ahead

Figure 5 shows the sequence of 1-step ahead forecasts for \( \Delta C_t \) and \( S_t \) over the last 20 data points, as well as the implied 1-step ahead forecasts for the level of consumption \( C_t \) and the change of income \( \Delta Y_t \).

The error bands centered around the forecasts represent the confidence intervals at a 95% level. The fact that almost all realized outcomes lie
within that band means that the deviation from the real data comes from the equations’ innovation error rather than wrong parameter estimates. This is confirmed by forecast accuracy-tests performed on the data forecasts.

5.3.2 Forecasting h-Steps Ahead

For the long run forecast, we will go back to equation (13). This will allow us show that, even with known parameters, with a model that passes all the tests of adequacy of its specification, that explains most of the variation in the data, and that does not suffer from any of the practical difficulties of measurement errors, parameter change or non-modelled variables, the ability to predict vanishes rapidly.

For a time \( T + j \), \( j > 0 \), we have:

\[
X_{T+j} = D + \Pi X_{T+j} + \epsilon_{T+j},
\]

so that by successive substitution, conditional on the outcome at time \( T \), we obtain:
\[ X_{T+j} = (I_n + \Pi)D + \Pi^2 X_{T+j-2} + \epsilon_{T+j} + \Pi \epsilon_{T+j-1} \]

\[
\vdots
\]

\[ = \left( \sum_{k=0}^{j-1} \Pi^k \right) D + \Pi^j X_T + \sum_{k=0}^{j-1} \Pi^k \epsilon_{T+j-k} \]  \hspace{1cm} (22)

Since \( \mathbb{E}[\epsilon_{T+j-k}] = 0 \) in all future periods, we compute the forecast using:

\[ \hat{X}_{T+j} = \left( \sum_{k=0}^{j-1} \Pi^k \right) D + \Pi^j X_T \]  \hspace{1cm} (23)

\( \Pi^j \) converges to 0 for stationary processes, so forecast also converges, to \((I_n - \Pi)^{-1} D\). Figure 6 however show that this convergence happens quite early.

![Figure 6: h-step ahead forecasts and confidence bands](image)

Furthermore, the error band confidence interval for the stationary Differences are constant, but for the non-stationary variable \( C_t \), the variances of the forecast become linear trends incorporating the accumulated noise. Eventually, the forecasts become almost totally uninformative due to the large variance.
6 Conclusion

The purpose of this paper is to introduce some basic ideas of econometric forecasting and to place the theory in a historical context. We have described a general framework and shown alternative methods of forecasts. Finally, we have briefly introduced a numerical example which contains a set of problems that econometrician are facing. We have discussed some solutions and techniques, without however going into details.

References

