

Forecasting Aggregated Time Series Variables A Survey

Helmut Lütkepohl ¹

Department of Economics, European University Institute, Via della Piazzola 43, I-50133
Firenze, ITALY, email: helmut.luetkepohl@eui.eu

Abstract. Aggregated times series variables can be forecasted in different ways. For example, they may be forecasted on the basis of the aggregate series or forecasts of disaggregated variables may be obtained first and then these forecasts may be aggregated. A number of forecasts are presented and compared. Classical theoretical results on the relative efficiencies of different forecasts are reviewed and some complications are discussed which invalidate the theoretical results. Contemporaneous as well as temporal aggregation are considered.

Key Words: Autoregressive moving-average process, temporal aggregation, contemporaneous aggregation, vector autoregressive moving-average process

JEL classification: C22, C32

Contents

1	Introduction and Overview	1
2	VARMA Processes	3
2.1	General Setup	3
2.2	Forecasting	5
3	Linear Aggregation of VARMA Processes	7
3.1	Contemporaneous Aggregation	7
3.2	Temporal Aggregation	11
3.3	Forecasting with Data of Different Sampling Frequencies	14
4	Implications of Estimation and Model Specification	15
4.1	General Results for Estimated Processes	16
4.2	Aggregated Processes	17
4.3	Model Specification	18
5	Practical Complications	19
5.1	Nonlinear Transformations	19
5.2	Non-Gaussian Processes	20
5.3	Aggregation with Time-Varying and Stochastic Weights	21
5.4	Structural Change	22
6	Conclusions	23

¹...

1 Introduction and Overview

The European Monetary Union (EMU) has stimulated the need for euro-area macroeconomic studies and forecasts of area wide aggregates. This has induced a revival of the discussion of aggregation of time series variables over the last years. Clearly, the EMU poses some challenges for constructing euro-area time series in particular because the treatment of the pre-EMU period is not obvious. Given that the European Central Bank (ECB) needs to base its policy on informative data, it is not surprising that some work on contemporaneously aggregating euro-area data has been done recently. Aggregation problems have already been discussed many years ago. Although they can still serve as a basis for the current discussion, it turns out that they are in many ways too limited. Also, in addition to contemporaneous aggregation, temporal aggregation continues to be a problem of considerable interest because the year is an important planning period for policy makers whereas many data are collected more frequently. In this study I will review some recent developments in temporal and contemporaneous aggregation with an emphasis on results related to forecasting aggregated variables.

When forecasting a contemporaneously and/or temporally aggregated variable is of interest, there are different possibilities to proceed. For example, one may directly use the aggregated series, construct a time series model for its data generation process (DGP) and use that for forecasting. Alternatively, one may construct a time series model for the DGP of the disaggregated data and forecast the disaggregate series. Predictions of the aggregate are then obtained by aggregating the disaggregate forecasts. A number of earlier studies on aggregating time series variables have investigated the relative efficiencies of these predictors. Some of the results will be reviewed in the following.

The basic model for the disaggregated series will usually be assumed to be an autoregressive integrated moving average (ARIMA) or vector autoregressive moving average (VARMA) process. The VARMA class has the advantage of being closed with respect to linear transformations, that is, a linearly transformed finite order VARMA process has again a finite order VARMA representation. Therefore linear aggregation issues can be studied within this class.

In theory a disaggregated optimal forecast will be at least as good as a forecast based on the aggregated information. No general ranking is possible in this framework, however, if disaggregate series are predicted with univariate time series models and then the forecasts are aggregated. In practice, the situation may be different from what theory implies, however. Reasons may be that the ideal assumptions underlying the theoretical results are not

satisfied in practice. For instance, the true DGPs will not be known and models specified and estimated on the basis of the available data only approximate the DGPs. Moreover, the actual variable of interest may be some transform of the variable for which a suitable linear time series model is found. For instance, log transformations are often considered. In that case nonlinear aggregation may be needed which is not covered by the theoretical results. Also, in the case of EMU mentioned earlier, it is conceivable if not likely that structural changes have occurred over the period of interest. Thus, assuming a time invariant DGP may be problematic in some situations. These issues will also be addressed in this study.

The structure of the paper is as follows. In the next section VARMA models will be presented as the basic forecasting tool. The section mainly serves to lay out the framework of analysis. In Section 3 the classical aggregation problems will be considered and standard results for predictors for aggregated variables will be presented. Extensions and complications of the basic results are treated in Sections 4 and 5. Conclusions follow in Section 6. Some of this paper is based on Lütkepohl (2006).

Notation, Terminology, Abbreviations

The following notation and terminology is used. A nonstationary variable is called integrated of order d ($I(d)$) if it is still nonstationary after taking differences $d - 1$ times but it can be made stationary or asymptotically stationary by differencing d times. In part of the following discussion the variables will be assumed to be stationary, sometimes referred to as $I(0)$. If they are integrated of order greater than zero they may be cointegrated. For instance, there may be linear combinations of $I(1)$ variables which are $I(0)$.

The *lag operator* also sometimes called *backshift operator* is denoted by L and it is defined as usual by $Ly_t \equiv y_{t-1}$. The *differencing operator* is denoted by Δ , that is, $\Delta y_t \equiv y_t - y_{t-1} \equiv (1 - L)y_t$. For a random variable or random vector x , $x \sim (\mu, \Sigma)$ signifies that its mean (vector) is μ and its variance (covariance matrix) is Σ . The notation $x \sim \mathcal{N}(\mu, \Sigma)$ is used if x is normally distributed. The $(K \times K)$ identity matrix is denoted by I_K and the determinant of a matrix A are denoted by $\det A$. The natural logarithm of a real number is signified by \log . The symbols \mathbb{Z} , \mathbb{N} and \mathbb{C} are used as usual for the sets of integers, the positive integers and the complex numbers, respectively.

DGP stands for data generation process. VAR, AR, MA, ARMA and VARMA are used as abbreviations for vector autoregressive, autoregressive, moving-average, autoregressive moving-average and vector autoregressive moving-average (process). MSE abbreviates mean squared error.

2 VARMA Processes

2.1 General Setup

Suppose the DGP of the K -dimensional multiple time series, y_1, \dots, y_T , is stationary, that is, its first and second moments are time invariant. It is a (finite order) VARMA process if it can be represented in the general form

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \dots + M_q u_{t-q}, \quad t \in \mathbb{Z}, \quad (2.1)$$

where A_1, \dots, A_p are $(K \times K)$ autoregressive parameter matrices while M_1, \dots, M_q are moving-average parameter matrices also of dimension $(K \times K)$. Defining the VAR and MA operators, respectively, as $A(L) = I_K - A_1 L - \dots - A_p L^p$ and $M(L) = I_K + M_1 L + \dots + M_q L^q$, the model can be written in more compact notation as

$$A(L)y_t = M(L)u_t, \quad t \in \mathbb{Z}. \quad (2.2)$$

Here u_t is a white-noise process with zero mean, nonsingular, time-invariant covariance matrix $E(u_t u_t') = \Sigma_u$ and zero covariances, $E(u_t u_{t-h}') = 0$ for $h = \pm 1, \pm 2, \dots$. To indicate the orders of the VAR and MA operators, the process (2.1) is sometimes called a VARMA(p, q) process. Notice, however, that so far we have not made further assumptions regarding the parameter matrices so that some or all of the elements of the A_i 's and M_j 's may be zero. In other words, there may be a VARMA representation with VAR or MA orders less than p and q , respectively.

The matrix polynomials in (2.2) are assumed to satisfy

$$\det A(z) \neq 0, |z| \leq 1, \quad \text{and} \quad \det M(z) \neq 0, |z| \leq 1 \quad \text{for} \quad z \in \mathbb{C}. \quad (2.3)$$

The first of these conditions ensures that the VAR operator is *stable* and the process is stationary. Then it has a pure MA representation

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} \quad (2.4)$$

with MA operator $\Phi(L) = I_K + \sum_{i=1}^{\infty} \Phi_i L^i = A(L)^{-1} M(L)$, i.e., $\Phi_0 = I_K$. The representation (2.4) is known as *Wold MA representation* of the process and, as we will see later, the u_t 's are just the one-step ahead forecast errors. Some of the forthcoming results are valid for more general stationary processes with Wold representation (2.4) which may not have a finite order VARMA representation. In that case, it is assumed that the Φ_i 's are absolutely summable so that the infinite sum in (2.4) is well-defined.

The second part of condition (2.3) is the usual *invertibility condition* for the MA operator which implies the existence of a pure VAR representation of the process,

$$y_t = \sum_{i=1}^{\infty} \Xi_i y_{t-i} + u_t, \quad (2.5)$$

where $\Xi(L) = I_K - \sum_{i=1}^{\infty} \Xi_i L^i = M(L)^{-1}A(L)$. Occasionally invertibility of the MA operator will not be a necessary condition. In that case, it is assumed without loss of generality that $\det M(z) \neq 0$, for $|z| < 1$. In other words, the roots of the MA operator are outside or on the unit circle. There are still no roots inside the unit circle, however. This assumption can be made without loss of generality because it can be shown that for an MA process with roots inside the complex unit circle an equivalent one exists which has all its roots outside and on the unit circle.

It may be worth noting that every pair of operators $A(L)$, $M(L)$ which leads to the same transfer functions $\Phi(L)$ and $\Xi(L)$ defines an equivalent VARMA representation for y_t . This nonuniqueness problem of the VARMA representation will not be of concern here. It is important in practice when it comes to parameter estimation because for estimation purposes a unique parametrization is required.

As specified in (2.1), it is assumed that the process is defined for all $t \in \mathbb{Z}$. For stable, stationary processes this assumption is convenient because it avoids consideration of issues related to initial conditions. Alternatively, one could define y_t to be generated by a VARMA process such as (2.1) for $t \in \mathbb{N}$, and specify the initial values $y_0, \dots, y_{-p+1}, u_0, \dots, u_{-p+1}$ separately. Under the present assumptions they can be defined such that y_t is stationary. Another possibility would be to define fixed initial values or perhaps even $y_0 = \dots = y_{-p+1} = u_0 = \dots = u_{-p+1} = 0$. In general, such an assumption implies that the process is not stationary but just *asymptotically stationary*, that is, the first and second order moments converge to the corresponding quantities of the stationary process obtained by specifying the initial conditions accordingly or defining y_t for $t \in \mathbb{Z}$.

Both the MA and the VAR representations of the process will be convenient to work with in particular situations. Another useful representation of a stationary VARMA process is the state-space representation which will not be used in this review, however. The relation between state-space models and VARMA processes is considered, for example, by Aoki (1987), Hannan & Deistler (1988), Wei (1990) and Harvey (2006).

In practice some of the variables may be integrated. In that case $\det A(z)$ has roots on the unit circle, e.g., $\det A(1) = 0$. There may also be roots on the unit circle other than one. For example, there may be seasonal unit roots. These features do not create additional problems for forecasting if the DGP is assumed to be known because the general forecasting

formulas for VARMA processes are still valid. For processes with integrated variables it is convenient, however, to define the process for $t \in \mathbb{N}$ with suitable starting values. There are important recent advances in specifying and estimating models with such features which are obviously important from a practical point of view. Also the levels VARMA form (2.1) is not necessarily the most convenient one for inference purposes. Since forecasting is the focus of the present exposition such issues are not discussed here.

Another important shortcoming of the present setup in practice is the absence of deterministic terms. Adding such terms to the stochastic part is a convenient way to account for them. Hence, the process

$$y_t = \mu_t + x_t,$$

may be considered, where μ_t is a deterministic term and x_t is the purely stochastic part with VARMA representation. The deterministic part may be a constant, $\mu_t = \mu_0$, a linear trend, $\mu_t = \mu_0 + \mu_1 t$, or a higher order polynomial trend. It may also include seasonal dummy variables or other dummies.

Future values of deterministic terms are precisely known. Therefore they are easy to handle from a forecasting point of view. In order to forecast y_t , a forecast of the purely stochastic process x_t is needed and then the deterministic part corresponding to the forecast period is simply added. The forecast errors and MSE matrices are the same as for the purely stochastic process. Therefore they will be ignored in the following. Extending the results to processes with deterministic terms is straightforward.

2.2 Forecasting

When forecasting a set of variables is the objective, it is useful to think about a loss function or an evaluation criterion for the forecast performance. Given such a criterion, optimal forecasts may be constructed. VARMA processes are particularly useful for producing forecasts that minimize the forecast MSE. Therefore this criterion will be used here and the reader is referred to Granger (1969) and Granger & Newbold (1977, Section 4.2) for a discussion of other forecast evaluation criteria. Some of the results for aggregated variables reviewed later also hold for more general loss functions.

Forecasts of the variables of the VARMA process (2.1) are obtained easily from the pure VAR form (2.5). Assuming an independent white noise process u_t , an optimal, minimum MSE h -step forecast at time τ is the conditional expectation given the y_t 's, $t \leq \tau$,

$$y_{\tau+h|\tau} \equiv E(y_{\tau+h} | y_\tau, y_{\tau-1}, \dots).$$

It may be determined recursively for $h = 1, 2, \dots$, as

$$y_{\tau+h|\tau} = \sum_{i=1}^{\infty} \Xi_i y_{\tau+h-i|\tau}, \quad (2.6)$$

where $y_{\tau+j|\tau} = y_{\tau+j}$ for $j \leq 0$. If the u_t do not form an independent but only uncorrelated white noise sequence, the forecast obtained in this way is still the best linear forecast although it may not be the best in a larger class of possibly nonlinear functions of past observations.

For given initial values, the u_t 's can also be determined under the present assumption of a known process. Hence, the h -step forecasts may be determined alternatively as

$$y_{\tau+h|\tau} = A_1 y_{\tau+h-1|\tau} + \dots + A_p y_{\tau+h-p|\tau} + \sum_{i=h}^q M_i u_{\tau+h-i}, \quad (2.7)$$

where, as usual, the sum vanishes if $h > q$.

Both ways of computing h -step forecasts from VARMA models rely on the availability of initial values. In the pure VAR formula (2.6) all infinitely many past y_t 's are in principle necessary if the VAR representation is indeed of infinite order. In contrast, in order to use (2.7), the u_t 's need to be known which are unobserved and can only be obtained if all past y_t 's or initial conditions are available. If only y_1, \dots, y_τ are given, the infinite sum in (2.6) may be truncated accordingly. For large τ , the approximation error will be negligible because the Ξ_i 's go to zero quickly as $i \rightarrow \infty$. Alternatively, precise forecasting formulas based on y_1, \dots, y_τ may be obtained via the so-called *Multivariate Innovations Algorithm* of Brockwell & Davis (1987, §11.4).

Under the present assumptions, the properties of the forecast errors for stable, stationary processes are easily derived by expressing the process (2.1) in Wold MA form (2.4). In terms of this representation the optimal h -step forecast may be expressed as

$$y_{\tau+h|\tau} = \sum_{i=h}^{\infty} \Phi_i u_{\tau+h-i}. \quad (2.8)$$

Hence, the forecast errors are seen to be

$$y_{\tau+h} - y_{\tau+h|\tau} = u_{\tau+h} + \Phi_1 u_{\tau+h-1} + \dots + \Phi_{h-1} u_{\tau+1}. \quad (2.9)$$

Thus, the forecast is unbiased (i.e., the forecast errors have mean zero) and the MSE or forecast error covariance matrix is

$$\Sigma_y(h) \equiv E[(y_{\tau+h} - y_{\tau+h|\tau})(y_{\tau+h} - y_{\tau+h|\tau})'] = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'.$$

If u_t is normally distributed (Gaussian), the forecast errors are also normally distributed,

$$y_{\tau+h} - y_{\tau+h|\tau} \sim \mathcal{N}(0, \Sigma_y(h)). \quad (2.10)$$

Hence, forecast intervals etc. may be derived from these results in the familiar way under Gaussian assumptions.

It is also interesting to note that the forecast error variance is bounded by the covariance matrix of y_t ,

$$\Sigma_y(h) \rightarrow_{h \rightarrow \infty} \Sigma_y \equiv E(y_t y_t') = \sum_{j=0}^{\infty} \Phi_j \Sigma_u \Phi_j'. \quad (2.11)$$

Hence, forecast intervals will also have bounded length as the forecast horizon increases.

The situation is different if there are $I(d)$ variables with $d > 0$. The formula (2.7) can again be used for computing the forecasts. Their properties will be different from those for stationary processes, however. Although the Wold MA representation does not exist for integrated processes, the Φ_j coefficient matrices can be computed in the same way as for stationary processes from the power series $A(z)^{-1}M(z)$ which still exists for $z \in \mathbb{C}$ with $|z| < 1$. Hence, the forecast errors can still be represented as in (2.9) (see Lütkepohl (2005, Chapters 6 and 14)). Thus, formally the forecast errors look quite similar to those for the stationary case. Now the forecast error MSE matrix is unbounded, however, because the Φ_j 's in general do not converge to zero as $j \rightarrow \infty$. Despite this general result, there may be linear combinations of the variables which can be forecasted with bounded precision if the forecast horizon gets large. This situation arises if there is cointegration.

As long as theoretical results are discussed and $I(1)$ series are under investigation one could consider the first differences of the process, Δy_t , which also have a VARMA representation. If there is genuine cointegration, then Δy_t is over-differenced in the sense that its VARMA representation has MA unit roots even if the MA part of the levels y_t is invertible.

3 Linear Aggregation of VARMA Processes

As mentioned in the introduction, a major advantage of the class of VARMA processes is that it is closed with respect to linear transformations. In other words, linear transformations of VARMA processes have again a finite order VARMA representation (see Lütkepohl (1984, 1987)). These transformations are the basis for studying aggregation problems. Contemporaneous aggregation will be considered in Section 3.1 and temporal aggregation will be treated in Section 3.2.

3.1 Contemporaneous Aggregation

In this section I present some forecasting results for linearly transformed and contemporaneously aggregated processes from Lütkepohl (1987) where also proofs and further references

can be found. Let y_t be a stationary VARMA process with pure, invertible Wold MA representation (2.4), that is, $y_t = \Phi(L)u_t$ with $\Phi_0 = I_K$, F is an $(M \times K)$ matrix with rank M . Suppose forecasting the transformed process $z_t = Fy_t$ is of interest. Moreover, suppose that the corresponding Wold MA representation is

$$z_t = v_t + \sum_{i=1}^{\infty} \Psi_i v_{t-i} = \Psi(L)v_t. \quad (3.1)$$

From (2.8) the optimal h -step predictor for z_t at origin τ , based on its own past, is then

$$z_{\tau+h|\tau} = \sum_{i=h}^{\infty} \Psi_i v_{\tau+h-i}, \quad h = 1, 2, \dots \quad (3.2)$$

Another predictor may be based on forecasting y_t and then transforming the forecast,

$$z_{\tau+h|\tau}^o \equiv Fy_{\tau+h|\tau}, \quad h = 1, 2, \dots \quad (3.3)$$

Before the two forecasts $z_{\tau+h|\tau}^o$ and $z_{\tau+h|\tau}$ are compared it may be of interest to draw attention to yet another possible forecast. If the dimension K of the vector y_t is large, it may be difficult to construct a suitable VARMA model for the underlying process and one may consider forecasting the individual components of y_t by univariate methods and then transforming the univariate forecasts. Because the component series of y_t can be obtained by linear transformations, they also have ARMA representations. Denoting the corresponding Wold MA representations by

$$y_{kt} = w_{kt} + \sum_{i=1}^{\infty} \theta_{ki} w_{k,t-i} = \theta_k(L)w_{kt}, \quad k = 1, \dots, K, \quad (3.4)$$

the optimal univariate h -step forecasts are

$$y_{k,\tau+h|\tau}^u = \sum_{i=h}^{\infty} \theta_{ki} w_{k,\tau+h-i}, \quad k = 1, \dots, K, \quad h = 1, 2, \dots \quad (3.5)$$

Defining $y_{\tau+h|\tau}^u = (y_{1,\tau+h|\tau}^u, \dots, y_{K,\tau+h|\tau}^u)'$, these forecasts can be used to obtain an h -step forecast

$$z_{\tau+h|\tau}^u \equiv Fy_{\tau+h|\tau}^u \quad (3.6)$$

of the variables of interest.

The three forecasts (3.2), (3.3) and (3.6) of the transformed process z_t are now compared. The MSE matrices corresponding to the three forecasts are denoted by $\Sigma_z(h)$, $\Sigma_z^o(h)$ and $\Sigma_z^u(h)$, respectively. Because $z_{\tau+h|\tau}^o$ uses the largest information set, it is not surprising that it has the smallest MSE matrix and is hence the best one out of the three forecasts,

$$\Sigma_z(h) \geq \Sigma_z^o(h) \quad \text{and} \quad \Sigma_z^u(h) \geq \Sigma_z^o(h), \quad h \in \mathbb{N}, \quad (3.7)$$

where “ \geq ” means that the difference between the left-hand and right-hand matrices is positive semidefinite. Thus, forecasting the original process y_t and then transforming the forecasts is generally more efficient in terms of MSE than forecasting the transformed process directly or transforming univariate forecasts. It is possible, however, that some or all of the forecasts are identical. Actually, for $I(0)$ processes, all three predictors always approach the same long-term forecast of zero. Consequently,

$$\Sigma_z(h), \Sigma_z^o(h), \Sigma_z^u(h) \rightarrow \Sigma_z \equiv E(z_t z_t') \quad \text{as } h \rightarrow \infty. \quad (3.8)$$

Moreover, it can be shown that if the one-step forecasts are identical, then they will also be identical for larger forecast horizons. More precisely we have,

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau} \Rightarrow z_{\tau+h|\tau}^o = z_{\tau+h|\tau} \quad h = 1, 2, \dots, \quad (3.9)$$

$$z_{\tau+1|\tau}^u = z_{\tau+1|\tau} \Rightarrow z_{\tau+h|\tau}^u = z_{\tau+h|\tau} \quad h = 1, 2, \dots, \quad (3.10)$$

and, if $\Phi(L)$ and $\Theta(L)$ are invertible,

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau}^u \Rightarrow z_{\tau+h|\tau}^o = z_{\tau+h|\tau}^u \quad h = 1, 2, \dots \quad (3.11)$$

Thus, one may ask whether the one-step forecasts can be identical and it turns out that this is indeed possible. The following proposition which summarizes results of Tiao & Guttman (1980), Kohn (1982) and Lütkepohl (1984), gives conditions for this to happen.

Proposition 1. Let y_t be a K -dimensional stochastic process with Wold MA representation as in (2.4) and F an $(M \times K)$ matrix with rank M . Then, defining $\Phi(L) = I_K + \sum_{i=1}^{\infty} \Phi_i L^i$, $\Psi(L) = I_K + \sum_{i=1}^{\infty} \Psi_i L^i$ as in (3.1) and $\Theta(L) = \text{diag}[\theta_1(L), \dots, \theta_K(L)]$ with $\theta_k(L) = 1 + \sum_{i=1}^{\infty} \theta_{ki} L^i$ ($k = 1, \dots, K$), the following relations hold:

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau} \iff F\Phi(L) = \Psi(L)F, \quad (3.12)$$

$$z_{\tau+1|\tau}^u = z_{\tau+1|\tau} \iff F\Theta(L) = \Psi(L)F \quad (3.13)$$

and, if $\Phi(L)$ and $\Theta(L)$ are invertible,

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau}^u \iff F\Phi(L)^{-1} = F\Theta(L)^{-1}. \quad (3.14)$$

□

There are several interesting implications of this proposition. First, if the forecasts are identical, then they are equally efficient for any choice of loss function and not just for the forecast MSE. In fact, Sbrana & Silvestrini (2008) show that the forecasts $z_{\tau+1|\tau}$ and $z_{\tau+1|\tau}^u$ can be equally efficient in terms of MSE even if the condition (3.13) does not hold. They analyze the relative efficiency of the two forecasts in terms of MSE in detail for the case of a bivariate MA(1) process. Second, if y_t consists of independent components ($\Phi(L) = \Theta(L)$) and z_t is just their sum, i.e., $F = (1, \dots, 1)$, then

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau} \iff \theta_1(L) = \dots = \theta_K(L). \quad (3.15)$$

In other words, forecasting the individual components and summing up the forecasts will result in a different predictor and may be strictly more efficient than forecasting the sum directly whenever the components are not generated by stochastic processes with identical temporal correlation structures. Third, forecasting the univariate components of y_t individually can be as efficient a forecast for y_t as forecasting on the basis of the multivariate process if and only if $\Phi(L)$ is a diagonal matrix operator. Related to this result is a well-known condition for Granger-noncausality. For a bivariate process $y_t = (y_{1t}, y_{2t})'$, y_{2t} is said to be Granger-causal for y_{1t} if the former variable is helpful for improving the forecasts of the latter variable. In terms of the previous notation this may be stated by choosing $F = (1, 0)$ and defining y_{2t} as being Granger-causal for y_{1t} if $z_{\tau+1|\tau}^o = Fy_{\tau+1|\tau} = y_{1,\tau+1|\tau}^o$ is a better forecast than $z_{\tau+1|\tau}$. From (3.12) it then follows that y_{2t} is not Granger-causal for y_{1t} if and only if $\phi_{12}(L) = 0$, where $\phi_{12}(L)$ denotes the upper right hand element of $\Phi(L)$. This characterization of Granger-noncausality is well-known in the related literature (e.g., Lütkepohl (2005, Section 2.3.1)).

It may also be worth noting that in general there is no unique ranking of the forecasts $z_{\tau+1|\tau}$ and $z_{\tau+1|\tau}^u$. Depending on the structure of the underlying process y_t and the transformation matrix F , either $\Sigma_z(h) \geq \Sigma_z^u(h)$ or $\Sigma_z(h) \leq \Sigma_z^u(h)$ will hold and the relevant inequality may be strict in the sense that the left-hand and right-hand matrices are not identical. As mentioned earlier, Sbrana & Silvestrini (2008) consider this case in more detail in the context of a bivariate MA(1) process.

Some but not all the results in this section carry over to nonstationary $I(1)$ processes. For example, the result (3.8) will not hold in general if some components of y_t are $I(1)$ because in this case the three forecasts do not necessarily converge to zero as the forecast horizon gets large. On the other hand, the conditions in (3.12) and (3.13) can be used for the differenced processes. For these results to hold, the MA operator may have roots on the unit circle and hence overdifferencing is not a problem.

The previous results on linearly transformed processes can also be used to compare different predictors for temporally aggregated processes. Some related results will be summarized next.

3.2 Temporal Aggregation

The results on linear transformations of VARMA processes can also be used to study temporal aggregation problems. Suppose aggregation of the variables y_t generated by (2.1) over m subsequent periods is desired. For instance, $m = 3$ if one wishes to aggregate monthly data to quarterly figures. To express the temporal aggregation as a linear transformation we define

$$\mathbf{y}_\vartheta = \begin{bmatrix} y_{m(\vartheta-1)+1} \\ y_{m(\vartheta-1)+2} \\ \vdots \\ y_{m\vartheta} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_\vartheta = \begin{bmatrix} u_{m(\vartheta-1)+1} \\ u_{m(\vartheta-1)+2} \\ \vdots \\ u_{m\vartheta} \end{bmatrix} \quad (3.16)$$

and specify the process

$$\mathcal{A}_0 \mathbf{y}_\vartheta = \mathcal{A}_1 \mathbf{y}_{\vartheta-1} + \cdots + \mathcal{A}_P \mathbf{y}_{\vartheta-P} + \mathcal{M}_0 \mathbf{u}_\vartheta + \mathcal{M}_1 \mathbf{u}_{\vartheta-1} + \cdots + \mathcal{M}_Q \mathbf{u}_{\vartheta-Q}, \quad (3.17)$$

where

$$\mathcal{A}_0 = \begin{bmatrix} I_K & 0 & 0 & \cdots & 0 \\ -A_1 & I_K & 0 & \cdots & 0 \\ -A_2 & -A_1 & I_K & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \\ -A_{m-1} & -A_{m-2} & -A_{m-3} & \cdots & I_K \end{bmatrix},$$

$$\mathcal{A}_i = \begin{bmatrix} A_{im} & A_{im-1} & \cdots & A_{im-m+1} \\ A_{im+1} & A_{im} & \cdots & A_{im-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{im+m-1} & A_{im+m-2} & \cdots & A_{im} \end{bmatrix}, \quad i = 1, \dots, P,$$

with $A_j = 0$ for $j > p$ and $\mathcal{M}_0, \dots, \mathcal{M}_Q$ defined in an analogous manner. The order $P = \min\{n \in \mathbb{N} | nm \geq p\}$ and $Q = \min\{n \in \mathbb{N} | nm \geq q\}$. Notice that the time subscript of \mathbf{y}_ϑ is different from that of y_t . The new time index ϑ refers to another observation frequency than t . For example, if t refers to months and $m = 3$, then ϑ refers to quarters.

Left-multiplying (3.17) by \mathcal{A}_0^{-1} and redefining the white noise process appropriately shows that \mathbf{y}_ϑ has a standard VARMA(P, Q) representation. Using that representation, temporal

aggregation over m periods can be analyzed as a linear transformation. In fact, different types of temporal aggregation can be handled. For instance, the aggregate may be the sum of subsequent values or it may be their average. Furthermore, temporal and contemporaneous aggregation can be dealt with simultaneously. In all of these cases the aggregate has a finite order VARMA representation if the original variables are generated by a finite order VARMA process and the structure of the aggregate can be analyzed using linear transformations. This approach may not be the most convenient one if the VARMA orders or specific structures are of interest. For another approach to study temporal aggregates exploiting properties of matrix polynomials see Marcellino (1999) and for a generalization of his approach see Gómez & Aparicio-Pérez (2009). A recent survey of temporal aggregation was given by Silvestrini & Veredas (2008).

Different forms of temporal aggregation are of interest, depending on the types of variables involved. If y_t consists of stock variables, then temporal aggregation is usually associated with *systematic sampling*, sometimes called *skip-sampling* or *point-in-time sampling*. In other words, the process

$$\mathbf{s}_\vartheta = y_{m\vartheta} \tag{3.18}$$

is used as an aggregate over m periods. For example, if $m = 3$, then \mathbf{s}_ϑ consists of every third member of the y_t process. This type of aggregation contrasts with temporal aggregation of flow variables where a temporal aggregate is typically obtained by summing up consecutive values. Thus, aggregation over m periods gives the aggregate

$$\mathbf{z}_\vartheta = y_{m\vartheta} + y_{m\vartheta-1} + \cdots + y_{m\vartheta-m+1}. \tag{3.19}$$

Now if, for example, t refers to months and $m = 3$, then three consecutive observations are added to obtain the quarterly value. For the moment it is assumed that the disaggregated process y_t is stationary and invertible and has a Wold MA representation as in (2.4), $y_t = \Phi(L)u_t$ with $\Phi_0 = I_K$. Given that temporal aggregation can be viewed as a linear transformation of some auxiliary process, this implies that \mathbf{s}_ϑ and \mathbf{z}_ϑ are also stationary and have Wold MA representations. Forecasting stock and flow variables is now discussed in turn. In other words, forecasts for \mathbf{s}_ϑ and \mathbf{z}_ϑ are considered.

Suppose first that one wishes to forecast \mathbf{s}_ϑ . Then the past aggregated values $\{\mathbf{s}_\vartheta, \mathbf{s}_{\vartheta-1}, \dots\}$ may be used to obtain an h -step forecast $\mathbf{s}_{\vartheta+h|\vartheta}$ as in (2.8) on the basis of the MA representation of \mathbf{s}_ϑ . If the disaggregate process y_t is available, another possible forecast results by systematically sampling forecasts of y_t which gives $\mathbf{s}_{\vartheta+h|\vartheta}^o = y_{m\vartheta+mh|m\vartheta}$. Using the results for linear transformations, the latter forecast generally has a lower MSE than $\mathbf{s}_{\vartheta+h|\vartheta}$ and the

difference vanishes if the forecast horizon $h \rightarrow \infty$. For special processes the two predictors are identical. It follows from relation (3.12) of Proposition 1 that the two predictors are identical for $h = 1, 2, \dots$, if and only if

$$\Phi(L) = \left(\sum_{i=0}^{\infty} \Phi_{im} L^{im} \right) \left(\sum_{i=0}^{m-1} \Phi_i L^i \right) \quad (3.20)$$

(Lütkepohl (1987, Proposition 7.1)). Thus, there is no loss in forecast efficiency under any loss function if the MA operator of the disaggregate process has the multiplicative structure in (3.20). This condition is, for instance, satisfied if y_t is a purely seasonal process with seasonal period m such that

$$y_t = \sum_{i=0}^{\infty} \Phi_{im} u_{t-im}. \quad (3.21)$$

It also holds if y_t has a finite order MA structure with MA order less than m . Interestingly, it also follows that there is no loss in forecast efficiency if the disaggregate process y_t is a VAR(1) process, $y_t = A_1 y_{t-1} + u_t$. In that case, the MA operator can be written as

$$\Phi(L) = \left(\sum_{i=0}^{\infty} A_1^{im} L^{im} \right) \left(\sum_{i=0}^{m-1} A_1^i L^i \right)$$

and, hence, it has the required structure.

Now consider the case of a vector of flow variables y_t for which the temporal aggregate is given in (3.19). For forecasting the aggregate \mathbf{z}_ϑ one may use the past aggregated values and compute an h -step forecast $\mathbf{z}_{\vartheta+h|\vartheta}$ as in (2.8) on the basis of the MA representation of \mathbf{z}_ϑ . Alternatively, one may again forecast the disaggregate process y_t and aggregate the forecasts. This forecast is denoted by $\mathbf{z}_{\vartheta+h|\vartheta}^o$, that is,

$$\mathbf{z}_{\vartheta+h|\vartheta}^o = y_{m\vartheta+mh|m\vartheta} + y_{m\vartheta+mh-1|m\vartheta} + \dots + y_{m\vartheta+mh-m+1|m\vartheta}. \quad (3.22)$$

Again the results for linear transformations imply that the latter forecast generally has a lower MSE than $\mathbf{z}_{\vartheta+h|\vartheta}$ and the difference vanishes if the forecast horizon $h \rightarrow \infty$. In this case equality of the two forecasts holds for small forecast horizons $h = 1, 2, \dots$, if and only if

$$\begin{aligned} & (1 + L + \dots + L^{m-1}) \left(\sum_{i=0}^{\infty} \Phi_i L^i \right) \\ &= \left(\sum_{j=0}^{\infty} (\Phi_{jm} + \dots + \Phi_{j(m-m+1)}) L^{jm} \right) \left(\sum_{i=0}^{m-1} (\Phi_0 + \Phi_1 + \dots + \Phi_i) L^i \right), \end{aligned} \quad (3.23)$$

where $\Phi_j = 0$ for $j < 0$ (Lütkepohl (1987, Proposition 8.1)). In other words, the two forecasts are identical and there is no loss in forecast efficiency from using the aggregate directly if

the MA operator of y_t has the specified multiplicative structure upon multiplication by $(1 + L + \dots + L^{m-1})$. This condition is also satisfied if y_t has the purely seasonal structure (3.21). However, in contrast to what was observed for stock variables, the two predictors are generally not identical if the disaggregate process y_t is generated by an MA process of order less than m .

It is perhaps also interesting to note that if there are both stock and flow variables in one system, then even if the underlying disaggregate process y_t is the periodic process (3.21), a forecast based on the disaggregate data may be better than directly forecasting the aggregate (Lütkepohl (1987, pp. 177-178)). This result is interesting because for the purely seasonal process (3.21) using the disaggregate process will not result in superior forecasts if a system consisting either of stock variables only or of flow variables only is considered.

So far temporal aggregation of stationary processes is discussed. Most of the results can be generalized to $I(1)$ processes by using the stationary process Δy_t instead of the original process y_t . Recall that forecasts for y_t can then be obtained from those of Δy_t . Moreover, in this context it may be worth taking into account that in deriving some of the conditions for forecast equality, the MA operator of the considered disaggregate process may have unit roots resulting from over-differencing. A result which does not carry over to the $I(1)$ case is the equality of long horizon forecasts based on aggregate or disaggregate variables. The reason is again that optimal forecasts of $I(1)$ variables do not settle down at zero eventually when $h \rightarrow \infty$.

Forecasting temporally aggregated processes has been discussed extensively in the literature. Early examples of treatments of temporal aggregation of time series are Abraham (1982), Amemiya & Wu (1972), Brewer (1973), Lütkepohl (1986a, b), Stram & Wei (1986), Telser (1967), Tiao (1972), Wei (1978) and Weiss (1984) among many others. More recently, Breitung & Swanson (2002) have studied the implications of temporal aggregation when the number of aggregated time units goes to infinity. As mentioned earlier, a recent survey with many references is given by Silvestrini & Veredas (2008).

3.3 Forecasting with Data of Different Sampling Frequencies

In practice it is not uncommon that data of different frequencies are available for different variables. For example, output data are usually collected quarterly or only annually whereas price indices, money stock variables or interest rate series are available at monthly or even higher observation frequency. Suppose that some quarterly series x_{θ} are considered which are related to the monthly series y_t . One may want to exploit the information in the quarterly series to predict the monthly series or vice versa. For that purpose the device used in (3.16)

may be extended and the quarterly process

$$\mathbf{z}_\vartheta = \begin{bmatrix} \mathbf{y}_\vartheta \\ x_\vartheta \end{bmatrix} \tag{3.24}$$

may be considered. Here \mathbf{y}_ϑ is the process defined in (3.16).

If the underlying monthly process is of VARMA or, more generally, infinite order MA type, then the previously mentioned results can be used to analyze the relative efficiencies of different forecasts by considering linear transformations of \mathbf{z}_ϑ . Again, using the disaggregate data which are available for y_t will in general be beneficial for improving the forecast efficiency. In fact, even if one is interested in monthly forecasts for all or some components of y_t , different forecasts can be compared using the previously mentioned results and the process (3.24) because, e.g., the 1-step ahead forecast of \mathbf{z}_ϑ will contain a 1-month ahead forecast for y_t .

Although this approach can be used to obtain some results of interest in the context of forecasting mixed frequency variables, there may again be issues for which other approaches are more convenient. For instance, a state-space representation is often helpful for analyzing forecasting problems. A state-space model can also be set up for mixed frequency variables (see Proietti & Moauro (2006)).

4 Implications of Estimation and Model Specification

Clements & Hendry (1998, Section 7.4) list a number of sources of forecast errors and, hence, forecast uncertainty. For example, in practice, the processes used for forecasting are not known but the parameters have to be estimated from data. Usually also the process orders and other characteristics are specified on the basis of the given time series and, hence, are uncertain. In addition, there may be parameter changes (structural change) during the estimation and/or forecast period which may contribute to forecast uncertainty. Also the available time series observations may be subject to measurement error. Moreover, the variables may not be normally distributed and may require nonlinear transformations to approximate their DGPs well by the standard forecasting models. Alternatively, the models may have to be augmented by nonlinear components. All these issues make forecasting more difficult and may invalidate the previously mentioned theoretical results which were derived under ideal conditions. Therefore it is important to understand the implications of these complications. Some of them are discussed now.

4.1 General Results for Estimated Processes

I denote by $\hat{y}_{\tau+h|\tau}$ the h -step forecast at origin τ given in Section 2.2, based on estimated rather than known coefficients. For instance, using the pure VAR representation of the process,

$$\hat{y}_{\tau+h|\tau} = \sum_{i=1}^{h-1} \hat{\Xi}_i \hat{y}_{\tau+h-i|\tau} + \sum_{i=h}^{\infty} \hat{\Xi}_i y_{\tau+h-i}, \quad (4.1)$$

where the $\hat{\Xi}_i$'s denote estimated parameter matrices. Of course, for practical purposes one may truncate the infinite sum at $i = \tau$ in (4.1). For the moment the infinite sum is considered and it is assumed that the model represents the DGP. Thus, there is no specification error. For this predictor the forecast error is

$$y_{\tau+h} - \hat{y}_{\tau+h|\tau} = (y_{\tau+h} - y_{\tau+h|\tau}) + (y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau}),$$

where $y_{\tau+h|\tau}$ is the optimal forecast based on known coefficients and the two terms on the right-hand side are uncorrelated if only data up to period τ are used for estimation. In that case the first term can be written in terms of u_t 's with $t > \tau$ and the second one contains only y_t 's with $t \leq \tau$. Hence, the two terms are uncorrelated and the forecast MSE becomes

$$\begin{aligned} \Sigma_{\hat{y}}(h) &= \text{MSE}(y_{\tau+h|\tau}) + \text{MSE}(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau}) \\ &= \Sigma_y(h) + E[(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau})(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau})']. \end{aligned} \quad (4.2)$$

The $\text{MSE}(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau})$ can be approximated by $\Omega(h)/T$, where

$$\Omega(h) = E \left[\frac{\partial y_{\tau+h|\tau}}{\partial \theta'} \Sigma_{\hat{\theta}} \frac{\partial y'_{\tau+h|\tau}}{\partial \theta} \right], \quad (4.3)$$

θ is the vector of coefficients to be estimated, and $\Sigma_{\hat{\theta}}$ is its asymptotic covariance matrix. Yamamoto (1980), Baillie (1981) and Lütkepohl (2005) give more detailed expressions for $\Omega(h)$ and Hogue, Magnus & Pesaran (1988) provide an exact treatment of the special case of an AR(1) process. The matrix $\Omega(h)$ is positive semidefinite and the forecast MSE,

$$\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \frac{1}{T} \Omega(h), \quad (4.4)$$

for estimated processes is larger (or at least not smaller) than the corresponding quantity for known processes, as one would expect. Because the additional term includes the asymptotic covariance matrix $\Sigma_{\hat{\theta}}$ of the parameter estimators in (4.3), more efficient parameter estimation also increases forecasting efficiency. On the other hand, for large sample sizes T , the term $\Omega(h)/T$ will be small or even negligible.

It may be worth noting that deterministic terms can be accommodated easily, as discussed earlier. In the present situation the uncertainty in the estimators related to such terms can also be taken into account like that of the other parameters. If the deterministic terms are specified such that the corresponding parameter estimators are asymptotically independent of the other estimators, an additional term for the estimation uncertainty stemming from the deterministic terms has to be added to the forecast MSE matrix (4.4). For deterministic linear trends in univariate models more details are presented in Kim, Leybourne & Newbold (2004).

In practice, choosing the estimators and computing estimates for VARMA processes is not a trivial task. A main problem is the choice of a unique (identified) parametrization or canonical form of the VARMA class. Although these problems can have a substantial impact on the estimates and, hence, on the forecasts obtained from estimated processes they are not considered here because they would require discussions away from the main focus of the present paper. The reader is referred to Lütkepohl (2006) for more discussion and further references on the issue. In the next subsection the implications of estimated processes for forecasting aggregates are discussed.

4.2 Aggregated Processes

In Section 3.1 it is mentioned that for contemporaneous aggregation generally forecasting the disaggregate process and aggregating the forecasts ($z_{\tau+h|\tau}^o$) is more efficient than forecasting the aggregate directly ($z_{\tau+h|\tau}$). In this case, if the sample size is large enough, the part of the forecast MSE due to estimation uncertainty will eventually be so small that the estimated $\hat{z}_{\tau+h|\tau}^o$ is again superior to the corresponding $\hat{z}_{\tau+h|\tau}$. There are cases, however, where the two forecasts are identical for known processes. Now the question arises whether in these cases the MSE term due to estimation errors will make one forecast preferable to its competitors. Indeed if estimated instead of known processes are used, it is possible that $\hat{z}_{\tau+h|\tau}^o$ loses its optimality relative to $\hat{z}_{\tau+h|\tau}$ because the MSE part due to estimation may be larger for the former than for the latter. Consider the case, where a number of series are simply added to obtain a univariate aggregate. Then it is possible that a simple parsimonious univariate ARMA model describes the aggregate well, whereas a large multivariate model is required for an adequate description of the multivariate disaggregate process. Clearly, it is conceivable that the estimation uncertainty in the multivariate case becomes considerably more important than for the univariate model for the aggregate. Lütkepohl (1987) shows that this may indeed happen in small samples. In fact, similar situations can not only arise for contemporaneous aggregation but also for temporal aggregation. Generally, if two

predictors based on known processes are nearly identical, the estimation part of the MSE becomes important and generally the predictor based on the smaller model is then to be preferred.

Various extensions of the previous results have been discussed in the literature. For example, Lewis & Reinsel (1985) and Lütkepohl (1985) consider the forecast MSE for the case where the true process, which may be of the mixed VARMA type, is approximated by a finite order VAR, thereby extending earlier univariate results by Bhansali (1978). Reinsel & Lewis (1987), Basu & Sen Roy (1987), Engle & Yoo (1987), Sampson (1991) and Reinsel & Ahn (1992) present results for processes with unit roots. Stock (1996) and Kemp (1999) obtain results by assuming that the forecast horizon h and the sample size T both go to infinity simultaneously.

4.3 Model Specification

So far only the effect of estimation uncertainty on the forecast MSE is taken into account. This analysis still assumes a known model structure and only allows for estimated parameters. In practice, model specification usually precedes estimation and usually there is additional uncertainty attached to this step in the forecasting procedure.

As mentioned earlier, it is also possible to explicitly take into account the fact that in practice models are only approximations to the true DGP by considering finite order VAR and AR approximations to infinite order processes. This has also been done by Lütkepohl (1987). Under these assumptions it is again found that the forecast $\hat{z}_{\tau+h|\tau}^o$ may lose its optimality and forecasting the aggregate directly or forecasting the disaggregate series with univariate methods and aggregating univariate forecasts may become preferable.

Therefore it is not surprising that recent empirical studies do not reach a unanimous conclusion regarding the value of using disaggregate information in forecasting aggregates. For example, Marcellino, Stock & Watson (2003) found disaggregate information to be helpful for forecasting several euro-area aggregates while Hubrich (2005) and Hendry & Hubrich (2005) concluded that disaggregation resulted in forecast deterioration in a comparison based on euro-area inflation data. Of course, there can be many more reasons than just estimation and specification issues for the empirical results to differ from the theoretical ones, as mentioned earlier.

5 Practical Complications

In practice there are a number of data features which are not in line with the theoretical framework assumed in the foregoing. Some of them are discussed in the following. For example, many variables are modelled with ARMA processes in logs rather than original levels. If forecasts are based on such nonlinear transformations the previously used linear aggregation framework is not sufficient for analyzing the implications for forecasting. Nonlinear transformations are discussed in Section 5.1. Such transformations are often considered to make the DGP more normal. Alternatively one may fit an ARMA or VARMA model and explicitly allow for a non-Gaussian distribution of the residuals. Related issues are discussed in Section 5.2. Another deviation from the basic framework arises from time-varying aggregation weights. The weights may change deterministically or stochastically. This situation may, for example, come up in aggregating European Union or euro-area data and is discussed in Section 5.3. Another problem in the context of constructing euro-area data is structural change in some component series. Again such a feature is not foreseen in the basic framework of the previous sections where time invariant models are assumed. The problem is briefly addressed in Section 5.4.

5.1 Nonlinear Transformations

The log transformation or more generally the Box-Cox transformation is often applied in practice to make the variance of a time series variable more homogeneous over time or to obtain a DGP which is better modelled by a Gaussian process. In econometric models the logs of variables are often considered to turn a multiplicative relation into a linear one. If a model for the logs of a variable is used for forecasting one may reverse the log transformation by applying the exponential function to the forecasts. If an aggregate is of interest one may aggregate the resulting forecasts. This approach has a couple of drawbacks, however. First, it is well-known that an instantaneous nonlinear transformation applied to the optimal forecast of a variable may not result in the optimal forecast of the transformed variable (Granger & Newbold (1976)). In particular, if optimal forecasts of the logs are available, converting them to forecasts for the original variable by applying the exponential function will in general not be optimal. Second, standard results regarding the optimality of aggregated forecasts refer to linear models and forecasts. They may not carry over to nonlinear processes. Third, it is not clear a priori that a forecast obtained via the log transformed variable is better than a direct forecast of the original variable even if the distribution of the log variable is closer to being Gaussian.

Suppose the variable of interest is y_t and define $x_t = \log y_t$. Moreover, suppose that x_t is generated by a Gaussian ARMA process. Granger & Newbold (1976) show that the optimal h -step forecast of $y_{\tau+h}$ at forecast origin τ is

$$y_{\tau+h|\tau} = \exp(x_{\tau+h|\tau} + \frac{1}{2}\sigma_x^2(h)),$$

where $x_{\tau+h|\tau}$ denotes the optimal forecast of $x_{\tau+h}$ and $\sigma_x^2(h)$ is the corresponding forecast error variance. Clearly, this forecast is in general different from

$$y_{\tau+h|\tau}^{naive} = \exp(x_{\tau+h|\tau})$$

which is called the naive forecast by Granger & Newbold (1976). Thus, one may want to consider the optimal forecast when aggregates are of interest.

Lütkepohl & Xu (2009) consider the question under which conditions it is useful to base a forecast on the log transformation if forecasting the original variable is of interest. They consider a number of economic variables and use univariate AR models to forecast the variables directly or their logs. Based on a range of stock market variables, gross domestic product (GDP) and consumption data from a number of countries their general finding is that forecasting the logs and converting the forecasts either by using the naive or the optimal predictor can lead to substantial gains in forecast precision as measured by the forecast MSE. It does not make much difference whether the optimal or the naive forecast is used. In fact, since in practice the optimal forecast also requires an estimator of the forecast error variance it may result in a slightly larger MSE than the naive forecast. This last result is driven by the fact that the forecast error variance of the log series is typically very small and, hence, the extra term in the optimal forecasting formula does not make much difference. Generally forecast gains from using logs are obtained for series for which the log transformation actually makes the variance more homogeneous. If this condition is not satisfied, the forecasts based on logs can even be much worse in terms of MSE than forecasts based directly on the original series.

For forecasting aggregates these results suggest that using the log transformation for individual components of an aggregate may be beneficial. Given that the log transformation can actually do damage for the forecast accuracy, it is not clear that necessarily all components of an aggregate should be transformed. Generally, it may be worth considering the implications of nonlinear transformations if linear or nonlinear aggregates are of interest.

5.2 Non-Gaussian Processes

If the DGP of a multiple time series is not normally distributed, point forecasts can be computed based on ARIMA or VARMA models as before. They may still be best *linear*

forecasts and may in fact be minimum MSE forecasts if u_t is independent white noise, as discussed in Section 2.2. For non-Gaussian processes nonlinear forecasts may be more accurate, however. In practice it may not be easy to beat the linear forecasts unless a specific nonlinearity is known to be present and, hence, can be modelled. If a linear model with non-Gaussian residuals is considered, the distribution has to be taken into account in setting up forecast intervals. If the distribution is unknown, bootstrap methods can be used to compute interval forecasts (e.g., Findley (1986), Masarotto (1990), Grigoletto (1998), Kabaila (1993), Kim (1999), Clements & Taylor (2001), Pascual, Romo & Ruiz (2004)).

5.3 Aggregation with Time-Varying and Stochastic Weights

Constructing EMU data poses aggregation problems which are not covered by the basic linear aggregation framework. Suppose, for example, that one wants to construct an unemployment rate series for the EMU by aggregating the unemployment rates of the individual EMU countries. The overall unemployment rate, u_t^{EU} , is a weighted average of the individual unemployment rates, $u_t^{(i)}$,

$$u_t^{EU} = \sum_{i=1}^N w_i u_t^{(i)}. \quad (5.1)$$

Here the w_i 's are the weights and N denotes the number of member states. The weights sum to one, $\sum_{i=1}^N w_i = 1$. The question is which weights to use. Clearly, w_i should be related to the relative size of country i . Because the relative country sizes may change, it is clear that the weights in (5.1) may vary over time. In fact, the number of EMU countries has changed over the years and is meant to change further in the future when new member states enter the EMU. Thus, it may be more appropriate to denote the weights by w_{it} rather than w_i . Since it is known beforehand when a new member state joins the EMU, one may think of the changes in the weights as deterministic. Alternatively, if the weights are strictly related to the working populations in the different member states, it may in fact be more appropriate to view the weights as stochastic.

Generally, there are many series where aggregation is done with stochastic weights. For example, there are several proposals for aggregating GDP growth rates based on stochastic weights. Suppose y_t^{EU} denotes euro-area GDP and $y_t^{(i)}$ is the corresponding figure for country i . Then Winder (1997) computes the EMU growth rate as

$$\Delta \log y_t^{EU} = \sum_{i=1}^N \frac{y_{t-1}^{(i)} / e_{TB}^{(i)}}{y_{t-1}^{EU}} \Delta \log y_t^{(i)}, \quad (5.2)$$

where $e_t^{(i)}$ denotes the exchange rate of country i in period t and TB signifies a fixed base

year. The exchange rate is necessary for the pre-EMU period, that is, when an aggregate series for the pre-EMU period is constructed. Winder considers a fixed exchange rate for the sample period. Alternatively, Beyer, Doornik & Hendry (2001) propose the aggregate

$$\Delta \log y_t^{EU} = \sum_{i=1}^N \frac{y_{t-1}^{(i)}/e_{t-1}^{(i)}}{y_{t-1}^{EU}} \Delta \log y_t^{(i)}, \quad (5.3)$$

which uses a flexible exchange rate. In (5.2) and (5.3) the weights are $w_{it} = \frac{y_{t-1}^{(i)}/e_{t-1}^{(i)}}{y_{t-1}^{EU}}$ and $w_{it} = \frac{y_{t-1}^{(i)}/e_{t-1}^{(i)}}{y_{t-1}^{EU}}$, respectively. Thus, in both cases the weights are clearly stochastic and thereby the aggregation scheme falls outside the basic linear framework.

The creation of euro-area data for the pre-EMU period has been discussed by a number of other authors as well (e.g., Fagan, Henry & Mestre (2001), Bosker (2006), Anderson, Dungey, Osborn & Vahid (2007), Angelini & Marcellino (2007), Brüggemann & Lütkepohl (2005, 2006)). It is of considerable interest for both forecasting euro-area aggregates and for economic analysis within the euro-area. Therefore it is interesting to know more about the theoretical properties of forecasts based on aggregate or disaggregate information. Generally, investigating the properties of forecasts for aggregates with stochastic weighting schemes is an interesting problem for future research. For the special case of constructing EMU data an additional practical problem arises which is discussed in the next subsection.

5.4 Structural Change

A change in the structure of the DGP is a particular problem for forecasting. Prediction relies on some time invariance to project the past into the future. In the previous discussions a time invariant DGP is assumed. In practice structural change is rather common, however. A case of particular importance is, for example, the creation of a single currency area in Europe. This event is likely to have caused structural changes in a number of EMU countries which had to adjust their economic systems to the Maastricht criteria. For example, some countries had to reduce their inflation rates and budget deficits drastically to be able to join the EMU. Therefore it is not at all clear that simply aggregating pre-EMU time series and using such data to predict future values of a variable is a good strategy. Brüggemann & Lütkepohl (2006) propose to use German data for the pre-EMU period. In a forecast comparison based on time series constructed in different ways for a number of euro-area variables Brüggemann, Lütkepohl & Marcellino (2008) also include this proposal. They find that this kind of approach works well for series which have a similar level when joining German and EMU data. Examples are price indices, interest rates and an exchange rate. Again this type of aggregation falls outside the basic linear aggregation framework although

it may be of considerable interest in practice and therefore may be a fruitful area for further research.

Generally, if some change in the structure of the DGP has occurred during the estimation and specification period or happens even during the forecast period, a more detailed investigation of the implications for the quality of the forecasts is necessary. The general results for time invariant processes discussed in the foregoing do not necessarily carry over to this case. There are some useful proposals which may be helpful in this situation. For instance, Clements & Hendry (1999, Chapter 5) investigate the possibility of robustifying forecasts against breaks by over-differencing the variables. Clearly, anything that works in general can also be used for forecasting aggregated variables. A deeper analysis of the implications for the different forecasts considered in the foregoing may be an interesting direction for future research.

6 Conclusions

VARMA models are standard tools for producing linear forecasts for a set of time series variables. Different forecasts for aggregated time series variables based on these models are compared. Both contemporaneous and temporal aggregation are considered. Classical results imply that forecasting the disaggregated process and aggregating the forecasts is more efficient in terms of MSE than forecasting the aggregate directly and thereby ignoring the disaggregate information. Moreover, for contemporaneous aggregation, forecasting the individual components with univariate methods and aggregating these forecasts is compared to the two previously mentioned forecasts. Forecasting univariate components separately may lead to better forecasts than forecasting the aggregate directly and the reverse result may also be obtained, depending on the DGP of the time series under consideration. Using univariate forecasts is in general inferior to aggregating forecasts of the fully disaggregated process, however. These results hold if the DGPs are known.

In practice the relevant model for forecasting a particular set of time series will not be known, of course, and it is necessary to use sample information to specify and estimate a suitable candidate model from the ARMA or VARMA class. If estimation and specification uncertainty are taken into account it turns out that forecasts based on a disaggregated multiple time series may not be better and may even be inferior to forecasting an aggregate directly. This situation is in particular likely to occur if the DGPs are such that efficiency gains from disaggregation do not exist or are small and the aggregated process has a simple structure which can be captured with a parsimonious model.

A number of other deviations from the idealized basic linear framework used in the theoretical analysis are also considered. In particular, problems related to nonlinear transformations of the variables of interest, non-Gaussian DGPs, time-varying aggregation weights and structural change are discussed. Some of these problems are especially relevant in aggregating euro-area variables and leave room for further research on forecasting aggregated time series variables.

Generally the focus of this survey is on predicting the levels of a time series and no consideration is given to second order moments such as conditional variances. These can be very important for constructing forecast intervals and assessing the forecast uncertainty. Such issues are heavily discussed in the current financial econometrics literature. They warrant a separate survey.

References

- Abraham, B. (1982). Temporal aggregation and time series, *International Statistical Review* **50**: 285–291.
- Amemiya, T. & Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model, *Journal of the American Statistical Association* **67**: 628–632.
- Anderson, H., Dungey, M., Osborn, D. R. & Vahid, F. (2007). Constructing historical euro area data, *Discussion Paper*, Australian National University.
- Angelini, E. & Marcellino, M. (2007). Econometric analysis with backdated data: Unified Germany and the euro area, *Working Paper series No 752*, European Central Bank.
- Aoki, M. (1987). *State Space Modeling of Time Series*, Springer-Verlag, Berlin.
- Baillie, R. T. (1981). Prediction from the dynamic simultaneous equation model with vector autoregressive errors, *Econometrica* **49**: 1331–1337.
- Basu, A. K. & Sen Roy, S. (1987). On asymptotic prediction problems for multivariate autoregressive models in the unstable nonexplosive case, *Calcutta Statistical Association Bulletin* **36**: 29–37.
- Beyer, A., Doornik, J. A. & Hendry, D. F. (2001). Reconstructing historical euro-zone data, *Economic Journal* **111**: 308–327.
- Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain, *Annals of Statistics* **6**: 224–231.
- Bosker, E. M. (2006). On the aggregation of eurozone data, *Economics Letters* **90**: 260–265.
- Breitung, J. & Swanson, N. R. (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models, *Journal of Time Series Analysis* **23**: 651–665.
- Brewer, K. R. W. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models, *Journal of Econometrics* **1**: 133–154.
- Brockwell, P. J. & Davis, R. A. (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York.

- Brüggemann, R. & Lütkepohl, H. (2005). Uncovered interest rate parity and the expectations hypothesis of the term structure: Empirical results for the U.S. and Europe, *Applied Economics Quarterly* **51**: 143–154.
- Brüggemann, R. & Lütkepohl, H. (2006). A small monetary system for the euro area based on German data, *Journal of Applied Econometrics* **21**: 683–702.
- Brüggemann, R., Lütkepohl, H. & Marcellino, M. (2008). Forecasting euro area variables with German pre-EMU data, *Journal of Forecasting* **27**: 465–481.
- Clements, M. P. & Hendry, D. F. (1998). *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.
- Clements, M. P. & Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge, Ma.
- Clements, M. P. & Taylor, N. (2001). Bootstrap prediction intervals for autoregressive models, *International Journal of Forecasting* **17**: 247–267.
- Engle, R. F. & Yoo, B. S. (1987). Forecasting and testing in cointegrated systems, *Journal of Econometrics* **35**: 143–159.
- Fagan, G., Henry, J. & Mestre, R. (2001). An area-wide model for the Euro area, *Working Papers 42*, European Central Bank.
- Findley, D. F. (1986). On bootstrap estimates of forecast mean square errors for autoregressive processes, in D. M. Allen (ed.), *Computer Science and Statistics: The Interface*, North-Holland, Amsterdam, pp. 11–17.
- Gómez, V. & Aparicio-Pérez, F. (2009). A new state-space methodology to disaggregate multivariate time series, *Journal of Time Series Analysis* **30**: 97–124.
- Granger, C. W. J. (1969). Prediction with a generalized cost of error function, *Operations Research Quarterly* **20**: 199–207.
- Granger, C. W. J. & Newbold, P. (1976). Forecasting transformed series, *Journal of the Royal Statistical Society B* **38**: 189–203.
- Granger, C. W. J. & Newbold, P. (1977). *Forecasting Economic Time Series*, Academic Press, New York.
- Grigoletto, M. (1998). Bootstrap prediction intervals for autoregressions: Some alternatives, *International Journal of Forecasting* **14**: 447–456.
- Hannan, E. J. & Deistler, M. (1988). *The Statistical Theory of Linear Systems*, Wiley, New York.
- Harvey, A. (2006). Forecasting with unobserved components time series models, in G. Elliott, C. W. J. Granger & A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier, Amsterdam, pp. 327–412.
- Hendry, D. F. & Hubrich, K. (2005). Forecasting aggregates by disaggregates, *discussion paper*, European Central Bank.
- Hogue, A., Magnus, J. & Pesaran, B. (1988). The exact multi-period mean-square forecast error for the first-order autoregressive model, *Journal of Econometrics* **39**: 327–346.
- Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?, *International Journal of Forecasting* **21**: 119–136.

- Kabaila, P. (1993). On bootstrap predictive inference for autoregressive processes, *Journal of Time Series Analysis* **14**: 473–484.
- Kemp, G. C. R. (1999). The behavior of forecast errors from a nearly integrated AR(1) model as both sample size and forecast horizon become large, *Econometric Theory* **15**: 238–256.
- Kim, J. H. (1999). Asymptotic and bootstrap prediction regions for vector autoregression, *International Journal of Forecasting* **15**: 393–403.
- Kim, T. H., Leybourne, S. J. & Newbold, P. (2004). Asymptotic mean-squared forecast error when an autoregression with linear trend is fitted to data generated by an I(0) or I(1) process, *Journal of Time Series Analysis* **25**: 583–602.
- Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past?, *Journal of Econometrics* **18**: 337–349.
- Lewis, R. & Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**: 393–411.
- Lütkepohl, H. (1984). Linear transformations of vector ARMA processes, *Journal of Econometrics* **26**: 283–293.
- Lütkepohl, H. (1985). The joint asymptotic distribution of multistep prediction errors of estimated vector autoregressions, *Economics Letters* **17**: 103–106.
- Lütkepohl, H. (1986a). Forecasting temporally aggregated vector ARMA processes, *Journal of Forecasting* **5**: 85–95.
- Lütkepohl, H. (1986b). Forecasting vector ARMA processes with systematically missing observations, *Journal of Business & Economic Statistics* **4**: 375–390.
- Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Lütkepohl, H. (2006). Forecasting with VARMA models, in G. Elliott, C. W. J. Granger & A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 1*, Elsevier, Amsterdam, pp. 287–325.
- Lütkepohl, H. & Xu, F. (2009). The role of the log transformation in forecasting economic variables, *Working paper*, European University Institute.
- Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis, *Journal of Business & Economic Statistics* **17**: 129–136.
- Marcellino, M., Stock, J. H. & Watson, M. W. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information, *European Economic Review* **47**: 1–18.
- Masarotto, G. (1990). Bootstrap prediction intervals for autoregressions, *International Journal of Forecasting* **6**: 229–239.
- Pascual, L., Romo, J. & Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes, *Journal of Time Series Analysis* **25**: 449–465.
- Proietti, T. & Moauro, F. (2006). Dynamic factor analysis with non-linear temporal aggregation constraints, *Journal of the Royal Statistical Society* **C55**: 281–300.
- Reinsel, G. C. & Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis* **13**: 353–375.

- Reinsel, G. C. & Lewis, A. L. (1987). Prediction mean square error for non-stationary multivariate time series using estimated parameters, *Economics Letters* **24**: 57–61.
- Sampson, M. (1991). The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models, *Journal of Applied Econometrics* **6**: 67–76.
- Sbrana, G. & Silvestrini, A. (2008). What do we know about comparing aggregate and disaggregate forecasts?, *mimeograph*, CORE, Université Catholique de Louvain, Belgium.
- Silvestrini, A. & Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey, *Journal of Economic Surveys* **22**: 458–497.
- Stock, J. H. (1996). VAR, error correction and pretest forecasts at long horizons, *Oxford Bulletin of Economics and Statistics* **58**: 685–701.
- Stram, D. O. & Wei, W. W. S. (1986). Temporal aggregation in the ARIMA process, *Journal of Time Series Analysis* **7**: 279–292.
- Telser, L. G. (1967). Discrete samples and moving sums in stationary stochastic processes, *Journal of the American Statistical Association* **62**: 484–499.
- Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series, *Biometrika* **59**: 525–531.
- Tiao, G. C. & Guttman, I. (1980). Forecasting contemporaneous aggregates of multiple time series, *Journal of Econometrics* **12**: 219–230.
- Wei, W. W. S. (1978). Some consequences of temporal aggregation in seasonal time series models, in A. Zellner (ed.), *Seasonal Analysis of Economic Time Series*, U.S. Department of Commerce, Bureau of the Census, pp. 433–444.
- Wei, W. W. S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, Redwood City, Ca.
- Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models, *Journal of Econometrics* **26**: 271–281.
- Winder, C. C. A. (1997). On the construction of European area-wide aggregates - a review of issues and empirical evidence, *IFC Bulletin* **1**: 15–23.
- Yamamoto, T. (1980). On the treatment of autocorrelated errors in the multiperiod prediction of dynamic simultaneous equation models, *International Economic Review* **21**: 735–748.