Time-scale transformations of discrete time processes*

Òscar Jordà
University of California, Davis
ojorda@ucdavis.edu

Massimiliano Marcellino†
Università Bocconi, IGIER and CEPR
massimiliano.marcellino@uni-bocconi.it

March 2003

Abstract

This paper investigates the effects of temporal aggregation when the aggregation frequency is variable and possibly stochastic. The results that we report include, as a particular case, the well-known results on fixed-interval aggregation, such as when monthly data is aggregated into quarters. A variable aggregation frequency implies that the aggregated process will exhibit time-varying parameters and non-spherical disturbances, even when these characteristics are absent from the original model. Consequently, we develop methods for specification and estimation of the aggregate models and show with an example how these methods perform in practice.

JEL Classification Codes: C22, C43, F31

Keywords: time aggregation, time-scale transformation, irregularly spaced data, autoregressive conditional intensity model.

*We thank an anonymous referee, Rob Engle, Tom Rothenberg, Jim Stock and seminar participants at the European University Institute, Harvard University, the Midwestern Econometrics Group, U.C. Berkeley, U.C. Davis and U.C. Riverside for useful comments on a previous draft.

†Corresponding author: Massimiliano Marcellino, IGIER - Università Bocconi, Via Salasco 5, 20136, Milano, Italy. Phone: +39-02-5836-3327. Fax: +39-02-5836-3302.

1. Introduction

There are two traditional approaches for dealing with time aggregation issues. The consequences of fixed-interval time aggregation in discrete-time models (for example, data generated monthly but recorded quarterly) have been thoroughly investigated by Telser (1967), Brewer (1973), Wei (1981), Weiss (1984), and Marcellino (1999), among others. The analysis of continuous-time models from fixed-interval observed data appears, for example, in the work of Sims (1971), Geweke (1978) and Stock (1987, 1988). This paper addresses a more general problem instead: the possibility that discrete-time processes evolve at irregular (usually stochastic) intervals, whether because this time scale is native to the data generating process (DGP) or whether because it reflects the frequency of data recording. This question is central in economics for example, where behavioral models often describe the timing of certain economic events as endogenously determined, but it arises in other disciplines as well.

A brief taxonomy of the universe of possibilities allowed by variable time intervals demonstrates its wide applicability. Denominate the time scale at which the data are generated *original-time scale* and use the subscript τ to index variables. In general, the recording schedule of the data will not coincide with the original-time scale. Denominate this data recording frequency as the *aggregate-time* scale and use the subscript t to index variables. Depending on the variability of the intervals in each of these time scales, we distinguish the following four types of aggregation:

	original-time-scale τ	**Aggregate time-scale t**
Type I	Regularly spaced	Regularly spaced
Type II	Irregularly spaced	Regularly spaced
Type III	Regularly spaced	Irregularly spaced
Type IV	Irregularly spaced	Irregularly spaced

The traditional analysis of time aggregation described in the first paragraph corresponds to type I aggregation. Examples of type II aggregation are common in finance, where time series are analyzed at daily or even weekly frequencies even though transaction intensities in these markets vary anywhere between seconds to hours. More formally, Jordà (1999) shows that partial adjustment models (such as a model of inventory adjustment) naturally generate irregularly spaced data in original-time although em-

pirical analyses necessarily rely on aggregated quarterly or monthly data at best. The classical literature on business cycle analysis initiated by Burns and Mitchell (1946), the literature on time deformation introduced by Stock (1987), and data recording with missing observations are examples of type III aggregation. Finally, type IV aggregation can be found in finance, where tick by tick financial data are often "thinned" by some statistical procedure (see Engle and Russell, 1998) to distinguish between uninformed and informed trades. The thinning process delivers a new series in which aggregate-time is also irregularly spaced. These examples illustrate the encompassing nature of the unifying framework we present for dealing with time aggregation.

The more interesting results in the paper correspond to situations in which the frequency of aggregation is variable, such as when it is the realization of a stochastic point process. In this case, the aggregate-time processes have time-varying parameters and non-spherical disturbances even when these characteristics are absent in original-time. These properties are important considerations in terms of forecasting, estimation and testing since they explain in which directions aggregate models are likely to be misspecified and in which way this misspecification can be corrected. Consequently, we derive the representation of the aggregated data generating process (DGP) and derive maximum likelihood (ML) estimators for its parameters. The resulting models improve the estimation of structural parameters and provide more accurate forecasts, as illustrated with an example on inventory decisions.

The paper is organized as follows. Section 2 introduces the general framework we use and presents the results on the representation of the aggregate-time DGP. Section 3 deals with estimation and inference. Section 4 proposes practical methods and models to deal with time aggregation problems. Section 5 presents the example on inventory decisions. Section 6 summarizes and concludes. All the technical derivations are gathered in the Appendix.

2. Time Scale Transformation of Discrete-Time Models

This section studies the transformation of a generic discrete-time ARMA process into the corresponding aggregate-time process. We begin by introducing the notation and framework to be used hereafter and then derive the conditional generating mechanisms with a simple example. The formal derivation of the general results can be found in the

appendix A.

Consider a generic stochastic process that evolves in *original-time* τ, namely, $y = \{y_\tau\}_{\tau=1}^{\infty}$. The available data, however, are the realizations of a different process, $\mathsf{x} = \{\mathsf{x}_t\}_{t=1}^{\infty}$, whose elements are functions of those of y and is said to evolve in *aggregate-time* t. As an example, if y is a quarterly *stock* variable observed at an annual frequency, then

$$\mathsf{x}_1 = y_4, \mathsf{x}_2 = y_8, \mathsf{x}_3 = y_{12}, ... \tag{2.1}$$

This type of aggregation will be referred to as *point-in-time sampling*. If we had considered a *flow* variable instead, the process for x would be

$$\mathsf{x}_1 = y_4 + y_3 + y_2 + y_1, \mathsf{x}_2 = y_8 + y_7 + y_6 + y_5, \mathsf{x}_3 = y_{12} + y_{11} + y_{10} + y_9, ... \tag{2.2}$$

Sometimes, the aggregated data could be the result of a weighted average of the original variable, possibly with time-varying weights. In general, let $w_{j,i}$ denote the weight corresponding to the aggregate-time period j for the i^{th} power of the original-time, lag operator, $Z^i y_\tau = y_{\tau-i}$. For example, if yearly data were generated by taking weighted averages of quarterly data, we could use the previous notation to indicate this as

$$\begin{aligned} \mathsf{x}_1 &= \left(w_{1,0}Z^0 + w_{1,1}Z^1 + w_{1,2}Z^2 + w_{1,3}Z^3\right) y_4 \\ &= w_{1,0}y_4 + w_{1,1}y_3 + w_{1,2}y_2 + w_{1,3}y_1; \\ \mathsf{x}_2 &= \left(w_{2,0}Z^0 + w_{2,1}Z^1 + w_{2,2}Z^2 + w_{2,3}Z^3\right) y_8 \\ & w_{2,0}y_8 + w_{2,1}y_7 + w_{2,2}y_6 + w_{2,3}y_5; \\ & ... \end{aligned} \tag{2.3}$$

Assuming that the weights are constant, as it is done in traditional research on time aggregation, it is easy to see that each of the $w_{j,i} = \frac{1}{4}$ in expression (2.3). We will refer to this type of aggregation as *phase-averaged sampling*.

The examples presented in expressions (2.1)-(2.3) correspond to a constant frequency of aggregation, k, which is fixed at $k = 4$ in the examples (four quarterly observations per year). In this paper we depart from this traditional case by allowing k to vary over time, such as when it is the realization of a stochastic process. Therefore, we will endow k with a time subscript and denote it as k_t. Notice that k_t is the number of original-time

intervals in the aggregate-time period t. Using the notational elements introduced thus far, it is natural to express the process x with the following generic notation,

$$\mathsf{x} = \{\mathsf{x}_t\}_{t=1}^{\infty} = \left\{W_t(Z)y_{\varphi(t)}\right\}_{t=1}^{\infty}; \qquad \varphi(t) = \sum_{j=1}^{t} k_j. \tag{2.4}$$

so that

$$\begin{aligned}
\mathsf{x}_1 &= W_1(Z)y_{k_1} = \sum_{i=0}^{k_1-1} w_{1,i}y_{(k_1-i)}, \\
\mathsf{x}_2 &= W_2(Z)y_{(k_1+k_2)} = \sum_{i=0}^{k_2-1} w_{2,i}y_{(k_1+k_2)-i}, \\
&\ldots
\end{aligned}$$

The general mapping from y to x described in expression (2.4) will allow us to treat time aggregation comprehensively and will be used to derive the representation of the DGP for x given that of y, for a general aggregation scheme.

Although one can derive the density of x from the finite dimensional cumulative density of y using standard techniques for the linear transformation of random variables (see Mood et al., 1974), this derivation is intractable in practice because it involves high-dimensional marginalization and integration. In this paper we follow an approach that is common in the literature, and assume that the original-time process y follows a generic ARMA process, and then derive the DGP for x using algebraic methods instead. Brewer (1973), Wei (1981), Weiss (1984) and Marcellino (1999) provide results for k constant. Here, we extend these results to allow for a time-varying k, the aggregation frequency.

Consider a simple example of an AR(1) process in original-time

$$y_\tau = \rho y_{\tau-1} + e_\tau, \quad e_\tau \sim i.i.d.(0, \sigma_e^2), \tag{2.5}$$

and assume that y is a stock variable and that the aggregation frequency is $k = 4$ with point-in-time sampling, so that the aggregate-time process x is as in (2.1). Without loss of generality, pre-multiply expression (2.5) by the auxiliary polynomial $(1 + \rho Z + \rho^2 Z^2 + \rho^3 Z^3)$ to obtain

$$y_\tau = \rho^4 y_{\tau-4} + e_\tau + \rho e_{\tau-1} + \rho^2 e_{\tau-2} + \rho^3 e_{\tau-3}, \tag{2.6}$$

thus allowing one to express x in aggregate-time as

$$\mathsf{x}_t = \psi \mathsf{x}_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \sigma^2), \tag{2.7}$$

with the obvious correspondences, $\psi = \rho^4$ and $\sigma^2 = (1 + \rho^2 + \rho^4 + \rho^6)\sigma_e^2$.

Relaxing the assumption that k is constant means we now have to premultiply both sides of (2.5) by a different auxiliary polynomial at each time period t. Specifically, when $t = 1$ and $\tau = k_1$, the polynomial is $(1 + \rho Z + ... + \rho^{k_1 - 1} Z^{k_1 - 1})$; when $t = 2$ and $\tau = k_1 + k_2$, the polynomial is $(1 + \rho Z + ... + \rho^{k_2 - 1} Z^{k_2 - 1})$; and so on. Therefore, the aggregate-time equivalent expression to (2.7) is now

$$\mathsf{x}_t = \psi_t \mathsf{x}_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \sigma_t^2), \tag{2.8}$$

with $\psi_t = \rho^{k_t}$ and $\sigma_t^2 = (1 + \rho^2 + ... + \rho^{2(k_t - 1)})\sigma_e^2$.

Several conclusions can be drawn from this simple example, which are properly generalized in appendix A. First, the coefficients of the aggregated process x are time-varying whenever k_t is variable. Second, although the order of the autoregressive polynomial is typically preserved, the aggregate-time process x will often have an MA component as well. For a point-in-time sampling scheme, the MA component will usually be of order $p-1$, where p is the AR order in original-time, while under phase-averaging the usual order is p (lower/higher values can be obtained when $p - q > k_t/q - p \geq k_t$, where q is the MA order in original-time). Third, these properties apply to stationary, integrated or even explosive original-time processes. Finally, while we derive all of the results assuming a variable k_t, the traditional results found in the literature with k constant can be obtained as special cases of our framework.

These conclusions raise a number of issues that we tackle below. In particular, the AR(1) example suggests that the aggregate-time process will have time-varying parameters and time-varying volatility. Both of these features have important implications from the point of view of specification and testing. It suggests that constant parameter specifications will provide unsatisfactory approximations to the structural parameters of the model and will forecast poorly – the conditional information relating k_t is unused. Even if we momentarily entertained that a constant parameter specification provided adequate estimates, the time-varying nature of the residual variance suggests one needs to be particularly careful in making inference. The next section derives the maximum likelihood estimator (MLE) for the aggregate-time process and establishes its asymptotic distribution. However, it is important to transcend the ideal theoretical results

into practical solutions of general time aggregation problems. This is the emphasis of section 4.

3. Maximum Likelihood Estimation

This section sketches the derivation of the maximum likelihood estimators for the parameters of the aggregate-time DGP, assuming that the sequence $\{k_t\}$ is observed. Appendix B provides full details and discusses conditions for the asymptotic normality of the estimators. When $k_t = k$, $\forall t$, the aggregate model will be an ARMA model with constant parameters, for which traditional modelling and estimation results are readily available. Thus, here we concentrate on situations where k_t fluctuates over time instead – this results in models with time-varying parameters, as we have seen in the previous section.

Let $\mathsf{X}_{t-1} = \{\mathsf{x}_{t-1}, \mathsf{x}_{t-2}, ...\}$, $K_{t-1} = \{k_{t-1}, k_{t-2}, ...\}$, and assuming fixed initial conditions, the joint likelihood for a sample of T observations can be written as the product of the conditional densities,

$$L(\Xi) = \prod_{t=p}^{T} f(\mathsf{x}_t | \mathsf{X}_{t-1}; K_{t-1}; \zeta_t). \tag{3.1}$$

However, this expression is ill suited for the derivation of the ML estimators for $\{\widehat{\zeta_t}$, $t = p, ..., T\}$ since the number of parameters is typically larger than the number of available observations. Alternatively, the likelihood can be reparametrized in terms of the native parameters of the original-time process y, say θ, and then the maximum likelihood estimators of θ can be used to obtain those of ζ_t by exploiting the mapping from θ to ζ_t derived in Appendix A.

Appendix B proposes a general Kalman filter based approach for the derivation of $L(\Xi(\theta)) = L(\theta)$, that is, the expression of the joint likelihood function in terms of the orginal-time parameters θ. This approach essentially consists in casting the original-time and aggregate-time ARMA processes in state space form. Then one can write the Kalman filter equations, derive the prediction errors, and use them to construct the likelihood function. These derivations are based on Harvey (1989, Ch. 6), and extend his results to the case of a time varying aggregation frequency and generic aggregation weights. The resulting likelihood is an expression that depends on the original-time parameters but is based on the aggregate-time data.

The next section explores alternative practical modeling strategies when the stochastic nature of k_t is also explicitely taken into consideration.

4. Practical Modelling Strategies

A transformation of the time scale from τ-time to t-time will yield an aggregate-time process x_t and a sequence $\{k_t\}$ corresponding to the variable frequency of aggregation. The next subsection assumes that both x_t and k_t are observable and that the practitioner's task is to estimate the parameters of interest by *jointly* modelling these two stochastic processes. At first, it may seem unrealistic to assume that k_t is observed, however, it is common in finance to operate with data that evolve at different time frequencies (for example, quotes, trades, financial information and macroeconomic information are recorded anywhere from a few seconds on average, all the way up to quarterly frequency). One solution is to aggregate all the data into a common time scale and use well established multivariate techniques for the analysis. In such a context, although it has been customary to disregard the manner in which the data is aggregated, there is nothing that prevents practitioners from recording the variable k_t. In fact, in some of our own work (Jordà and Marcellino, in press) we have found this variable to be very useful indeed.

The second subsection discusses methods for modelling x_t when the k_t are unobservable instead. This scenario is more closely related to the classical discussion of time aggregation in the literature. However, the distinct possibility that in certain contexts k_t cannot be regarded as constant, suggests that one cannot rely on traditional models but instead one has to rely on specifications that can accommodate the time-varying nature of the conditional mean and variance coefficients. As we will see, one practical solution is to rely on Hamilton's (1989) Markov switching-regimes model.

4.1. Stochastic and observable k_t : The ACI Model

Given observations on the aggregate-time process x_t and the aggregation frequency k_t, it will be of interest to model their joint distribution conditional on past information, the distribution of y, and the aggregation weighting polynomial $W_t(Z)$. Since k_t represents the number of original-time observations over which aggregation takes place at aggregate-time t, it is natural to think of k_t as taking a finite number of integer val-

ues so that $k_t \in \{0, 1, 2, ..., N\}$. Momentarily entertain the simplifying assumption that $P(k_t = j|\mathsf{X}_{t-1}, K_{t-1}; \theta_k) = P(k_t = j) = p_j$, then the joint distribution of x_t and k_t can be factored by the product rule of probability as

$$f(\mathsf{x}_t, k_t = j|\mathsf{X}_{t-1}, K_{t-1}; \theta) = g(\mathsf{x}_t|k_t = j, \mathsf{X}_{t-1}, K_{t-1}; \theta_x) \cdot p_j \tag{4.1}$$

Assuming the x_t are Gaussian, their conditional distribution becomes

$$g(\mathsf{x}_t|k_t = j, \mathsf{X}_{t-1}, K_{t-1}; \theta_x) = \frac{1}{\sqrt{2\pi}\sigma(j)_t} exp\left\{\frac{-(\mathsf{x}_t - \mu(j)_t)^2}{2\sigma(j)_t^2}\right\} \tag{4.2}$$

where $\mu(j)$ and $\sigma(j)$ are indexed by j to indicate that they depend on the value of the frequency of aggregation for that period as in expression (2.8), for example. Given expressions (4.1) and (4.2), the joint likelihood of the data can be expressed as

$$f(\mathsf{x}_t, k_t = j|\mathsf{X}_{t-1}, K_{t-1}; \theta) = \frac{p_j}{\sqrt{2\pi}\sigma(j)_t} exp\left\{\frac{-(\mathsf{x}_t - \mu(j)_t)^2}{2\sigma(j)_t^2}\right\} \tag{4.3}$$

If k_t is observable but with a one period delay instead, then the unconditional density of x_t has to be used. The end result will be a mixture of normal distributions, which is obtained by summing up (4.3) over all the possible values of k_t, specifically

$$P(\mathsf{x}_t|\mathsf{X}_{t-1}, K_{t-1}; \theta_x) = \sum_{j=0}^{N} f(\mathsf{x}_t, k_t = j|\mathsf{X}_{t-1}, K_{t-1}; \theta_x). \tag{4.4}$$

Expressions (4.3) and (4.4) describe the basic intuition behind the steps necessary to jointly model x_t and k_t so it is time to become more ambitious with regard to the simplifying assumption $P(k_t = j) = p_j$. As we discussed, k_t records the number of original-time observations aggregated per aggregate-time interval t and is best thought of as an integer-valued variable with positive support. Therefore, a natural distributional assumption for this type of variable is the Poisson distribution. For example, the typical Poisson regression framework would specify the conditional mean (or intensity) of this Poisson process, say λ_t, as a function of x_{t-1} with the following simple expression,

$$\begin{aligned} P(k_t &= j|\mathsf{X}_{t-1}, K_{t-1}; \theta_k) = \frac{e^{-\lambda_t}\lambda_t^j}{j!} \\ \log(\lambda_t) &= \omega + \delta \mathsf{x}_{t-1} \end{aligned} \tag{4.5}$$

thus ensuring that the parameter space (in this case ω and δ) is unconstrained. Such a model has a long tradition and its MLE properties are well developed. Expression (4.5) is not completely satisfactory because it restricts the nature of the time series process for k_t. Consequently, one could consider adding lags of both k_t and x_t into more general expressions of (4.5). In Jordà and Marcellino (in press) we have had success with a more general specification that we denominated the *autoregressive conditional intensity* model (ACI). This specification seems more natural for a time series process and for a one-lag specification, it consists in specifying the conditional mean of the Poisson process as

$$log(\lambda_t) = \omega + \alpha log(\lambda_{t-1}) + \beta k_{t-1} + \delta \mathsf{x}_{t-1} \qquad ACI(1,1) \tag{4.6}$$

This specification of the conditional intensity function ensures that the conditional mean remains strictly positive without restricting the parameter space. The term $log(\lambda_{t-1})$ parsimoniously endows the conditional mean with an exponentially declining dependence on infinite lags of k_t and x_t and can be seen as a natural analog to a typical ARMA model in the linear time series tradition. Thus, stationarity requires that $|\alpha + \beta| < 1$, when $\delta = 0$. Jordà and Marcellino (in press) apply this ACI model to explain the behavior of price-quote spreads in the foreign exchange market as indicators of market liquidity. We find that by allowing the parameters of the time series process that explains the size of the price-quote spread to depend on the arrival intensity of these quotes, the model fit improves significantly relative to competing specifications. We attribute this improved performance in large part on being more careful in accounting for the type of aggregation that is the central topic of this paper. The MLE for the ACI model is disarmingly simple and we refer the reader to that paper for more details. Once the marginal density of k_t has been specified with the ACI model for example, one can rely on the product rule of probability and exogeneity arguments (see Engle et al., 1983) to estimate the conditional model for x_t (given by aggregation formulae similar to (2.8)) and the marginal model for k_t separatly, giving consistent and efficient estimates of all the parameters.

4.2. Stochastic, non-observable k_t : The Markov Switching-Regimes Model

In many situations the practitioner will observe x_t but not k_t, yet suspect that the observed data are the result of time aggregation with variable frequency. This would

prevent us from directly using the ACI model introduced above. Instead, we need a flexible model that will allow for variation in the parameters of the conditional mean and variance, dependence in the unobserved frequency of aggregation k_t, yet be restrictive enough that the model can be estimated in practice. A specification that meets all of these requirements is based on assuming that the integer-valued k_t can take on a small number of values, say $k_t \in \{1, 2, ...N\}$ such that $P(k_t = j|\mathsf{X}_{t-1}, K_{t-1}; \theta_k)$ is an N-state Markov chain, that is

$$P(k_t = j|\mathsf{X}_{t-1}, K_{t-1}; \theta_k) = P(k_t = j|k_{t-1} = i) = p_{ij} \tag{4.7}$$

for $i, j = 1, 2, ..., N$. This assumption can be generalized as in Lam (1990), Durland and McCurdy (1994), and Diebold, Lee and Weinbach (1994) but we restrict our attention to the basic formulation for clarity. Based on (4.3) and (4.7), it is immediately apparent that this specification can be viewed as a special case of the popular Markov switching-regimes (MSR) model proposed in Hamilton (1989), and the doubly stochastic model proposed by Tjøstheim (1986). A simple example will illustrate the particulars of this correspondence.

Consider the following, original-time, ARMA(2,0) model

$$y_\tau = \rho_1 y_{\tau-1} + \rho_2 y_{\tau-2} + \varepsilon_\tau \qquad \varepsilon_\tau \sim N(0, \sigma_\varepsilon^2) \tag{4.8}$$

and assume $k_t = \{1, 2\}$, that is, for every original-time period there is some probability that the corresponding observation will be recorded or that it will be skipped. The two-state Markov chain that describes k_t is $P(k_t = j|k_{t-1} = i) = p_{ij}$ for $i, j = 1, 2$. Consequently, the resulting aggregated-time process is as follows for each of the four possible combinations of events:

for $k_t = 1$, and $k_{t-1} = 1$,

$$\begin{aligned} \mathsf{x}_t &= \rho_1 \mathsf{x}_{t-1} + \rho_2 \mathsf{x}_{t-2} + u_t \qquad ARMA(2,0) \\ u_t &= \varepsilon_\tau; \qquad E(u_t) = 0; \qquad E(u_t^2) = \sigma_\varepsilon^2 \end{aligned}$$

for $k_t = 1$, and $k_{t-1} = 2$,

$$\begin{aligned} \mathsf{x}_t &= (\rho_1^2 + \rho_2)\mathsf{x}_{t-1} + \rho_1\rho_2 \mathsf{x}_{t-2} + u_t \qquad ARMA(2,0) \\ u_t &= \varepsilon_\tau + \rho_1 \varepsilon_{\tau-1}; \qquad E(u_t) = 0; \qquad E(u_t^2) = \sigma_\varepsilon^2(1 + \rho_1^2) \end{aligned} \tag{4.9}$$

for $k_t = 2$, and $k_{t-1} = 1$,

$$\begin{aligned} \mathsf{x}_t &= \left(\frac{\rho_1^2 + \rho_2}{\rho_1}\right)\mathsf{x}_{t-1} - \frac{\rho_2}{\rho_1}\mathsf{x}_{t-2} + u_t - \frac{\rho_2}{\rho_1}u_{t-1} \qquad ARMA(2,1) \qquad (4.10) \\ u_t &= \varepsilon_\tau; \qquad E(u_t) = 0; \qquad E(u_t^2) = \sigma_\varepsilon^2 \end{aligned}$$

and for $k_t = 2$, and $k_{t-1} = 2$,

$$\begin{aligned} \mathsf{x}_t &= (\rho_1^2 + 2\rho_2)\mathsf{x}_{t-1} - \rho_2^2\mathsf{x}_{t-2} + u_t - \rho_2 u_{t-1} \qquad ARMA(2,1) \qquad (4.11) \\ E(u_t) &= 0; \qquad E(u_t^2) = \sigma_\varepsilon^2(1 + \rho_1^2 + \rho_2^2) \end{aligned}$$

Following Hamilton (1994), define the new variable s_t which characterizes the regime at date t as follows

$$\begin{aligned} &s_t = 1 \text{ if } k_t = 1 \text{ and } k_{t-1} = 1, \qquad s_t = 2 \text{ if } k_t = 2 \text{ and } k_{t-1} = 1, \\ &s_t = 3 \text{ if } k_t = 1 \text{ and } k_{t-1} = 2, \qquad s_t = 4 \text{ if } k_t = 2 \text{ and } k_{t-1} = 2, \end{aligned}$$

so that, given our assumptions, s_t follows a four state Markov chain with transition matrix

$$P = \begin{bmatrix} p_{11} & 0 & p_{11} & 0 \\ p_{12} & 0 & p_{12} & 0 \\ 0 & p_{21} & 0 & p_{21} \\ 0 & p_{22} & 0 & p_{22} \end{bmatrix}.$$

Appendix C contains the expression for the conditional densities and the form that the estimation algorithm proposed by Hamilton (1994) takes in this case. However, except for the parametric restrictions implied by time aggregation on the coefficients of the conditional mean and variance, estimation of the model poses no additional difficulties. The next section illustrates some of the nuances of the theoretical discussion carried thus far with a simple example.

5. An example: Time aggregation and structural inference

Time aggregation issues are not popular with empirical practitioners. The usual justification for this attitude is that the DGP is never observed and therefore, econometric

models need always be approximations to the true model. Thus, time aggregation is but one of the dimensions of this approximation problem. Furthermore, the prescriptions deriving from traditional time aggregation studies amount to few useful empirical remedies. From the point of view of forecasting, there is little advantage in knowing the process is time-aggregated since conditional mean forecasts still rely on constant parameter specifications that are well understood. Once one accounts for the additional serial correlation features induced by time aggregation, little else can be done.

In contrast, we have introduced the notion that time aggregation can occur over a variable number of original-time observations. Such a scenario has significant implications for empirical work, as we have justified in the previous sections. The resulting aggregated processes will exhibit time-varying parameters in the conditional mean and variance, even if these features were absent in the original-time process. Consequently, we have recommended ways in which to specify statistical models that account for these features. Accounting for the variability in the aggregation frequency is helpful in improving forecast performance and is important for inference. In this section we put these ideas to work with a simple version of an inventory control problem.

The example we analyze is based on a classical stock-adjustment model (see Caballero and Engel, 1993) of the form

$$z_\tau = \mu + (1-\alpha) z_{\tau-1} + \varepsilon_\tau \qquad \varepsilon_\tau \overset{iid}{\sim} WN(0, \sigma^2); \qquad \alpha \in [0,1] \tag{5.1}$$

where z_τ denotes a disequilibrium variable, whose specific definition will become clear momentarily; and α is the speed of adjustment parameter. For example, if this adjustment process is an (S,s) type of adjustment, then $\alpha = 1$ but the timing of such adjustments will be stochastic – it will be determined by the crossing of the barriers (S,s). If the adjustment process is linear/quadratic instead, then $\alpha < 1$, depending on the underlying nature of the adjustment costs. The model in expression (5.1) has a long tradition and has been applied widely to explain inventory behavior, investment dynamics, short-run changes in employment, pricing policies, and other economic phenomena.

The example we investigate here is an inventory model based on glass container data from the Census Bureau's monthly survey Manufacturers' Shipments, Inventories and Orders – also known as the M-3 report. The sample ranges from January 1991

to December 2001 and is not seasonally adjusted. Three main advantages justify our choice: (1) the data is based on 16 manufacturing plants which diminishes cross-sectional aggregation issues; (2) the data is reported in physical units of one thousand gross (one thousand gross = 144,000) rather than in dollars, thus avoiding accounting and questionable valuation methods that tend to introduce measurement error; and (3) the data are not seasonally adjusted, which unlike most economic applications, is preferable when investigating inventory behavior.

A natural definition for z in this context is the ratio of inventories to shipments, which we will adopt here. Figure 1 displays this ratio for the whole sample of 132 observations. Estimation of expression (5.1) by conventional methods in aggregate-time, without further considerations, yields the following estimates

$$z_t = \underset{(0.12)}{0.54} + \underset{(0.06)}{0.71}\, z_{t-1} + 0.18 u_t \qquad u_t \sim N(0,1) \tag{5.2}$$

with a Durbin-Watson statistic of 2.13 and a log-likelihood value of 43.50 (under the assumption of Gaussianity). These estimates imply an estimate of $\widehat{\alpha} = 0.29$ or 29% adjustment per period, which is considerably higher than other estimates of inventory behavior in the literature (Jordà 1999 and references therein report values in the neighbourhood of 5%), but much closer to what economic theory would predict.

Based on economic fundamentals, there are strong reasons to suspect that the original-time scale and the observed-time scale do not coincide. If inventories are adjusted according to (S,s) rules for example, then the observed data will be aggregated over time-varying intervals of original-time observations. Accordingly, it is natural to experiment with some of the solutions advanced in section 4. Specifically, since we do not observe the frequency of aggregation, we will use the Markov switching-regimes model presented in subsection 4.2. Furthermore, since we are interested in showing that the coefficients of the model adhere to the predictions of time aggregation, we do not impose the cross-regime coefficient restrictions described in that subsection. Consequently, we estimated a version of the model with three states (because we have 132 observations, we have to be careful not to estimate regimes with very few observations), whose estimates we report below,

$$
\begin{aligned}
\text{State 1:} \quad & z_t = \underset{(0.17)}{0.13} + \underset{(0.09)}{0.99}\, z_{t-1} + \underset{(0.12)}{0.37}\, u_t \\
\text{State 2:} \quad & z_t = \underset{(0.08)}{0.39} + \underset{(0.05)}{0.72}\, z_{t-1} + \underset{(0.09)}{0.17}\, u_t \\
\text{State 3:} \quad & z_t = \underset{(0.13)}{0.60} + \underset{(0.06)}{0.58}\, z_{t-1} + \underset{(0.12)}{0.33}\, u_t
\end{aligned}
\tag{5.3}
$$

with log-likelihood 60.83, substantially larger than the value of 43.50 obtained for the estimates in (5.2). The estimates in (5.3) suggest that state 1 correspond to aggregate-time periods in which there are no adjustments, so that $\widehat{\alpha} = 0$. State 2 suggests these periods correspond to one original-time adjustment, with $\widehat{\alpha} = 0.28$ or 28%. State 3 corresponds to two original-time adjustments. If this regime were to be the result of time aggregation, as we have hypothesized, then we would expect the coefficient estimate on z_{t-1} to be $0.72^2 = 0.52$, which is indeed within a conventional 95% confidence interval around the estimated value of 0.58. Notice also that the implied estimate for α from state 3 is $\widehat{\alpha} = 0.24$ or 24% (derived from $(1-\widehat{\alpha})^2 = 0.58$), very close to the 28% estimate from state 2.

The results of the Markov switching-regimes model are very encouraging and lend support to our views on time aggregation. However, the estimates of α under either the model in (5.2) or (5.3) are very similar, suggesting that there is little gain in going through the trouble of estimating a more complicated model, other than possibly for forecasting improvement (from jointly forecasting the state and the value of z_t). The explanation for this can be readily found by noticing that the steady-state probabilities for each regime imply that the average aggregation frequency is 0.87 original-time intervals per aggregate-time interval (this is calculated as 53% of the time there is no adjustment over the aggregate-time period; 7% of the time there is one adjustment; and 40% of the time there are two adjustments), which is rather close to a one-to-one match of original-time and aggregate-time intervals. In other situations or if the data had been reported at a quarterly frequency instead, we would expect a more significant mismatch between time-scales and therefore, a more significant role for the techniques we advocate.

6. Conclusions

Time scale transformations are quite common, since there is often a mismatch between the generation and the collection of the data. This mismatch poses serious problems for estimation of structural parameters, testing of hypotheses of interest, and forecasting with standard time series models. The effects can be even more dramatic when the frequency of aggregation varies over time, perhaps because it is itself a random variable.

In this paper we have highlighted these problems, but also suggested solutions by explicitly keeping into account the presence of time scale transformations. We have developed maximum likelihood techniques for estimation and inference on the original parameters of interest, suggested new models for the aggregated process – such as the autoregressive conditional intensity model – and proposed alternative explanations for adopting already existing nonlinear specifications, such as the Markov regime-switching model. An example on an inventory adjustment model highlights the potential for the techniques and considerations we have presented.

7. Appendix A - Aggregate DGP, ARMA case

Assume the original-time process y evolves according to a generic stochastic linear difference equation

$$\Phi(Z)y_\tau = \Psi(Z)\varepsilon_\tau \tag{7.1}$$

where $\Phi(Z) = 1 - \phi_1 Z - \phi_2 Z^2 - ... - \phi_p Z^p$; $\Psi(Z) = 1 - \psi_1 Z - \psi_2 Z^2 - ... - \psi_q Z^q$ and $\varepsilon_\tau \sim WN(0, \sigma^2)$.

The corresponding aggregate-time process, x, can be derived by finding, at each aggregate-time period t, the polynomial $B_t(Z)$, such that $B_t(Z)\Phi(Z) = C_t(L)$, just as we did to obtain expression (2.8). The MA component of x_t can then be easily derived from the autocovariance of $B_t(Z)\Psi(Z)\varepsilon_\tau$ in aggregate-time.

We introduce some additional notation to find the coefficients of the polynomial $B_t(Z)$. Define

$$\begin{aligned} \underset{g_t \times 1}{\gamma_t} &= (-\phi_1, -\phi_2, ..., -\phi_p, 0, ..., 0)' \\ \underset{b_t \times 1}{\beta_t} &= (\beta_{t,1}, \beta_{t,2}, ..., \beta_{t,b_t})' \end{aligned} \tag{7.2}$$

so that $B_t(Z) = \left(1 + \beta_{t,1}Z + \beta_{t,2}Z^2 + ... + \beta_{t,b_t}Z^{b_t}\right)$, with $b_t = \sum_{j=0}^{p-1} k_{t-j} - p$ and $g_t = b_t + p$. Next, define the rectangular matrix Γ_t

$$\underset{g_t \times b_t}{\Gamma_t} = \begin{bmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ -\phi_1 & 1 & 0 & ... & 0 & 0 & 0 \\ -\phi_2 & -\phi_1 & 1 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ -\phi_p & -\phi_{p-1} & -\phi_{p-2} & ... & ... & ... & ... \\ 0 & -\phi_p & -\phi_{p-1} & ... & ... & ... & ... \\ 0 & 0 & -\phi_p & & & & \\ ... & ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & 1 & 0 & 0 \\ ... & ... & ... & ... & -\phi_1 & 1 & ... \\ 0 & 0 & 0 & ... & -\phi_2 & -\phi_1 & 1 \end{bmatrix}. \tag{7.3}$$

and denote $\underset{b_t \times b_t}{\Gamma_t^*}$ and $\underset{b_t \times 1}{\gamma_t^*}$ as the counterparts to Γ_t and γ_t in expressions (**??**) and (7.3) obtained by deleting the rows k_{t-j} for $j = 0, 1, 2, ..., p-1$. This notation now permits us to introduce two propositions that characterize the dynamic properties of aggregate-time processes.

Proposition 1. *If y is the original-time process in expression (7.1), $k = \{k_t\}_{t=1}^{\infty}$ and x is the aggregate time process obtained from a point-in-time sampling scheme, such that $\mathsf{x} = \{\mathsf{x_t}\}_{t=1}^{\infty} = \{y_{k_1}, y_{(k_1+k_2)}, y_{(k_1+k_2+k_3)}, ...\}$ then x follows the linear stochastic difference equation*

$$C_t(L)\mathsf{x_t} = \mathsf{H_t(L)}\upsilon_{\mathsf{t}} \qquad \upsilon_{\mathsf{t}} \sim \mathsf{WN}(\mathsf{0}, \xi_{\mathsf{t}}^2) \tag{7.4}$$

The coefficients of $C_t(L) = (1 - c_{t,1}L - c_{t,2}L^2 - ... - c_{t,p}L^p)$ are the k_{t-j+1} rows of $-\Gamma_t(\Gamma_t^*)^{-1}\gamma_t^* + \gamma_t$ for $j = 1, ..., p$ The coefficients of $H_t(L) = (1 - h_{t,1}L - ... - h_{t,r_t}L^{r_t})$ and ξ_t^2 are the solutions to the non-linear system

$$\begin{aligned} \sum_{i=0}^{r_t} h_{t,i}^2 \xi_{t-i}^2 &= \sum_{i=0}^{b_t+q} \pi_{t,i}^2 \sigma^2 \\ -h_{t,j}\xi_{t-j}^2 + \sum_{i=1}^{(r_t-j)} h_{(t-j),i}\xi_{t-j-i}^2 h_{t,(j+i)} &= -\pi_{t,l}\sigma^2 + \sum_{i=1}^{b_t+q-l} \pi_{(t-j),i}\sigma^2\pi_{t,(l+i)} \end{aligned} \tag{7.5}$$

for $j = 1, ..., r_t$, where the πs are the coefficients of $\Pi_t(Z) = B_t(Z)\Psi(Z)$; and $l = \sum_{m=1}^{j} k_{t+1-m}$.

Proof. First derive the AR component of x from that of y given that the aggregation scheme is point-in-time. As mentioned before, for each period t, we want to find a polynomial, $B_t(Z)$, such that

$$B_t(Z)\Phi(Z) = 1 - c_{t,1}Z^{k_t} - c_{t,2}Z^{(k_t+k_{t-1})} - ... - c_{t,p}Z^{(k_t+k_{t-1}+...+k_{t-p+1})} = C_t(L) \tag{7.6}$$

Without placing any restriction on $B_t(Z)$, the coefficient of Z^i in $B_t(Z)\Phi(Z)$ coincides with the i^{th} element of the vector $\Gamma_t\beta_t + \gamma_t$. In order for (7.6) to hold, so that the appropriate coefficients in $B_t(Z)\Phi(Z)$ are zero, it must be that $\Gamma_t^*\beta_t+\gamma_t^* = 0$ from where it follows that the coefficients of $B_t(Z)$ are $\beta_t = -(\Gamma_t^*)^{-1}\gamma_t^*$. Notice that the columns of Γ_t^* are linearly independent so this matrix is full rank and its inverse always exists. The coefficients of $C_t(L)$ are the rows k_{t-j+1} of the vector $\Gamma_t\beta_t + \gamma_t = -\Gamma_t(\Gamma_t^*)^{-1}\gamma_t^* + \gamma_t$ for $j = 1, ..., p$. Therefore, in general, the order of $C_t(L)$ (the AR component of x) is at most p, the same as the order of $\Phi(Z)$ (the AR component of y).

Next, derive the MA component by defining the following variables

$$\begin{aligned} \zeta_t &= B_t(Z)\Psi(Z)\varepsilon_{(b_t+q)} = \Pi_t(Z)\varepsilon_{(b_t+q)} \\ \zeta_{t-1} &= B_{t-1}(Z)\Psi(Z)\varepsilon_{(b_{t-1}+q)} = \Pi_{t-1}(Z)\varepsilon_{(b_{t-1}+q)} \\ &... \end{aligned} \tag{7.7}$$

such that

$$\begin{aligned} cov(\zeta_t, \zeta_t) &= \sum_{i=0}^{b_t+q} \pi_{t,i}^2\sigma^2 \\ cov(\zeta_t, \zeta_{t-j}) &= -\pi_{t,l}\sigma^2 + \sum_{i=1}^{(b_t+q-l)} \pi_{(t-j),i}\sigma^2\pi_{t,(l+i)} \qquad \text{for } j = 1, ..., r_t \\ cov(\zeta_t, \zeta_{t-j}) &= 0 \qquad \text{for } j > r_t \end{aligned} \tag{7.8}$$

where in general $r_t = p-1$ (lower/higher values can be obtained when $p-q > k_t/q-p \geq k_t$). Expression (7.8) is the autocovariance function of a time-varying MA process. The corresponding autocorrelation function therefore has to be equal to that of the MA component in the generating mechanism of x and its coefficients have to satisfy (7.5). In practice, the MA coefficients can be obtained from the corresponding autocovariance function through a Kalman filter approach (see Hamilton, 1994, p. 391) or using the method in Wilson (1972).

Finally, it is easy to show that υ_t is the residual of a projection of ζ_t on $\upsilon_{t-1}, \upsilon_{t-2}, \upsilon_{t-3}, ...$ This ensures that the error terms from the aggregate-time-scale process are serially uncorrelated. The υ_t are random linear combinations of independently and identically distributed $WN(0, \sigma)$ random variables ε_τ. ■

Before analyzing aggregation by phase averaging, let γ_t, β_t and Γ_t be defined as in (7.2) and (7.3) but with $b_t = \sum_{j=0}^{p} k_{t-j} - p - 1$ instead. In addition, define

$$\underset{g_t \times 1}{\lambda_t} = \left(\frac{1}{k_t} e_{k_t}, -\frac{d_{t,1}}{k_{t-1}} e_{k_{t-1}}, -\frac{d_{t,2}}{k_{t-2}} e_{k_{t-2}}, ..., -\frac{d_{t,p}}{k_{t-p}} e_{k_{t-p}} \right)',$$

where e_{k_t} is a $1 \times k_t$ vector of ones and the $d_{t,i}$ are the coefficients of L^i in the aggregate AR polynomial $D_t(L)$. Let λ_t^* be the $b_t \times 1$ vector obtained by deleting the rows k_{t-j} of λ_t for $j = 0, 1, ..., p-1$. This notational considerations allow us to introduce the next proposition.

Proposition 2. *If x is generated by (7.1),* $k = \{k_t\}_{t=1}^{\infty}$, $W_t(L) = (1 + Z + Z^2 + ... + Z^{(k_t-1)})/k_t$ *and* $\mathsf{x} = \{\mathsf{x_t}\}_{t=1}^{\infty} = \left\{W_t(L) y_{\varphi(t)}\right\}_{t=1}^{\infty}$ *with* $\varphi(t) = \sum_{i=1}^{t} k_i \ \forall t$ *then*

$$D_t(L)\mathsf{x_t} = \mathsf{M_t(L)u_t} \qquad \text{for } \mathsf{u_t} \sim \mathsf{WN}(0, \nu_{\mathsf{t}}^2) \tag{7.9}$$

The coefficients of $D_t(L)$ *are the solutions to the linear system of* p *equations corresponding to the rows* k_{t-j+1} *of* $\Gamma_t(\Gamma_t^*)^{-1}(\lambda_t^* - \gamma_t) + \gamma_t = \lambda_t$ *for* $j = 1, ..., p$ *and* $\beta_t = (\Gamma_t^*)^{-1}(\lambda_t^* - \gamma_t^*)$. *The coefficients of* $M_t(L) = (1 - m_{t,1}L - ... - m_{t,s_t}L^{s_t})$ *are the solutions to the non-linear system*

$$\begin{aligned} \sum_{i=0}^{s_t} m_{t,i}^2 \nu_{t-i}^2 &= \sum_{i=0}^{b_t+q-l} \theta_{t,i}^2 \sigma^2 \\ -m_{t,j}\nu_{t-j}^2 + \sum_{i=1}^{s_t-j} m_{(t-j),i}\nu_{t-j-i}^2 m_{t,(j+i)} &= -\theta_{t,l}\sigma^2 + \sum_{i=1}^{b_t+q-l} \theta_{(t-j),i}\sigma^2\theta_{t,(l+i)}; \end{aligned} \tag{7.10}$$

for $j = 1, ..., s_t$, *where* $\Theta_t(Z) = B_t(Z)W_t(Z)\Psi(Z)$; *and* $l = \sum_{n=1}^{j} k_{t+1-n}$.

Proof. This time, we want to determine the polynomial $B_t(Z)$ such that

$$B_t(Z)\Phi(Z) = \begin{pmatrix} \omega_{k_t}(Z) - d_{t,1}Z^{k_t}\omega_{k_{t-1}}(Z) - d_{t,2}Z^{(k_t+k_{t-1})}\omega_{k_{t-2}}(Z) - ... \\ -d_{t,p}Z^{(k_t+...+k_{t-p+1})}\omega_{k_{t-p}}(Z) \end{pmatrix}$$

where $\omega_{k_t} = \left(\sum_{i=0}^{(k_t-1)} Z^i\right)/k_t$ so that $B_t(Z)\Phi(Z)y_\tau = D_t(L)\mathsf{x}_t$. For such a restriction to be satisfied, it must be that $\Gamma_t^*\beta_t + \gamma_t^* = \lambda_t^*$ which requires $\beta_t = (\Gamma_t^*)^{-1}(\lambda_t^* - \gamma_t^*)$. This in turn implies that $\Gamma_t(\Gamma_t^*)^{-1}(\lambda_t^* - \gamma_t) + \gamma_t = \lambda_t$. Now we need to determine the coefficients of $D_t(L)$. Given the expressions for λ_t and λ_t^* this can be accomplished by solving the linear system of p equations which correspond to the rows k_{t-j+1} of $\Gamma_t(\Gamma_t^*)^{-1}(\lambda_t^* - \gamma_t^*) + \gamma_t = \lambda_t$, $j = 1, .., p$. The proof for the coefficients of the MA component is similar to that of Proposition 1. ■

When $k_t = k\ \forall t$, propositions 1 and 2 simplify to the results obtained by Brewer (1973), Wei (1981), Weiss (1984) and Marcellino (1999). Following Marcellino (1999), propositions 1 and 2 can be readily extended to multivariate processes as long as the aggregation frequency, $\{k_t\}_{t=1}^{\infty}$, is common to all the elements of the vector process.

8. Appendix B - ML estimation

This appendix describes ML estimation of the aggregate-time process x_t using the Kalman filter. We begin with the state space expression for the orginal-time ARMA(p,q) process in (7.1), namely

$$\begin{aligned} y_\tau &= z'\alpha_\tau\ , \quad \alpha_\tau = S\alpha_{\tau-1} + e_\tau\ , \\ E(\alpha_0) &= \alpha_0, \quad V(\alpha_0) = P_0, \quad E(e_\tau\alpha_0) = 0\ \forall\tau\ , \\ z' &= [1\ -\psi_1\ -\psi_2\ ...\ -\psi_{r-1}]\,, \\ S &= \begin{bmatrix} \phi_1 & \phi_2 & ... & \phi_{r-1} & \phi_r \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ ... & & & & \\ 0 & 0 & ... & 1 & 0 \end{bmatrix}, \quad e_\tau = \begin{bmatrix} \varepsilon_\tau \\ 0 \\ 0 \\ ... \\ 0 \end{bmatrix}, \end{aligned} \tag{8.1}$$

where $r = max(p, q+1)$, α_τ is an r-dimensional vector of state variables, $\tau = 1, ..., T$, and we further assume that ε_τ is Normally distributed.

Next, define the variables

$$s_i = \sum_{j=1}^{i} k_j, \quad i = 1, ..., N, \tag{8.2}$$

where N is the number of aggregate-time periods, $s_0 = 0$, $s_N = T$, and

$$\beta_{s_{i-1}+r_i} = \sum_{j=1}^{r_i} w_{i,r_i-j} y_{s_{i-1}+j}, \quad \beta_0 = 0, \quad r_i = 1, ..., k_i,$$

where w_{i,r_i-j} are the weights in $W_t(Z)$ in (2.5), so that

$$\begin{aligned} \beta_\tau &= \varphi_\tau \beta_{\tau-1} + z'\alpha_\tau = \varphi_\tau \beta_{\tau-1} + z'S\alpha_{\tau-1} + z'e_\tau, \\ \varphi_\tau &= \begin{cases} 0 & \tau = s_{i-1}+1, \\ 1 & \text{otherwise.} \end{cases} \end{aligned}$$

The state space representation for the aggregated process in τ time (SSR(τ)) is

$$\begin{aligned} y_\tau &= g'\gamma_\tau, \quad \gamma_\tau = D_\tau \gamma_{\tau-1} + R\eta_\tau, \quad \tau = s_i \ , \ i = 1, ..., N, \\ g' &= \left[\underset{1\times r}{0} \ 1 \right], \quad \gamma_\tau = [\alpha_\tau \ \beta_\tau]', \\ D_\tau &= \begin{bmatrix} S & 0 \\ z'S & \varphi_\tau \end{bmatrix}, \quad R = \begin{bmatrix} I & 0 \\ z' & 0 \end{bmatrix}, \quad \eta_\tau = \begin{bmatrix} e_\tau \\ 0 \end{bmatrix}. \end{aligned} \tag{8.3}$$

From SSR(τ) we can also derive a state space representation for the aggregated process in aggregate-time, i.e. in t time (SSR(t)). It is

$$\mathsf{x}_t = g'\gamma_t, \quad \gamma_t = \begin{bmatrix} S^{k_t} & 0 \\ z'(W_{k_t} - I) & 0 \end{bmatrix} \gamma_{t-1} + \begin{bmatrix} I & 0 \\ 0 & z' \end{bmatrix} \begin{bmatrix} \eta_t^\alpha \\ \eta_t^\beta \end{bmatrix},$$

with $W_j = \sum_{s=0}^{j} S^s$, $\eta_t^\alpha = \sum_{j=1}^{r_t} S^{r_t-j} e_{s_{t-1}+j}$, $\eta_t^\beta = \sum_{j=1}^{k_t} W_{k_t-r_t} e_{s_{t-1}+r_t}$.

Derivation of the ML estimators for the original-time parameters is made more convenient by adopting the SSR(τ) in (8.3). Defining the optimal estimators of γ_τ by c_τ, with covariance matrix Σ_τ, the Kalman filter equations are:

$$\begin{aligned} c_{\tau|\tau-1} &= D_\tau c_{\tau-1} \\ \Sigma_{\tau|\tau-1} &= D_\tau \Sigma_{\tau-1} D_\tau' + RQR' \\ c_\tau &= \begin{cases} c_{\tau|\tau-1} & \tau \neq s_i, \quad i = 1, ..., N \\ c_{\tau|\tau-1} + \Sigma_{\tau|\tau-1} g' h_\tau^{-1} g (x_\tau - g' c_{\tau|\tau-1}) & \text{otherwise} \end{cases} \\ \Sigma_\tau &= \begin{cases} \Sigma_{\tau|\tau-1} & \tau \neq s_i, \quad i = 1, ..., N \\ \Sigma_{\tau|\tau-1} + \Sigma_{\tau|\tau-1} g' h_\tau^{-1} g \Sigma_{\tau|\tau-1} & \text{otherwise} \end{cases} \\ c_0 &= \begin{bmatrix} \alpha_0 \\ 0 \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned} \tag{8.4}$$

where $h_\tau = g'\Sigma_{\tau|\tau-1}g$, and Q is the variance of η_τ. The relevant prediction errors are

$$v_\tau = y_\tau - \widehat{y}_{\tau|\tau-1} = g'(\gamma_\tau - c_{\tau|\tau-1}), \quad \tau = s_i, \quad i = 1, ..., N.$$

Hence, the likelihood can be written as

$$\log L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \log f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2N}\sum_{i=1}^{N}\log h_{s_i} - \frac{1}{2N}\sum_{i=1}^{N}\frac{v_{s_i}^2}{h_{s_i}}.$$

Maximization of this expression with respect to $\theta = (\phi_i,\ i = 1, ..., p,\ \psi_j,\ j = 1, ..., q,\ \sigma)$ yields the ML estimators of the parameters of the original-time model, $\widehat{\theta}$. The formulae in propositions 1 and 2 can then be used to recover the ML estimators of the parameters of the aggregated process.

The following assumption collects all the conditions required to derive the properties of the ML estimators $\widehat{\theta}$:

ML Assumptions

i) $\theta \in \Theta$ and Θ is a compact subset of R^k.

ii) $f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)$ is a random function continuously differentiable of order 2 on Θ a.s., $i = 1, 2, ...$.

iii) (a) $\{f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$, (b) $\{\nabla_\theta f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$, (c) $\{\nabla^2_\theta f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$ are a.s. Lipschitz-L_1.

iv) The elements of (a) $\{f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$, (b) $\{\nabla_\theta f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$, (c) $\{\nabla^2_\theta f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$ are near epoch dependent of size -1 on (Θ, ρ), where ρ is any convenient norm on R^k.

v) The elements of (a) $\{f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$, (b) $\{\nabla_\theta f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$, (c) $\{\nabla^2_\theta f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$ are $r-$dominated on Θ uniformly in $i = 1, 2, ...,\ r > 2$.

vi) The sequence $\{\overline{Q}_N(\theta)\}$=$\{N^{-1}\sum_{i=1}^{N} E(\log f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta))\}$, has identifiably unique maximizers $\{\theta^*\}$ on Θ, interior to Θ uniformly in N.

vii) Defining $\{Q_N(\theta)\}$=$\{N^{-1}\sum_{i=1}^{N}\log(f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta))\}$, then (a) $\{B_N^*\} = \{Var[N^{1/2}\nabla_\theta Q_N(\theta^*)]\}$, (b) $\{A_N^*\} = \{\nabla^2_\theta \overline{Q}_N(\theta^*)\}$ are uniformly positive definite. ■

Conditions i) and ii) guarantee the existence of an estimator $\widehat{\theta}$, such that

$$Q_N(\widehat{\theta}) = \inf_{\theta \in \Theta} Q_N(\theta), \quad a.s.$$

Conditions iii)-(a), iv)-(a), and v)-(a) impose, respectively, smoothness, memory, and moment conditions on $\{f(\mathsf{x}_i|\mathsf{X}_{i-1};\theta)\}$ to ensure that $Q_N(\theta) - \overline{Q}_N(\theta) \to 0$ a.s. uniformly in Θ (Gallant and White,1988, Theorem 3.18). Under the additional condition vi), $\widehat{\theta} - \theta^* \overset{a.s.}{\to} 0$ (Gallant and White, 1988, Theorem 3.19) which ensures the estimator $\widehat{\theta}$ is consistent for θ^*. Under the additional conditions iii)-(b), iv)-(b), v)-(b), and vii)-(a), the asymptotic distribution of $B_N^{*-1/2} N^{1/2} \nabla_\theta Q_N(\theta^*)'$ is $N(0, I_k)$ (Gallant and White, 1988, Corollary 5.5). Further conditions on the matrix of second derivatives in iii)-(c), iv)-(c), v)-(c) and vii)-(b), and from a mean value expansion of $\nabla_\theta Q_N(\widehat{\theta})$ around θ^*, ensure the asymptotic distribution of $B_N^{*-1/2} A_N^* N^{1/2}(\widehat{\theta} - \theta^*)$ is $N(0, I_k)$ (Gallant and White, 1988, Theorem 5.7). Moreover, given that the model is correctly specified and the information matrix equality holds, $A_N^{*1/2} N^{1/2}(\widehat{\theta} - \theta^*)$ is asymptotically distributed as $N(0, I_k)$ (White, 1994, Theorem 6.5). If we complement our original hypothesis of an ARMA process with i.i.d. normal errors for y_τ with the assumption that the eigenvalues of S in (8.1) are inside the unit circle, then the conditions ii) to v) and vii) are satisfied. Actually, x_t also follows an ARMA process, as we saw in Appendix A, and its conditional distribution is also normal.

Condition vi) deserves more discussion. It requires the parameters in original-time, θ, to be globally identifiable (Rothenberg, 1971) in aggregate-time. Standard conditions for identification of ARMA models, e.g. Hannan (1971), are necessary but not sufficient for vi) to hold, because temporal aggregation can transform globally identifiable parameters into locally identifiable or non identifiable parameters. For example, two AR(1) processes in original-time with parameters β and $-\beta$, after point-in-time sampling with $k_t = k$ and k even, are both transformed into an AR(1) with parameter β^k, so that only $|\beta|$ is identifiable. An even worse case is aggregation by point-in-time sampling of an MA(q) process with $k_t > q$: the aggregated process becomes white noise. See, e.g., Marcellino (1998) for conditions that preserve identification through temporal aggregation, and Hinnich (1999) for a discussion of aliasing in the frequency domain.

9. Appendix C - Aggregate Markov Switching model

We now discuss the details of the estimation of the Markov Switching model in Section 4.2.

The four conditional densities corresponding to each of the four states are given by

$$
\begin{aligned}
f(\mathsf{x}_t|\mathsf{X}_{t-1}, s_t &= 1, \theta) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} exp\left\{\frac{-\left(\mathsf{x}_t - \rho_1\mathsf{x}_{t-1} - \rho_2\mathsf{x}_{t-2}\right)^2}{2\sigma_\varepsilon^2}\right\} \\
f(\mathsf{x}_t|\mathsf{X}_{t-1}, s_t &= 2, \theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2(1+\rho_1^2)}} exp\left\{\frac{-\left(\mathsf{x}_t - (\rho_1^2+\rho_2)\mathsf{x}_{t-1} - \rho_1\rho_2\mathsf{x}_{t-2}\right)^2}{2\sigma_\varepsilon^2(1+\rho_1^2)}\right\} \\
f(\mathsf{x}_t|\mathsf{X}_{t-1}, s_t &= 3, \theta) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} exp\left\{\frac{-\left(\mathsf{x}_t - \left(\frac{\rho_1^2+\rho_2}{\rho_1}\right)\mathsf{x}_{t-1} + \frac{\rho_2}{\rho_1}\mathsf{x}_{t-2} + \frac{\rho_2}{\rho_1}u_{t-1}\right)^2}{2\sigma_\varepsilon^2}\right\} \\
f(\mathsf{x}_t|\mathsf{X}_{t-1}, s_t &= 4, \theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2(1+\rho_1^2+\rho_2^2)}} \\
&\quad exp\left\{\frac{-\left(\mathsf{x}_t - (\rho_1^2+2\rho_2)\mathsf{x}_{t-1} + \rho_2^2\mathsf{x}_{t-2} - \rho_2^2 u_{t-1}\right)^2}{2\sigma_\varepsilon^2(1+\rho_1^2+\rho_2^2)}\right\}
\end{aligned}
$$

Let η_t denote a 4×1 vector that collects the above four densities. Collect the conditional probabilities $P(s_t = l|I_t;\theta)$ for $l = 1,2,3,4$ in a 4×1 vector denoted $\widehat{\xi}_{t|t}$. Further, denote $\widehat{\xi}_{t+1|t}$ as a 4×1 vector whose l^{th} element represents $P(s_{t+1} = l|I_t,\theta)$. Hamilton (1994) shows that optimal inference and forecasts for each date t in the sample can be found by iterating on the following pair of equations

$$
\widehat{\xi}_{t|t} = \frac{\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right)}{\mathbf{1}'\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right)} \tag{9.1}
$$

$$
\widehat{\xi}_{t+1|t} = P \cdot \widehat{\xi}_{t|t} \tag{9.2}
$$

where $\mathbf{1}'$ is a 4×1 vector of ones and the symbol $\odot$ denotes element by element multiplication. Given a starting value, $\widehat{\xi}_{1|0}$ and an assumed value for the population parameter θ, one can iterate on (9.1) and (9.2) for $t = 2,3,...,T$ calculate the values of $\widehat{\xi}_{t|t}$ and $\widehat{\xi}_{t+1|t}$ for each date t in the sample.

Furthermore, Hamilton (1994) shows that the log-likelihood for the observed data evaluated at the value θ that was used to perform the iterations, can also be calculated

as a by-product of this algorithm from

$$L(\theta) = \sum_{t=2}^{T} \log f(\mathsf{x}_t | I_{t-1}, \theta), \quad f(\mathsf{x}_t | I_{t-1}, \theta) = \mathbf{1}' \left(\widehat{\xi}_{t|t-1} \odot \eta_t \right) \tag{9.3}$$

For a given θ, the value of the log-likelihood implied by that value of θ is given by (9.3). The value of θ that maximizes the log-likelihood can be determined numerically. Further details on the estimation algorithm just described, inference on the transition probabilities p_{ij}, and forecasting can be found in Hamilton (1994).

References

[1] Brewer, K. R. W. (1973) "Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX models", *Journal of Econometrics*, 1, 133-154.

[2] Burns, Arthur F. and Wesley C. Mitchell (1946) *Measuring Business Cycles,* New York: National Bureau of Economic Research.

[3] Caballero, Ricardo J. and Eduardo M. R. A. Engel (1993) "Microeconomic Adjustment Hazards and Aggregate Dynamics," *Quarterly Journal of Economics,* 108, 313-358.

[4] Diebold, Francis. X., Joon H. Lee and Gretchen C. Weinbach (1994) "Regime Switching with Time-Varying Transition Probabilities," in *Nonstationary Time Series Analysis and Cointegration*. Advanced Texts in Econometrics, C. P Hargreaves (Ed.), Oxford University Press: Oxford.

[5] Durland, J. Michael and Thomas H. McCurdy (1994) "Duration-Dependent Transitions in a Markov Model of U.S. GNP Growth," *Journal of Business Economics and Statistics*, 12(3), 279-88.

[6] Engle, Robert F., David F. Hendry and Jean François Richard (1983) "Exogeneity," *Econometrica*, 51, 277-304.

[7] Engle, Robert F. and Jeffrey R. Russell (1998) "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66(5), 1127-1162.

[8] Gallant, A.R. and H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford: Basil Blackwell.

[9] Geweke, John B. (1978) "Temporal Aggregation in Multivariate regression Model," *Econometrica*, 46, 643-662.

[10] Govaerts, Bernadette, David F. Hendry, and Jean François Richard (1995) "Encompassing in Stationary Linear Dynamic Models", *Journal of Econometrics*, 63(1), 245-270

[11] Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-84.

[12] Hamilton, James D. (1994) *Time Series Analysis*, Princeton, New Jersey: Princeton University Press.

[13] Hannan, E. J. (1971) "The Identification Problem for Multiple Equation Systems with Moving Average Errors," *Econometrica*, 39, 751-765.

[14] Harvey, Andrew C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

[15] Hinich, Melvin J. (1999) "Sampling Dynamical Systems," *Macroeconomic Dynamics*, 3, 602-609.

[16] Jordà, Òscar (1999) "Random Time Aggregation in Partial Adjustment Models," *Journal of Business Economics and Statistics*, 17(3), 382-395.

[17] Jordà, Òscar and Massimiliano Marcellino (in press) "Modeling High-Frequency FX Data Dynamics", *Macroeconomic Dynamics*, forthcoming.

[18] Lam, Pok-sang (1990) "The Hamilton Model with a General Autoregressive Component. Estimation and Comparison with Other Models of Economic Time Series," *Journal of Monetary Economics*, 26, 409-32.

[19] Marcellino, Massimiliano (1998) "Temporal Disaggregation, Missing Observations, Outliers, and Forecasting A Unifying Non-Model Based Approach", *Advances in Econometrics*, vol. 13, 181-202.

[20] Marcellino, Massimiliano (1999) "Some Consequences of Temporal Aggregation in Empirical Analysis", *Journal of Business and Economic Statistics*, 17(1), 129-136.

[21] Mood, A. M., Graybill F. A.. e Boes D.C. (1974), *Introduction to the Theory of Statistics*, New York: McGraw-Hill.

[22] Rothenberg, Thomas J. (1971) "Identification in Parametric Models", *Econometrica*, 39, 577-591.

[23] Sims, Christopher A. (1971) "Discrete Approximations to Continuous Time Lag Distributions in Econometrics," *Econometrica*, 38, 545-564.

[24] Stock, James H. (1987) "Measuring Business Cycle Time," *Journal of Political Economy*, 95, 1240-1261.

[25] Stock, James H. (1988) "Estimating Continuous Time Processes Subject to Time Deformation," *Journal of the American Statistical Association*, 83, 77-84.

[26] Telser, L. G. (1967) "Discrete Samples and Moving Sums in Stationary Stochastic Processes," *Journal of the American Statistical Association*, 62, 484-499.

[27] Tjøstheim, Dag (1986) "Some Doubly Stochastic Time Series Models," *Journal of Time Series Analysis*, 7, 51-72.

[28] Wilson, G. Tunnicliffe (1972) "The Factorization of Matricial Spectral Densities," *SIAM Journal on Applied Mathematics*, 23, 420-426.

[29] Wei, W. W. S. (1981) "Effect of Systematic Sampling on ARIMA Models," *Communications in Statistics - Theory and Methods*, 10, 2389-2398.

[30] Weiss, Andrew A. (1984) "Systematic Sampling and Temporal Aggregation in Time Series Models," *Journal of Econometrics*, 271-281.

[31] White, Halbert (1994) *Estimation, Inference and Specification Analysis*, Cambridge, Cambridge University Press.

Figure 1 – Ratio of Inventories to Shipments of Glass Containers. Sample, January 1991 to December, 2001, Not Seasonally Adjusted.