April 2000

Stochastic Processes Subject to Time-Scale Transformations: An Application to High-Frequency FX Data*

Abstract

This paper is a general investigation of temporal aggregation in time series analysis. It encompasses traditional research on time aggregation as a particular case and extends the analysis to irregular intervals of aggregation. The Data Generating Process is allowed to evolve at regular, deterministic-irregular or even stochastic intervals of time (*operational time*). The time scale of this process is then transformed to generate the *observational time* process. This transformation can be deterministic (such as the familiar aggregation of monthly data into quarters) or more generally, stochastic (such as aggregating stock market quotes by the hour). In general, the observational time model exhibits persistence, time-varying parameters and non-spherical disturbances. Consequently, we review detection, specification, estimation and structural inference in this context, provide new solutions to these issues, and apply our results to high frequency, FX data.

- *JEL Classification Codes:* C13, C22, C43
- *Keywords:* time aggregation, time-scale transformation, irregularly spaced data, autoregressive conditional intensity model.

Oscar Jordá
University of California, Davis
ojorda@ucdavis.edu

Massimiliano Marcellino
Università Bocconi
massimiliano.marcellino@uni-bocconi.it

*Corresponding author: Oscar Jordá, Department of Economics, U.C. Davis, One Shields Ave., Davis, CA 95616-8578. Phone: (530) 752 7021. Fax: (530) 752 9382. e-mail: ojorda@ucdavis.edu.

1 Introduction

This paper explores the following question: What happens to dynamic econometric analysis when the time scale at which the data is generated does not coincide with the time scale at which the data is observed? Traditional research on time aggregation is almost entirely concerned with the following two problems: (1) the consequences of fixed-interval time aggregation[1] (for example, data generated monthly but recorded quarterly), and (2) analysis of continuous-time models from fixed-interval observed data.[2] This paper, however, treats these lines of research as particular cases of a more general problem, namely, that the time scales native to the data generating process (D.G.P.) and the frequency of data recording possibly evolve at irregular (usually stochastic) intervals of time.

The motivation for this analysis is natural in economics. Theoretical models typically result in a set of dynamic equations that describe some aspect of economic behavior. The dynamic evolution of the economic system and the timing with which changes in the variables occur are therefore determined internally by this behavioral model. This is the time scale at which the data will be generated. We call this the *operational time scale*, data based scale or economic time, following the nomenclature in Stock (1987) and use the subscript τ to index variables. However, empirical validation of the hypotheses put forth by the economic model necessarily relies on available data. The timing with which the data is recorded will, in general, not coincide with the operational time scale. We denominate the frequency of data collection as the *observational time* scale and use the subscript t to index variables.

Depending on whether each of these time scales evolves at regular intervals of time (i.e., the time elapsed between realizations is a fixed time unit, such as a week, a month, a quarter, and so on) or irregular intervals (when the time elapsed between realizations follows a stochastic process

[1] See, e.g., Telser (1967), Brewer (1973), Wei (1981), Weiss (1984), Christiano and Eichenbaum (1987) and Marcellino (1999).

[2] See, e.g., Sims (1971), Geweke (1978), Bergstrom (1984) and Stock (1987).

instead), we contemplate four types of time aggregation

	Operational time-scale τ	**Observational 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 literature on time aggregation typically assumes that the true economic decision interval is finer than the data sampling interval and in addition, assumes that economic decision-making is done at fixed intervals. For example, the observational time scale is in quarters while the operational time scale is in months. This example of type I aggregation has been well studied in the literature. Examples of type II aggregation are common in finance, where time series are analyzed at daily or even weekly frequencies although transactions in these markets happen anywhere from a second apart to hours apart and beyond. More formally, Jordá (1999) shows that partial adjustment models naturally generate irregularly spaced data in operational time although empirical analyses necessarily rely on observational quarterly or monthly data at best.

The analysis of Friedman and Schwarz (1982) in which monthly and quarterly data are phase-averaged across different stages of the business cycle; the literature on time deformation introduced by Stock (1987); and situations in which the data have missing observations can be viewed as examples of type III aggregation. Finally, type IV aggregation is common in finance. To distinguish between uninformed and informed traders, tick by tick financial data are often "thinned" by some statistical procedure (see Engle and Russell (1998)), the result of which is a new series in which observational time is also irregularly spaced. The findings in this paper thus encompass the traditional results on time aggregation. Some of the cases studied can be interpreted as the discrete-time analog to Stock's (1987) time deformation model while previous work on unequally spaced data by Robinson (1977) and Dunsmuir (1983) can also be viewed as particular cases in our framework.

The most interesting results in the paper correspond to situations in which the frequency of aggregation is stochastic. This produces observational time processes that have time-varying parameters and non-spheric disturbances. We provide propositions to support these claims and introduce new methods to model the frequency of aggregation as a stochastic point process from the maximum likelihood principle. These models are well suited to capture the specific non-linearities introduced by generic transformations of the time scale, improving the estimation of structural parameters and providing more accurate forecasts. In addition, the autoregressive conditional intensity model (ACI) that we introduce here can be used in a general context to deal with dynamic count-data problems.

The paper is organized as follows. Sections 2 and 3 introduce the general framework we use, and present the theoretical results on representation, estimation, and inference. Section 4 analyzes in further detail the implications of the results in sections 2 and 3, and proposes practical methods and models to deal with time aggregation problems. Section 5 reviews the main characteristics of high frequency foreign exchange (FX) data, and illustrates how they can be generated by time scale transformations, both theoretically and by means of simulation experiments. Section 6 presents a simple dealer's inventory model to highlight the problems with structural inference in the presence of unaccounted time scale transformations. Section 7 applies the models of section 4 to study the behavior of the bid-ask spread in the US Dollar - Deutsche Mark FX market. Finally, Section 8 summarizes and concludes.

2 Time Scale Transformation of Discrete-Time Models

This section studies the transformation of a generic, operational, discrete-time ARIMA process into the corresponding observational time process. We begin by introducing the notation and framework to be used hereafter and then derive the conditional generating mechanisms. Consider a generic stochastic process that evolves in *operational time* τ, namely, $x = \{x_\tau\}_{\tau=1}^{\infty}$. 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 x. x is said to evolve in *observational time* t. The transformation from the operational time scale τ into the observational time scale t is given by

$$\tau = \varphi(t) = \varphi(k(t)) = \sum_{j=1}^{t} k_j \qquad \text{for } k = \{k_t\}_{t=1}^{\infty} \tag{1}$$

k is termed the *frequency of aggregation* which can be thought of as a sequence of numbers or more generally a stochastic process itself. Note that $\varphi(t) - \varphi(t-1) = k_t$, that is, the number of operational time observations per sampling interval $(t-1, t]$. In order to accommodate different types of aggregation, define the lagged polynomial $W_t(Z)$, termed the *aggregation scheme*, as follows

$$W_t(Z) = w_{t,0} + w_{t,1}Z + w_{t,2}Z^2 + ... + w_{t,(k_t-1)}Z^{(k_t-1)} \tag{2}$$

where Z is the lag operator in operational time such that $x_\tau Z = x_{\tau-1}$. Traditionally, $W_t(Z) = 1$ for point-in-time aggregation schemes while $W_t(Z) = (1 + Z + Z^2 + \ ... + Z^{(k_t-1)})/k_t$ for phase averaging aggregation schemes, although in principle one need not restrict attention to these two alternatives.

In this paper, we analyze the following transformation of the data

$$\mathsf{x} = \{\mathsf{x}_t\}_{t=1}^{\infty} = \left\{W_t(Z)x_{\varphi(t)}\right\}_{t=1}^{\infty}; \qquad \varphi(t) = \sum_{j=1}^{t} k_t \tag{3}$$

or more specifically

$$\begin{aligned} \mathsf{x}_1 \quad &= \quad W_1(Z)x_{k_1} = \sum_{j=0}^{(k_1-1)} w_{1,j}x_{(k_1-j)} = \\ &\quad w_{1,0}x_{k_1} + w_{1,1}x_{(k_1-1)} + ... + w_{1,(k_1-1)}x_1 \end{aligned} \tag{4}$$

$$\mathsf{x}_2 = W_2(Z)x_{(k_1+k_2)} = \sum_{j=0}^{(k_2-1)} w_{2,j}x_{(k_1+k_2)-j} =$$

$$w_{2,0}x_{(k_1+k_2)} + w_{2,1}x_{(k_1+k_2)-1} + ... + w_{2,(k_2-1)}x_{(k_1+1)}$$

$$...$$

This framework allows us to treat time aggregation comprehensively. For example, type I aggregation from monthly to quarterly data implies $k_t = 3\ \forall t$ and $W_t(Z) = 1\ \forall t$ (point-in-time) or $W_t(Z) = \left(1 + Z + Z^2\right)/3$ (phase averaging). Alternatively, type II aggregation corresponds to allowing k_t to be a stochastic process (in Jordá 1999, k_t is a Poisson random variable). Other types of aggregation are easily accommodated as well.

In principle, given the finite dimensional cumulative density function of x, one can derive the density of x simply by application of standard techniques for the linear transformation of random variables (see e.g. Mood et Al. (1974)). However, this involves marginalization and integration with respect to several variables, thus making the problem intractable. In this paper we follow an approach that is common in the literature. Assume that the operational time process x follows a general ARIMA process, and that $k = \{k_t\}_{t=1}^{\infty}$ is a generic stochastic process which in some cases is related to the process that generates x.

Therefore, x evolves according to the following stochastic linear difference equation

$$\Phi(Z)x_\tau = \Psi(Z)\varepsilon_\tau \tag{5}$$

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 ε_τ is a white noise error, $\varepsilon_\tau \sim WN(0, \sigma^2)$. Under these assumptions, we define the auxiliary vectors

$$\underset{g_t \times 1}{\gamma_t} = (-\phi_1, -\phi_2, ..., -\phi_p, 0, ..., 0)' \tag{6}$$

$$\underset{b_t \times 1}{\beta_t} = (\beta_{t,1}, \beta_{t,2}, ..., \beta_{t,b_t})'$$

where $g_t = \sum_{j=0}^{p-1} k_{t-j}$; $b_t = \sum_{j=0}^{p-1} k_{t-j} - p$. The $\beta_{t,i}$ for $i = \{1, 2, ..., b_t\}$ are the coefficients of the polynomial $B_t(Z) = \left(1 + \beta_{t,1} Z + \beta_{t,2} Z^2 + ... + \beta_{t,b_t} Z^{b_t}\right)$ and the matrix Γ_t is defined as

$$\underset{g_t \times b_t}{\Gamma_t} = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ -\phi_1 & 1 & 0 & ... & 0 \\ -\phi_2 & -\phi_1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ -\phi_p & -\phi_{p-1} & -\phi_{p-2} & ... & ... \\ 0 & -\phi_p & -\phi_{p-1} & ... & ... \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

In addition, denote Γ_t^* to be a $b_t \times b_t$ matrix and γ_t^* to be a $b_t \times 1$ vector obtained by deleting the k_{t-j} rows of Γ_t and γ_t respectively for $j = 0, 1, 2, ..., p-1$. Then, we introduce the following two propositions:

Point-in-Time Sampling

Proposition 1 *If x is the operational time process generated according to the general discrete time model defined in equation (5), $k = \{k_t\}_{t=1}^{\infty}$ and x is the observational time process obtained from a point-in-time sampling scheme such that $\mathsf{x} = \{\mathsf{x}_t\}_{t=1}^{\infty} = \{x_{k_1},\ x_{(k_1+k_2)},\ x_{(k_1+k_2+k_3)},\ ...\}$ then x follows the linear stochastic difference equation*

$$C_t(L)\mathsf{x}_t = H_t(L)\upsilon_t \qquad \upsilon_t \sim WN(0, \xi_t^2) \tag{8}$$

where L is the lag operator in observational time such that $L\mathsf{x}_t = \mathsf{x}_{t-1}$. 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 $j = 1, ..., p$ while 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

$$\begin{aligned} \sum_{i=0}^{r_t} h_{t,i}^2 \xi_{t-i}^2 &= \sum_{i=0}^{b_t+q} \pi_{t,i}^2 \sigma^2 \\ -h_{t,j}\xi_{t-j}^2 + \sum_{i=1}^{(r_t-j)} h_{(t-j),i}\xi_{t-j-i}^2 h_{t,(j+i)} &= -\pi_{t,l}\sigma^2 + \sum_{i=1}^{b_t+q-l} \pi_{(t-j),i}\sigma^2\pi_{t,(l+i)} \end{aligned} \tag{9}$$

for $j = 1, ..., r_t$, where $\Pi_t(Z) = B_t(Z)\Psi(Z)$; $\beta_t = -(\Gamma_t^*)^{-1}\gamma_t^*$ and $l = \sum_{m=1}^{j} k_{t+1-m}$.

Proof. See Appendix.

Phase Averaging

Let γ_t, β_t and Γ_t be defined as in (6) and (7) but with $g_t = \sum_{j=0}^{p} k_{(t-j)} - 1$; $b_t = \sum_{j=0}^{p} k_{(t-j)} - p - 1$. 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 polynomial $D_t(L)$. Let λ_t^* be the $b_t \times 1$ vector obtained by deleting the k_{t-j} rows of λ_t for $j = 0, 1, ..., p-1$.

Proposition 2 *If x is generated by (5), $k = \{k_t\}_{t=1}^{\infty}$, $W_t(L) = (1 + Z + Z^2 + ... + Z^{(k_t - 1)})/k_t$ and $\mathsf{x} = \{\mathsf{x}_t\}_{t=1}^{\infty} = \{W_t(L) x_{\varphi(t)}\}_{t=1}^{\infty}$ with $\varphi(t) = \sum_{i=1}^{t} k_i \; \forall t$ then*

$$D_t(L)\mathsf{x}_t = M_t(L) u_t \qquad \text{for } u_t \sim WN(0, \nu_t^2) \tag{10}$$

The coefficients of $D_t(L)$ are the solutions of the linear system of p equations which correspond to the k_{t-j+1} rows of $\Gamma_t (\Gamma_t^*)^{-1} (\lambda_t^* - \gamma_t) + \gamma_t = \lambda_t$ for $j = 1, ..., p$. 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

$$\sum_{i=0}^{s_t} m_{t,i}^2 \nu_{t-i}^2 = \sum_{i=0}^{b_t + q - l} \theta_{t,i}^2 \sigma^2 \tag{11}$$

$$-m_{t,j} \nu_{t-j}^2 + \sum_{i=1}^{s_t - j} m_{(t-j),i} \nu_{t-j-i}^2 m_{t,(j+i)} = -\theta_{t,l} \sigma^2 + \sum_{i=1}^{b_t + q - l} \theta_{(t-j),i} \sigma^2 \theta_{t,(l+i)};$$

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

Proof. See Appendix.

These propositions show that the coefficients of the aggregated process x are, in general, time-varying (whenever k is non-constant) and rather different from the coefficients of the original process x. The order of the autoregressive polynomial is typically preserved although the observational time process x can now exhibit a moving average component. For a point-in-time sampling scheme, the MA component will usually be of order $p - 1$, while under phase-averaging the usual order is p (lower/higher values can be obtained when $p - q > k_t / q - p \geq k_t$, a well known result in the context of fixed interval time aggregation). We note that the proofs in propositions 1 and 2 make no assumptions regarding the roots of $\Phi(Z)$ in (5) and therefore, they apply to stationary, integrated or even explosive processes.

In the particular case where $k_t = k \ \forall t$, 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.

3 Maximum Likelihood Estimation

This section derives the maximum likelihood estimators for the observational time parameters from the state space representation of the operational time process assuming that the sequence $\{k_t\}$ is observed. The asymptotic distribution of the maximum likelihood estimators is then derived. In conjunction with propositions 1 and 2, one can then recover the parameters of the operational time model. When $k_t = k$, $\forall t$, the aggregated model will be an ARIMA model with constant parameters, for which traditional modelling and estimation results are readily available. Thus, here we concentrate on situations where k_t changes over time. Let $\breve{\mathsf{x}}_{t-1} = \{\mathsf{x}_{t-1}, \mathsf{x}_{t-2}, ...\}$, $\breve{k}_{t-1} = \{k_{t-1}, k_{t-2}, ...\}$, then, assuming fixed initial conditions, the joint likelihood for a sample of T observations can be written as

$$L(\Gamma) = \prod_{i=p}^{T} f(\mathsf{x}_i | \breve{\mathsf{x}}_{i-1}; \gamma_i). \tag{12}$$

Yet, this expression is not suited for the derivation of the ML estimators, $\{\widehat{\gamma}_i\}$, $i = p, ..., T$, since the number of parameters is typically larger than the available observations. Alternatively, the likelihood can be reparametrized in terms of the parameters of the disaggregate process x_τ, say θ. Therefore, we propose a general Kalman filter based approach for the derivation of $L(\theta)$. We begin by casting the operational time τ ARMA process (5) in state space form. Next we consider the state space form for the aggregated process. Then we 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 state space form for the ARMA(p,q) process in (5) is

$$\begin{aligned} x_\tau &= z'\alpha_\tau \ , \\ \alpha_\tau &= S\alpha_{\tau-1} + e_\tau \ , \\ E(\alpha_0) &= \alpha_0, \quad V(\alpha_0) = P_0, \quad E(e_\tau \alpha_0) = 0 \ \forall \tau \ , \end{aligned} \tag{13}$$

with

$$z' = [1 \ -\psi_1 \ -\psi_2 \ ... \ -\psi_{r-1}] \, ,$$

$$S = \begin{bmatrix} \phi_1 & \phi_2 & ... & \phi_{r-1} & \phi_r \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ ... & & & & \\ 0 & 0 & ... & 1 & 0 \end{bmatrix}, \quad e_\tau = \begin{bmatrix} \varepsilon_\tau \\ 0 \\ 0 \\ ... \\ 0 \end{bmatrix},$$

where $r = max(p, q+1)$, α_τ is an r dimensional vector of state variables, $\tau = 1, ..., T$. With respect to (5) we further assume that ε_τ is Normally distributed.

Let us now define the variables $s_i = \sum_{j=1}^{i} k_j$, $i = 1, ..., N$, with $s_0 = 0$, $s_N = T$, and

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

where w_{i,r_i-j} are the weights in $W_t(Z)$ in (2), 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

$$\begin{aligned} x_\tau &= g'\gamma_\tau, \quad \tau = s_i \ , \ i = 1, ..., N, \\ \gamma_\tau &= D_\tau \gamma_{\tau-1} + R\eta_\tau, \quad \tau = 1, ..., T, \end{aligned} \tag{14}$$

with

$$
\begin{aligned}
g' &= \begin{bmatrix} \underset{1\times r}{0} & 1 \end{bmatrix}, \quad \gamma_\tau = [\alpha_\tau \ \beta_\tau]', \\
D_\tau &= \begin{bmatrix} S & 0 \\ z'S & \varphi_\tau \end{bmatrix}, \quad R = \begin{bmatrix} I & 0 \\ z' & 0 \end{bmatrix}, \quad \eta_\tau = \begin{bmatrix} e_\tau \\ 0 \end{bmatrix}.
\end{aligned}
$$

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

$$
\begin{aligned}
\mathsf{x}_t &= g'\gamma_t, \\
\gamma_t &= \begin{bmatrix} S^{k_t} & 0 \\ z'(W_{k_t} - I) & 0 \end{bmatrix} \gamma_{t-1} + \begin{bmatrix} I & 0 \\ 0 & z' \end{bmatrix} \begin{bmatrix} \eta_t^\alpha \\ \eta_t^\beta \end{bmatrix},
\end{aligned}
\tag{15}
$$

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

To derive the ML estimators of the disaggregate parameters, it is more convenient to adopt the SSR(τ). Defining the optimal estimators of γ_τ by c_τ, with covariance matrix Σ_τ, the Kalman filter equations are:

$$
\begin{aligned}
c_{\tau|\tau-1} &= D_\tau c_{\tau-1} \\
\Sigma_{\tau|\tau-1} &= D_\tau \Sigma_{\tau-1} D_\tau' + RQR' \\
c_\tau &= \begin{cases} c_{\tau|\tau-1} & \tau \neq s_i, \quad i = 1, ..., N \\ c_{\tau|\tau-1} + \Sigma_{\tau|\tau-1} g' f_\tau^{-1} g(x_\tau - g' c_{\tau|\tau-1}) & \text{otherwise} \end{cases}, \\
\Sigma_\tau &= \begin{cases} \Sigma_{\tau|\tau-1} & \tau \neq s_i, \quad i = 1, ..., N \\ \Sigma_{\tau|\tau-1} + \Sigma_{\tau|\tau-1} g' f_\tau^{-1} g \Sigma_{\tau|\tau-1} & \text{otherwise} \end{cases}, \\
c_0 &= \begin{bmatrix} \alpha_0 \\ 0 \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix},
\end{aligned}
\tag{16}
$$

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

$$
v_\tau = x_\tau - \widehat{x}_{\tau|\tau-1} = g'(\gamma_\tau - c_{\tau|\tau-1}), \quad \tau = s_i, \quad i = 1, ..., N. \tag{17}
$$

Hence, the likelihood can be written as

$$\log L(\theta) = \log \prod_{i=1}^{N} f(\mathsf{x}_i | \breve{\mathsf{x}}_{i-1}; \theta) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{N}\log f_{s_i} - \frac{1}{2}\sum_{i=1}^{N}\frac{v_{s_i}^2}{f_{s_i}}. \tag{18}$$

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 operational 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.

In order to derive the properties of the ML estimators, we make the following additional assumptions:

(a) The eigenvalues of S are inside the unite circle.

(b) The true parameter values, θ^0, are in an interior point of the parameter space.

(c) θ^0 is globally identifiable in the sense of Rothenberg (1971).

Now, it can be easily checked that all the conditions in Theorem 4.2 and 6.5 in White (1994, pp. 42 and 94) are satisfied and thus,

$$\widehat{\theta} \xrightarrow{p} \theta^0,$$

$$\sqrt{N}(\widehat{\theta} - \theta^0) \xrightarrow{d} N(0, IA^{-1}),$$

where IA is the asymptotic Information matrix. Assumption (a) can be relaxed to allow for non-stationary processes, but this substantially complicates inference, see e.g. Sims *et al.* (1990). Assumption (b) is a standard assumption to guarantee asymptotic normality of the normalized estimators. Assumption (c) is needed to ensure convergence to θ^0. Notice that standard conditions for identification of ARMA models, e.g. Hannan (1971), are necessary but not sufficient for assumption (c) to hold, see e.g. Marcellino (1998). Temporal aggregation can transform globally identifiable parameters into locally identifiable ones (e.g. when an AR(1) process is subject to point-in-time sampling with $k_t = k$ and k is even), or into non-identifiable ones (e.g. when an MA(q) process is subject to point-in-time sampling with $k_t > q$). The pre-existing hypothesis of an ARMA process with i.i.d. normal errors makes the other assumptions in White's (1994) theorems valid.

4 Practical Modelling Strategies

A transformation of the time scale from τ-time to t-time will yield an observational time process x_t and a sequence $\{k_t\}$ which corresponds to the frequency of aggregation. We begin this section by assuming 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. Consequently, we propose a number of models and estimation strategies. In practice there are also cases where k_t is not observable. The second subsection presents solutions when this situation occurs.

4.1 Stochastic and observable k_t : The ACI Model

Consider modelling the joint probability distribution of x_t and k_t conditional on past information, given the distribution of x_τ, the aggregation scheme $W_t(Z)$, and possibly a vector of exogenous variables, $\underset{\sim}{z}_{t-1}$. Assume for simplicity that k_t can only take a finite number of integer values, $k_t \in \{0, 1, 2, ..., N\}$ with $P(k_t = j| \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta) = P(k_t = j) = \pi_j$. These assumptions will be relaxed later but make the exposition clearer here. Then the joint distribution of x_t and k_t can be factored as follows

$$f(\mathsf{x}_t, k_t = j| \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \delta) =$$

$$g(\mathsf{x}_t| k_t = j, \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta) \cdot P(k_t = j) \tag{19}$$

Under the assumption of Gaussianity, the conditional distribution of x_t from this expression becomes

$$g(\mathsf{x}_t| k_t = j, \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}\sigma(j)_t} exp\left\{\frac{-(\mathsf{x}_t - \mu(j)_t)^2}{2\sigma(j)_t^2}\right\} \tag{20}$$

were $\mu(j)$ and $\sigma(j)$ are indexed by j to indicate their dependence on the frequency of aggregation. From (19), the joint likelihood is

$$f(\mathsf{x}_t, k_t = j| \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \delta) = \frac{\pi_j}{\sqrt{2\pi}\sigma(j)_t} exp\left\{\frac{-(\mathsf{x}_t - \mu(j)_t)^2}{2\sigma(j)_t^2}\right\} \tag{21}$$

If k_t is weakly exogenous for θ (see Engle et al. (1983)), that is, if the parameters of interest are a function of θ, and θ and the π's are variation free, then joint estimation is not required, and θ can be estimated by maximizing (20) alone. >From a computational point of view, the Kalman filter based approach proposed in section 3 can be used to estimate this likelihood. Under the additional assumption of Granger non-causality of x_t for k_t (which holds in this case), inference on θ can also be conducted without any reference to the distribution of k_t (see e.g. Govearts *et al.* (1995)). Otherwise, the joint likelihood for x_t and k_t has to be used, and it can be easily obtained from (19).

When k_t is observable but with a one period delay, the unconditional density of x_t has to be used. Assuming the operational time process is Gaussian, this is a mixture of normal distributions, which is obtained by summing up (21) over all the possible values of k_t,

$$P(\mathsf{x}_t|\ \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta) = \sum_{j=0}^{N} f(\mathsf{x}_t, k_t = j|\ \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta). \tag{22}$$

Consequently, the joint unconditional log-likelihood can be written as

$$L(\delta) = \sum_{t=max(p,q)}^{T} log\left(P(\mathsf{x}_t|\ \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta)\right) \tag{23}$$

One natural generalization of $P(k_t = j) = \pi_j$ is to assume that k_t is Poisson distributed. Jordá (1999), based on the Law of Rare Events, presents formal arguments that justify this choice. In recent work, Engle and Russell (1998) and Hamilton and Jordá (1999) have introduced time series models for the analysis of duration data. Based on the correspondence between stochastic point processes and duration data analysis, a natural parametrization for the distribution of k_t under the Poisson assumption is

$$\begin{aligned} P(k_t &= j|\ \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta) = \frac{e^{-\lambda_t}\lambda_t^j}{j!} \\ log(\lambda_t) &= \omega + \alpha log(\lambda_{t-1}) + \beta\left(\frac{k_{t-1}}{\lambda_{t-1}}\right) + \delta'\underline{\mathsf{x}}_{t-1} \end{aligned} \tag{24}$$

where $\underline{\mathsf{x}}_{t-1} = (\mathsf{x}_{t-1}, \mathsf{x}_{t-2}, ..., \mathsf{x}_{t-r})'$. λ_t is the conditional intensity of the Poisson process, in our application, the average number of operational time periods over which we aggregate in the interval of observational time $[t, t-1)$. Consequently, it is natural to term the model in (24) as the *autoregressive conditional intensity model,* ACI(1,1). The specification of the conditional intensity function ensures that the intensity remains strictly positive without restricting the parameter space. The term $log(\lambda_{t-1})$ makes explicit the autoregressive nature of the specification, while the term $log\left(\frac{k_{t-1}}{\lambda_{t-1}}\right)$ captures the dynamic effect of surprises in k_t. This ratio will be approximately an i.i.d. sequence, with the approximation becoming exact as $\delta' \underline{\mathsf{x}}_{t-1}$ becomes negligible. The role played by the term $log\left(\frac{k_{t-1}}{\lambda_{t-1}}\right)$ is similar in nature to the moving average term in a typical ARMA model. The model will be stationary for $|\alpha + \beta| < 1$, when $\delta' = 0$.

The unconditional density of x_t is just a mixture of normals given by the product of expressions (22) and (24), so that the joint likelihood can be maximized given an initial guess for λ_0, which for convenience can be set equal to the unconditional intensity. Section 7 will apply the ACI model to explain the price quotes in the FX market and their arrival rate. The ACI model is obviously not limited to the range of problems that we pursue here. Rather, it is a very general, time-series formulation for dynamic, count-data problems.

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

A situation often encountered in applied work is that in which k_t is not observable. This prevents us from directly using the ACI model introduced above. An alternative assumption that we recommend for this situation is to let $P(k_t = j | \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta)$ be an N-state Markov chain, that is

$$P(k_t = j | \breve{\mathsf{x}}_{t-1}, \breve{k}_{t-1}; \theta) = P(k_t = j | k_{t-1} = i) = p_{ij} \tag{25}$$

for $i, j = 1, 2, ..., N$. This assumption can be generalized as in Lam (1990), Durland and McCurdy (1994), Filardo (1993) and Diebold, Lee and Weinbach (1994). We restrict our attention to the basic formulation for clarity of exposition. Based on (21) and (25), it is immediately apparent that this specification boils down to that proposed in Hamilton (1989) for the popular Markov switching regimes (MSR) model, and Tjøstheim's (1986) doubly stochastic model, even if here it arises because of aggregation issues. An example illustrates the particulars of this technique.

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

$$x_\tau = \rho_1 x_{\tau-1} + \rho_2 x_{\tau-2} + \varepsilon_\tau \qquad \varepsilon_\tau \sim N(0, \sigma_\varepsilon^2) \tag{26}$$

For simplicity, assume $k_t = \{1, 2\}$, that is, every operational 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 observational-time process is described as follows

1. $i = 1; j = 1$

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

2. $i = 1; j = 2$

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

3. $i = 2; j = 1$

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

4. $i = 2; j = 2$

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

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

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

Recall, since $p_{ij} = P(k_t = j | k_{t-1} = i)$, then s_t follows a four state Markov chain with transition matrix

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

The appendix contains the expression for the conditional densities and the form that the estimation algorithm proposed by Hamilton (1994) takes in this case.

5 Time-scale Transformations in Practice: The FX Market

This section explains several stylized facts peculiar to the intra-daily foreign exchange market (FX) as possibly stemming from a time scale transformation. The next section investigates the pitfalls in structural inference of the dealer's inventory control models in the presence of time aggregation while section 7 estimates the dynamic behavior of the bid-ask spread in the U.S. Dollar - Deutsche Mark, FX market. These data provide unique opportunities for understanding the behavior of financial markets and concepts like risk and market efficiency as well as revealing the micro-structure and decision making of traders in this market.[3] However, the econometric analysis of high frequency data (for example, there are an average of 3,100 quotes per day for the USD-DM FX rate in a typical business day[4]) has required the development of a variety of new

[3] Surveys on the FX market at daily or weekly frequencies include Hsieh (1988), Baillie and McMahon (1989), and Guillaume et al. (1995).

[4] See Dacorogna et al. (1993).

econometric techniques that range from long-memory models in volatility[5] to developing time scales that depend on the geographical peculiarities of the three main trading regions[6] (Europe, East Asia and America) to developing new time series models for irregularly spaced data[7] . The next three subsections will: (1) enumerate the principal statistical stylized facts, documented elsewhere in the literature; (2) discuss the role of time aggregation in theory; and (3) present a Monte-Carlo study as an illustration.

5.1 Stylized facts

Denote the *price* at time τ or tick τ as

$$x_\tau \equiv log(fx_\tau) \tag{27}$$

that is, the log of an exchange rate quote. For simplicity, we do not distinguish between "asks" and "bids" in which case, x_τ is typically taken to be the average of the log ask and log bid quotes. Consequently, the *change of price* or *return* is defined as

$$r_\tau \equiv [x_\tau - x_{\tau-1}]. \tag{28}$$

The volatility associated with this process will be defined as

$$v_\tau \equiv \frac{1}{n}\sum_{k=1}^{n} |r_{\tau-k}| \tag{29}$$

where n will depend on the size of the operational time interval. The absolute value of the returns is preferred to the more traditional squared value because it captures better the autocorrelation and seasonality of the data (Taylor, 1988; Müller et al., 1990; Granger and Ding, 1996). There are a variety of other quantities of interest such as the relative spread, the tick frequency, the volatility ratio, etc. that fall beyond the scope of this discussion.

[5] See Baillie (1996) for a survey.

[6] See Dacorogna et al. (1993).

[7] See Engle and Russell (1998).

The following is a list of the most salient stylized facts:

1. The data is non-normally distributed with "fat tails." However, temporal aggregation tends to diminish these effects. At a weekly frequency, the data appears normal.

2. The data is leptokurtic although temporal aggregation reduces the excess kurtosis.

3. Seasonal patterns corresponding to the hour of the day, the day of the week and the presence of traders in the three major geographical trading zones can be observed for returns and particularly for volatility. However, Dacorogna et al. (1993, 1996), via a change in the time-scale based on a business time-scale, transform the data such that seasonality and conditional heteroskedasticity are eliminated. Their time-scale basically consists on "expanding" periods of high volatility and "contracting" those of low volatility. The reader is referred to the references for a detailed discussion.

4. Let the *scaling law* reported in Müller et al. (1990) be defined as:

$$\overline{|x_\tau - x_{\tau-1}|} = \left(\frac{\Delta\tau}{k}\right)^D \tag{30}$$

   where k is a constant that depends on the FX rate and $D = 1/E$ is the drift exponent. For a Gaussian random walk, the theoretical value of $D = 0.5$. However, it is observed that $D \simeq 0.58$ for the major FX rates. The scaling law holds with a similar value of D for the volatility.

5. The volatility is decreasingly conditionally heteroskedastic with the frequency of aggregation.

6. Seasonally filtered absolute returns data exhibits long-memory effects, that is, autocorrelations that decay at a slower than exponential rate.

5.2 The Role of Random Time Aggregation

Under common forms of market efficiency, it is natural to assume that the price process follows a martingale. We will assume that the driving process for prices is a random walk – a more stringent assumption than a martingale in that it does not allow dependence in higher moments. Accordingly, let

$$x_\tau = f_\tau + \rho x_{\tau-1} + \varepsilon_\tau \qquad \varepsilon_\tau \sim WN(0, \sigma_\varepsilon^2), \tag{31}$$

where the random walk condition would imply $\rho = 1$ and $f_\tau = 0$, a deterministic, possibly time varying drift which is useful to accommodate for the fact that prices might grow over time.

Consider a simple scenario in which the frequency of aggregation is deterministic and cyclical, i.e., $k = k_1, k_2, ..., k_j, k_1, k_2, ..., k_j, ...$ This is a convenient way to capture the seasonal levels of activity in different hours, or days of the week and serves to illustrate some important results. The (point-in-time) aggregated process resulting from (31) and the frequency of aggregation described above is therefore a time-varying seasonal AR(1):

$$\begin{aligned}
\mathsf{x}_t &= \rho^{k_1}\mathsf{x}_{t-1} + u_t \qquad u_t \sim (0, \sigma_{u,t}^2), \\
\mathsf{x}_{t+1} &= \rho^{k_2}\mathsf{x}_t + u_{t+1} \qquad u_{t+1} \sim (0, \sigma_{u,t+1}^2), \\
&\ldots \\
\mathsf{x}_{t+j-1} &= \rho^{k_j}\mathsf{x}_{t+j-2} + u_{t+j-1} \qquad u_{t+j-1} \sim (0, \sigma_{u,t+j-1}^2), \\
\mathsf{x}_{t+j} &= \rho^{k_1}\mathsf{x}_{t+j-1} + u_{t+j} \qquad u_{t+j} \sim (0, \sigma_{u,t}^2), \\
&\ldots,
\end{aligned} \tag{32}$$

where the errors are uncorrelated, $\sigma_{u,t+(i-1)}^2 = (1 + \rho^2 + ... + \rho^{2(k_i-1)})\sigma_\varepsilon^2$, $i = 1, ..., j$, and t is measured in hours or days. Further aggregation by point-in-time sampling with $\widetilde{k} = \sum_{i=1}^j k_i, \sum_{i=1}^j k_i, ...,$ yields the constant parameter AR(1) process

$$\mathsf{x}_T = \rho^{\widetilde{k}}\mathsf{x}_{T-1} + e_T \qquad e_T \sim WN(0, \sigma_e^2), \tag{33}$$

with $\sigma_e^2 = \sum_{i=0}^{j-1} \rho^{2\sum_{l=0}^{i} k_l}\sigma_{u,t-i}^2$, $k_0 = 0$. Time (T) is measured in days or weeks.

Most of the properties described in the previous subsection are referred to the first differences of the variables, so that we also derive their generating mechanism. From (32), after some rearrangements, we get:

$$\begin{aligned}\Delta \mathsf{x}_{t+1} &= \frac{\rho^{k_2}-1}{\rho^{k_1}-1}\rho^{k_1}\Delta \mathsf{x}_t + u_{t+1} - \left(\frac{\rho^{k_2}-1}{\rho^{k_1}-1}\rho^{k_1} - \rho^{k_2} + 1\right) u_t, \\ \Delta \mathsf{x}_{t+2} &= \frac{\rho^{k_3}-1}{\rho^{k_2}-1}\rho^{k_2}\Delta \mathsf{x}_{t+1} + u_{t+2} - \left(\frac{\rho^{k_3}-1}{\rho^{k_2}-1}\rho^{k_2} - \rho^{k_3} + 1\right) u_{t+1}, \\ & \ldots,\end{aligned} \tag{34}$$

that is, a time-varying seasonal ARMA(1,1) process, except for $\rho = 1$ (the model collapses to a random walk with time-varying variance). Instead, from (33), it simply becomes:

$$\Delta \mathsf{x}_T = \rho^{\widetilde{k}}\Delta \mathsf{x}_{T-1} + \Delta e_T. \tag{35}$$

Consider the stylized facts from the previous subsection in light of this simple example.

1. Non normality of $\Delta \mathsf{x}_t$ and normality of $\Delta \mathsf{x}_T$ is coherent with the fact that u_t is a weighted sum of a smaller number of original errors (ε_τ) than e_T. The time-varying nature of (34) can also contribute to the generation of outliers, that in turn can determine the leptokurtosis in the distribution of $\Delta \mathsf{x}_t$.

2. (34) can also explain why the value of D in (30) is not 0.5: x_t is not a pure Gaussian random walk. It is more difficult to determine theoretically whether (34) can generate a value of D close to the empirical value 0.59. We will provide more evidence on this in the simulation experiment of the next subsection.

3. The long memory of $\Delta \mathsf{x}_t$ can be a spurious finding due to the assumption of a constant generating mechanism, even if particular patterns of aggregation can generate considerable persistence in the series.

4. The presence of seasonality in the behavior of $\Delta \mathsf{x}_t$ is evident from (34). (35) illustrates that this feature can disappear when further aggregating the data.

5. Conditional heteroskedasticity can also easily emerge when a constant parameter model is used instead of (34). That it disappears with temporal aggregation is a well known result,

see e.g. Diebold (1988), but (35) provides a further reason for this to be the case, i.e., the aggregated model is no longer time-varying.

6. The time-scale transformations by Dacorogna *et al.* (1993, 1996) can be interpreted in our framework as a clever attempt to homogenize the aggregation frequencies, i.e., from $k = k_1, k_2, ..., k_j, k_1, k_2, ..., k_j, ...$ to $k = \widetilde{k}, \widetilde{k}, ...$, by redistributing observations from more active to less active periods. This changes the t time scale, which can be still measured in standard units of time, and makes the parameters of the $\Delta \mathsf{x}_t$ process stable over time. This transformation attenuates several of the mentioned peculiar characteristics of intra daily or intra weekly exchange rates.

5.3 A Monte Carlo Study

This subsection analyzes the claims presented above and illustrates some of the theoretical results just derived via Monte-Carlo simulations. The D.G.P. we consider for the *price* series is the following operational time AR(1) model:

$$x_\tau = \mu_\tau + \rho x_{\tau-1} + \varepsilon_\tau$$

where $\varepsilon_\tau \sim N(0,1)$. Under a strong version of market efficiency, it is natural to experiment with $\mu = 0$ and $\rho = 1$. However, we also consider $\mu = 0.000005$ and $\rho = 0.99$ to study the consequences of slight deviations from the random walk ideal. We simulated series of 50,000 observations in length. The first 100 observations of each series are disregarded to avoid initialization problems.

The operational time D.G.P. is aggregated three different ways:

1. **Deterministic, seasonal, irregularly spaced aggregation:** Following the discussion of the previous subsection, let the auxiliary variable $s_\tau = 1$ if observation τ is recorded, 0 otherwise. Then, the following deterministic sequence determines the point-in-time aggregation scheme:

$$\begin{cases} s_r = 1 \quad if \quad r \in \{1,2,3;26,27,28;36,37;41,42;56,57,58;76,77\} \\ s_r = 0 \qquad\qquad otherwise \end{cases}$$

and $s_{r+100n} = s_r$ for $r \in \{1, 2, ..., 100\}$ and $n \in \{1, 2, ...\}$. In other words, the aggregation scheme repeats itself in cycles of 100 observations. Within the cycle there are periods of high frequency of aggregation and low frequency of aggregation that mimic the intensity in trading typical of the FX market. Note that from the sequence $\{s_\tau\}_{\tau=1}^{50,000}$ it is straight forward to obtain the sequence $\{k_t\}_{t=1}^{T}$. For example, the first few terms are: 1, 1, 23, 1, 1, 8, ...

2. **Deterministic, fixed interval aggregation:** This consists on a simple sampling scheme with $k_t = 100\ \forall t$ or in terms of the auxiliary variable, $s_\tau = 1$ if $r \in \{100, 200, ...\}, 0$ otherwise.

3. **Random, seasonal, irregularly spaced aggregation:** Let $h_\tau \equiv P(s'_\tau = 1)$ which can be interpreted as a discrete time hazard.[8] Accordingly, the expected duration between recorded observations is $\psi_\tau = h_\tau^{-1}$. Think of the underlying "innovations" for the process that generates s'_τ as being an i.i.d. sequence of continuous-valued logistic variables denoted $\{v_\tau\}$. Further, suppose there exists a latent process $\{\lambda_\tau\}$ such that:

$$P(s'_\tau = 1) = P(v_\tau > \lambda_\tau) = (1 + e^{\lambda_\tau})^{-1}$$

Notice, $\lambda_\tau = log(\psi_\tau - 1)$. Hamilton and Jordá (1999) show that one can view this mechanism as a discrete-time approximation that generates a frequency of aggregation that is Poisson distributed. For the sake of comparability, we choose λ_τ to reproduce the same seasonal pattern as in bullet point 1 above but in random time. Accordingly:

$$\lambda_\tau = \lambda - 1.5\lambda s_\tau$$

where $\lambda = log(15 - 1)$, since 15 is the average duration between non-consecutive records described by the deterministic, irregular aggregation scheme introduced above. In other

[8] We use the notation s'_τ to distinguish it from its deterministic counterpart introduced in bullet point 1 above.

words, the probability of an observation being recorded is usually 0.07 except when $s_\tau = 1$ in which case this probability jumps to 0.8.

Tables 1-4 report the following information for the original and aggregated data: (1) the coefficient of kurtosis of the simulated *price* series; (2) the p-value of the null hypothesis of normality from the Jarque-Bera statistic; (3) the estimated coefficient D of the scaling law; (4) the presence of ARCH in absolute returns ($|r_t|$ in (28)) ; and (5) the presence of ARCH in volatility for averages over 5 periods (v_t in (29)).

Table 1 - Operational Time Data

	Kurtosis	**Jarque-Bera**	**D**	**ARCH** $\vert r_t\vert$	**ARCH** v_t
$\rho = 1; \mu = 0$	3.0041	0.4283	0.5002	No	No
$\rho = 1; \mu = 5 \times 10^{-6}$	3.0019	0.5797	0.5044	No	No
$\rho = 0.99; \mu = 0$	3.0108	0.4213	0.4842	No	No
$\rho = 0.99; \mu = 5 \times 10^{-6}$	2.9985	0.4214	0.4832	No	No

Table 2 - Deterministic, Seasonal, Irregularly Spaced Aggregated Data

	Kurtosis	**Jarque-Bera**	**D**	**ARCH** $\vert r_t\vert$	**ARCH** v_t
$\rho = 1; \mu = 0$	7.7214	0.0000	0.5506	Yes	Yes
$\rho = 1; \mu = 5 \times 10^{-6}$	7.7385	0.0000	0.5716	Yes	Yes
$\rho = 0.99; \mu = 0$	7.3839	0.0000	0.4464	Yes	Yes
$\rho = 0.99; \mu = 5 \times 10^{-6}$	7.4858	0.0000	0.4483	Yes	Yes

Table 3 - Deterministic, Fixed Interval, Aggregated Data

	Kurtosis	**Jarque-Bera**	**D**	**ARCH** $\|r_t\|$	**ARCH** v_t
$\rho = 1; \mu = 0$	2.9406	0.5149	0.5105	No	No
$\rho = 1; \mu = 5 \times 10^{-6}$	2.9371	0.5457	0.7246	No	No
$\rho = 0.99; \mu = 0$	2.8761	0.5000	0.0467	Yes*	No
$\rho = 0.99; \mu = 5 \times 10^{-6}$	2.9657	0.5414	0.0502	Yes*	No

* Barely significant at the conventional 5% level.

Table 4 - Random, Seasonal, Irregularly Spaced, Aggregated Data

	Kurtosis	**Jarque-Bera**	**D**	**ARCH** $\|r_t\|$	**ARCH** v_t
$\rho = 1; \mu = 0$	6.4773	0.0000	0.5351	Yes	No
$\rho = 1; \mu = 5 \times 10^{-6}$	6.4035	0.0000	0.5531	Yes	No
$\rho = 0.99; \mu = 0$	6.1091	0.0000	0.4454	Yes	No
$\rho = 0.99; \mu = 5 \times 10^{-6}$	6.2055	0.0000	0.4462	Yes	Yes*

* Barely significant at the conventional 5% level.

Several patterns are worth noting from the tables. Both forms of irregularly spaced data generate fat tailed distributions away from the normal with excess kurtosis and ARCH in absolute returns. The coefficient for D is very close to the analytical level of 0.5 for the original and the regularly spaced data but it takes on values of approximately 0.55 for irregularly spaced data for both cases of $\rho = 1$. This is close to the 0.58 reported for most FX series. In addition, the seasonal patterns induced through the deterministic, irregularly spaced aggregation, are readily apparent in the shape of the autocorrelation function of absolute returns but not for the returns series per se, in a manner that is also characteristic of FX markets. This simple experiment thus demonstrates that time aggregation may be behind many of the stylized facts common to high frequency FX data.

6 Time Aggregation and Structural Inference

The previous section illustrated that many of the stylized properties of the returns in FX markets can be explained as the outcome of time scale transformations, that is, as an artifact of the irregular arrival of price quotes in time rather than as an intrinsic property of returns. Another important element in most financial markets concerns the microstructure of the dealer's preferred inventory levels. This section presents a simplified version of such inventory control problems to illustrate the potential pitfalls of structural inference when the practitioner assumes fixed interval aggregation (which includes the no-aggregation case $k = 1$ as a particular class) when the aggregation is over stochastic intervals of time.

Consider the following stock adjustment model. Let z_τ describe the disequilibrium existing at time τ, that is, the distance between the desired and the actual levels of inventories. Under rather general assumptions, Jordá (1999) shows that the dynamics of z_τ are described by the first order autoregressive process

$$z_\tau = (1 - \alpha) z_{\tau-1} + \varepsilon_\tau \qquad \varepsilon_\tau \sim N(0, \sigma^2); \qquad \alpha \in (0, 1] \tag{36}$$

The assumption of normality is not fundamental but simplifies the derivation of the maximum likelihood estimator (MLE) below.

The interpretation of the model is fairly intuitive. Let the operational time-scale τ denote the timing with which the dealer decides to adjust the disequilibrium variable z. The disequilibrium existing at the decision node τ is the sum of two components: $(1 - \alpha) z_{\tau-1}$ indicates the amount of disequilibrium left unadjusted at the decision node $\tau - 1$, and ε_τ denotes a stochastic disequilibrium component that is accumulated since the previous decision node, $\tau - 1$. When $\alpha = 1$ the model becomes an (s, S) type model, while values of α between 0 and 1 are indicative of linear-quadratic inventory adjustment costs. Accordingly, α plays a critical role in describing the "speed of adjustment" and constitutes the structural parameter of interest.

Next, assume that the inventory position z is not recorded at each decision node but rather at regular intervals of time. In this context, k_t denotes the number of decision nodes per sampling interval $[t, t-1)$. Under general conditions, Jordá (1999) shows that k_t is best described as a stochastic Poisson point process.

In the language of this paper, the process z_τ is said to be time-scale transformed with frequency of aggregation k_t under a point-in-time sampling scheme. According to proposition 1, the observational-time process, z_t, is described by,

$$\begin{aligned} z_t &= \mu(j)_t + \sigma(j)_t u_t \qquad u_t \sim N(0,1); \qquad j = 1, 2, ..., N \\ \mu(j)_t &= (1-\alpha)^j \\ \sigma(j)_t^2 &= (1 + (1-\alpha)^2 + ... + (1-\alpha)^{2j})\sigma^2 = \sigma^2 \frac{1-(1-\alpha)^{2j+1}}{\alpha} \equiv \sigma_j^2 \end{aligned} \tag{37}$$

Following equation (21), the joint likelihood for this problem becomes,

$$f(z_t, k_t = j | \breve{z}_{t-1}, \breve{k}_{t-1}) = \frac{\pi_j}{\sqrt{2\pi}\sigma_j} exp\left\{ \frac{-(z_t - (1-\alpha)^j z_{t-1})^2}{2\sigma_j^2} \right\} \tag{38}$$

Summing up over all possible values of j to obtain the unconditional density of z_t, the joint likelihood can be written as in (23). Suppose the π_j were constant (for expositional convenience) and let I_t denote the information set at time t, that is, $\{\breve{z}_t, \breve{k}_t\}$. Then, it is easy to check that,

$$E(z_t | I_{t-1}) = \left(\pi_1(1-\alpha) + \pi_2(1-\alpha)^2 + ... + \pi_N(1-\alpha)^N\right) z_{t-1} \tag{39}$$

Now, consider that one correctly understands that the observable process z_t is the outcome of fixed-interval sampling of a process that evolves in economic time. However, rather than assuming that k_t is stochastic, traditional time-aggregation imposes that economic time also evolves in fixed intervals of time, or in our notation, $k_t = \overline{k}\ \forall t$. Thus, z_t is incorrectly perceived to be,

$$\begin{aligned} z_t &= \phi z_{t-1} + v_t \qquad v_t \sim N(0, \xi^2) \\ \xi^2 &= \frac{1-(1-\alpha)^{2\overline{k}+1}}{\alpha}\sigma^2 \\ \phi &= (1-\alpha)^{\overline{k}} \end{aligned} \tag{40}$$

with unconditional density,

$$f(z_t|I_{t-1},\theta) = \frac{1}{\sqrt{2\pi}\xi} exp\left\{\frac{-(z_t - \phi z_{t-1})^2}{2\xi^2}\right\}. \tag{41}$$

Estimation of the structural parameters would then proceed under traditional MLE techniques or more simply by least squares on (40). Both methods yield estimates of $\widehat{\phi}$ and $\widehat{\xi}$ from which, unconditionally with respect to k_t, we obtain,

$$E(z_t|I_{t-1}) = \widehat{\phi} z_{t-1} = (\pi_1(1-\alpha) + \pi_2(1-\alpha)^2 + ... + \pi_N(1-\alpha)^N)z_{t-1}$$

Naturally, the π_i and the α cannot be separately identified, even if one were to know that $E(k_t) = \overline{k}$. Assuming fixed interval time aggregation, it is common to thus set $\widehat{\phi} = (1-\widehat{\alpha})^{\overline{k}}$ and then calculate $\widehat{\alpha}$ from this expression. However, this estimate of the speed of adjustment is obviously downward biased. Consider a numerical example. Let $N = 4$, with $\pi_1 = \pi_4 = 0.5$, $\pi_2 = \pi_3 = 0$ and thus $\overline{k} = 2.5$. Suppose $\alpha = 0.75$, then $\sum_{i=1}^{N} \pi_i(1-\alpha)^i = 0.13$, or in other words, $\widehat{\phi} \simeq 0.13$. Let $\widehat{\alpha}$ be incorrectly computed from $1 - \widehat{\phi}^{1/\overline{k}} = 0.56$. Compare this to $\alpha = 0.75$, the original assumption. The estimated "speed of adjustment" parameter is almost 1/4 lower.

In general, a time-scale transformed process x_t where k_t is stochastic will yield a model like in equations (37) and (38), that is,

$$\mathsf{x}_t = \mu(j)_t + \sigma(j)_t \varepsilon_t \qquad \varepsilon_t \sim WN(0,1)$$

Estimation of a constant parameter model will yield estimates of $E(\mathsf{x}_t|I_{t-1})$ unconditionally with respect to k_t where it is important to note that

$$\sum_{j=1}^{N} \pi_j \mu(j)_t \neq \mu\left(\sum_{j=1}^{N} j\pi_j\right)_t.$$

7 The Bid-Ask Spread

This section illustrates with an empirical example some of the techniques developed in previous sections. In particular, we use high frequency data on market makers' FX, USD - DM quotes. The data corresponds to the HFDF-93 data set available from Olsen & Associates, originally released to researchers for the 1993 High Frequency Data in Finance Conference in 1993. Although most of the research done on this data has focused on forecasting, here we explore the dynamics of the bid-ask spread as a function of information flows. The explanations for the width of the spread vary widely (see O'Hara, 1995), ranging from "market failure" and "market power" explanations (less likely to apply in the FX market) to more transactions, cost-related "dealer risk-aversion" and "gravitational pull" theories. The simple approach we take here is to investigate the dynamics of the bid-ask spread as a function of information flows measured by the level of activity in the market and cast the problem as a time aggregation exercise.

The FX market is a 24 hours global market although the activity pattern throughout the day is dominated by three major trading centers: East Asia, with Tokyo as the major trading center; Europe, with London as the major trading center; and America, with New York as the major trading center. Figure 1 displays the activity level in a regular business day as the number of quotes per half hour interval. The seasonal pattern presented is calculated non-parametrically with a set of 48 time-of-day dummies. Figure 2 illustrates the weekly seasonal pattern in activity by depicting a sample week of raw data.

The original data set spans one year beginning October 1, 1992 and ending September 30, 1993, approximately 1.5 million observations. The data has a two second granularity and it is pre-filtered for possible coding errors and outliers at the source (approximately 0.36% of the data is therefore lost). The subsample that we consider contains 3500 observations of half hour intervals (approximately 300,000 ticks) constructed by counting the number of quotes in half hour intervals throughout the day. For each individual half-hour observation we then record the corresponding bid-ask spread. A comprehensive survey of the stylized statistical properties of the data can be found in Guillaume et al. (1995). Here, we report only some of the salient features of the data.

The average intensity is approximately 120 quotes/half hour during regular business days although during busy periods this intensity can reach 250 quotes/half hour. The activity level significantly drops over the weekend although not completely. The bid-ask spread displays a similar seasonal pattern, with weekends exhibiting larger spreads (0.00110) relative to regular business days (0.00083).

Let k_t denote the number of quotes per half hour interval (a count variable which we have termed at times as the aggregation frequency) and let x_t denote the bid-ask spread that corresponds to the half hour interval t. Thus, the problem consists on estimating the joint density of k_t and x_t which can be decomposed without loss of generality as in equation (19) in section 4.1,

$$P(k_t = j | \breve{x}_{t-1}, \breve{k}_{t-1}, \theta_1) = \frac{e^{-\lambda_t} \lambda_t^j}{j!}$$

and

$$g(x_t | k_t = j, \breve{x}_{t-1}, \breve{k}_{t-1}, \theta_2) \tag{42}$$

More specifically, the ACI model described in section 4 can be used to express λ_t as

$$\log(\lambda_t) = seasonals + \theta(L) \log(\lambda_{t-1}) + \Psi(L) k_{t-1} + \Pi(L) x_{t-1} \tag{43}$$

where the *seasonals* are a collection of dummies for time-of-day effects, day-of-week effects, and holiday effects. The corresponding lag polynomials are

$$\begin{aligned} \theta(L) &= \left(\theta_1 + ... + \theta_7 L^7\right) \left(1 - \theta_d L^{48}\right) \left(1 - \theta_w L^{336}\right) \qquad (44) \\ \Psi(L) &= \left(\psi_1 + ... + \psi_7 L^7\right) \left(1 - \psi_d L^{48}\right) \left(1 - \psi_w L^{336}\right) \\ \Pi(L) &= \left(\pi_1 + ... + \pi_7 L^7\right) \end{aligned}$$

that is, the dynamic formulation of the intensity allows for deterministic as well as multiplicative, stochastic, time-of-day and day-of-week seasonal effects. We include up to 7 lags to capture some of the periodicity in the "lunch" and other breaks that recur across the trading areas. The model for x_t is the following:

$$x_t = seasonals(1 + F_0(k_t)) + \Phi(L, k_t) x_{t-1} + \varepsilon_t$$

with

$$\Phi(L, k_t) = \phi_1(1 + F_1(k_t)) + \phi_2(1 + F_2(k_t))L + \phi_3(1 + F_3(k_t))L^2 \quad (45)$$

where $F_i(k_t)$ for $i = 0, 1, 2, 3$ is a non-parametric estimate based on a sixth order polynomial designed to capture the effects of the intensity level on the short-run dynamics of the spread variable. There are at least two ways in which the formulation of the model in (45) may appear incomplete. One is that we do not consider multiplicative, stochastic seasonal effects. The second is that we do not specify the variance in a dynamic way. However, to avoid distractions from the central issue that concerns this exercise, we opted for a more parsimonious model. The residuals of the fitted model did not show any evidence of ARCH effects which indicates that modelling the variance may not be central to learning about the short-run dynamics of the bid-ask spread.

Table 5 below compares the estimates of a simple Poisson count regression model exclusively based on the seasonal dummies against the ACI model in equations (42) and (43).

Table 5. Poisson and ACI Estimates

	Poisson	**ACI**
Log-Likelihood	-28562.37	-18145.52
# parameters	55	80
AIC	19.729	12.565
SIC	19.842	12.730
Ljung-Box Q_5	1608	99
Ljung-Box Q_{10}	1903	128
Ljung-Box Q_{50}	2398	372
LR test ACI vs. Poisson (p-value)	0.000	

These results are rather encouraging. The improvement on the overall fit of the data is quite remarkable by any measure. The Ljung-Box statistics reveal that the ACI model dramatically reduces the amount of left-over serial correlation in the residuals although there seems to be room left for improvement. The data has a mean of 87.33 and a standard deviation of 88.31 which suggests that the existing minimal amount of overdispersion is not a concern and that the Poisson assumption is a reasonable one for these data.

The second part of the exercise analyses the dynamics of the spread as a function of the level of activity. Figure 3 depicts the estimated autoregressive parameters as a function of the intensity. In the limit, as $k_t \to 0$ then $\phi_1 \to 1, \phi_2 \to 0$, and $\phi_3 \to 0$ as we should expect when the sampling frequency is so high as to record observations were no activity has elapsed. However, as the aggregation frequency becomes higher, the parameter estimates display a fair amount of non-monotonic variation, ranging from high persistence to negative correlation and back into higher levels of persistence. Figure 4 reports the fluctuation in the average, seasonally-adjusted residual spread as a function of the intensity. After accounting for the intra-day trading patterns, the spread exhibits two well defined peaks: One at low levels of activity and another when the intensity reaches 140 quotes per half-hour (recall that the average trading intensity in a regular business day is around 120 quotes per half hour). These results highlight the enormous amount of variability in the dynamic behavior of the spread when one accounts for the activity level in the market, a result that is consistent with the view that different intensity levels are related to different information flows, thus affecting the spread.

8 Conclusions

Time scale transformations are quite common in econometrics since there is often a mismatch between the decision time of agents and the data collecting time of statistical agencies, the former is usually substantially finer than the latter. Even when the original disaggregated data are available, the econometrician often chooses to analyze aggregated 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 switching model. Our empirical investigation of the USD-DM FX market illustrated how some of these techniques work in practice.

An interesting topic for further research is testing for time-scale transformations. In the case of misspecification tests, the aim is to check the assumptions underlying a constant parameter linear dynamic model that should be violated in the presence of time-scale transformations. Of course, they could be not satisfied for other reasons, such as intrinsic non-linearities or structural breaks, so that rejection of the null hypotheses should not be interpreted as direct evidence in favor of the presence of time-scale transformations. Examples include tests for parameter constancy, nonlinearity, autocorrelation, heteroskedasticity, normality, and ARCH effects. Despite these techniques, however, time aggregation problems and their solutions will heavily depend on the practitioner's knowledge of the economic problem that provides the backdrop for how the data was generated.

9 Appendix

Proof. Proposition 1: First we derive the AR component of x from that of x given that the aggregation scheme is point-in-time. For each period t, we want to find a polynomial, $B_t(Z)$, such that

$$B_t(Z)\Phi(Z) = 1 - c_{t,1}Z^{k_t} - c_{t,2}Z^{(k_t+k_{t-1})} - ... - c_{t,p}Z^{(k_t+k_{t-1}+...+k_{t-p+1})} = C_t(L) \tag{46}$$

Without placing any restriction on $B_t(Z)$, the coefficient on the i^{th} power of the lagged polynomial corresponding to the product $B_t(Z)\Phi(Z)$ coincides with the i^{th} element of the vector $\Gamma_t\beta_t + \gamma_t$. In order for (46) to hold, it must be that $\Gamma_t^*\beta_t + \gamma_t^* = 0$ from where $\beta_t = -(\Gamma_t^*)^{-1}\gamma_t^*$. Notice that the columns of Γ_t^* are linearly independent, so the matrix is full rank and its inverse always exists. The coefficients of $C_t(L)$ are the k_{t-j+1} rows of the vector $-\Gamma_t(\Gamma_t^*)^{-1}\gamma_t^* + \gamma_t$ 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 x).

Next, we derive the MA component. Define the following variables

$$\begin{aligned}\zeta_t &= B_t(Z)\Psi(Z)\varepsilon_{(b_t+q)} = \Pi_t(Z)\varepsilon_{(b_t+q)} \\ \zeta_{t-1} &= B_{t-1}(Z)\Psi(Z)\varepsilon_{(b_{t-1}+q)} = \Pi_{t-1}(Z)\varepsilon_{(b_{t-1}+q)} \\ &\dots\end{aligned} \tag{47}$$

such that

$$\begin{aligned}cov(\zeta_t, \zeta_t) &= \sum_{i=0}^{b_t+q} \pi_{t,i}^2\sigma^2 \\ cov(\zeta_t, \zeta_{t-j}) &= -\pi_{t,l}\sigma^2 + \sum_{i=1}^{(b_t+q-l)} \pi_{(t-j),i}\sigma^2\pi_{t,(l+i)} \qquad \text{for } j = 1, ..., r_t \\ cov(\zeta_t, \zeta_{t-j}) &= 0 \qquad \text{for } j > r_t\end{aligned} \tag{48}$$

where in general $r_t = p - 1$ (lower/higher values can be obtained when $p - q > k_t / q - p \geq k_t$). This is the autocorrelation function of a time-varying MA process, which has to be equal to that of the MA component in the generating mechanism of x. Therefore, the coefficients of the latter have to satisfy (9). The MA coefficients from the corresponding autocovariance function can be obtained through a Kalman filter approach (see Hamilton (1994), p. 391), or using the method in Tunnicliffe Wilson (1972). In addition, 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 observational 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. ■

Proof. Proposition 2: 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 $B_t(Z)\Phi(Z)x_\tau = D_t(L)\mathsf{x}_t$. 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 k_{t-j+1} rows 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. ■

The observational time Markov Switching model

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

$$f(\mathsf{x}_t | \breve{\mathsf{x}}_{t-1}, s_t = 1, \theta) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} exp\left\{\frac{-\left(\mathsf{x}_t - \rho_1\mathsf{x}_{t-1} - \rho_2\mathsf{x}_{t-2}\right)^2}{2\sigma_\varepsilon^2}\right\}$$

$$f(\mathsf{x}_t | \breve{\mathsf{x}}_{t-1}, s_t = 2, \theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2(1+\rho_1^2)}} exp\left\{\frac{-\left(\mathsf{x}_t - (\rho_1^2+\rho_2)\mathsf{x}_{t-1} - \rho_1\rho_2\mathsf{x}_{t-2}\right)^2}{2\sigma_\varepsilon^2(1+\rho_1^2)}\right\}$$

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

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

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

$$\widehat{\xi}_{t|t} = \frac{\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right)}{\mathbf{1}'\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right)} \tag{A2}$$

$$\widehat{\xi}_{t+1|t} = P \cdot \widehat{\xi}_{t|t} \tag{A3}$$

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 (49) and (49) for $t = 2, 3, ..., T$ to 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

$$L(\theta) = \sum_{t=2}^{T} \log f(\mathsf{x}_t|I_{t-1}, \theta)$$

where

$$f(\mathsf{x}_t|I_{t-1}, \theta) = \mathbf{1}'\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right) \tag{A4}$$

For a given θ, the value of the log-likelihood implied by that value of θ is given by (49). The value of θ that maximizes the log-likelihood can be maximized numerically. Further details on the estimation algorithm just described, inference on the transition probabilities p_{ij}, and forecasting can be found in Hamilton (1994).

Acknowledgements The authors wish to thank 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. Marcellino is grateful to the NBER and Harvard University for kind hospitality while working on this project.

References

Baillie, Richard T. and Patrick C. McMahon, (1989), **The Foreign Exchange Market: Theory and Econometric Evidence, Cambridge**; New York and Melbourne: Cambridge University Press.

Baillie, Richard T. (1996), "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73(1), 5-59.

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

Christiano, Lawrence J. and Martin Eichenbaum, (1987), "Temporal Aggregation and Structural Inference in Macroeconomics," *Carnegie-Rochester Conference Series on Public Policy*, 26, 63-130.

Dacorogna, Michel M., Ulrich A. Müller, R. J. Nagler, Richard B. Olsen, and Olivier V. Pictet, (1993), "A Geographical Model for the Daily and Weekly Seasonal Volatility in the FX Market," *Journal of International Money and Finance*, 12(4), 413-438.

Dacorogna, Michel M., C. L. Gauvreau, Ulrich A. Müller, Richard B. Olsen, and Olivier V. Pictet, (1996), "Changing Time Scale Short-Term Forecasting in Financial Markets," *Journal of Forecasting*, vol. 15, 203-227.

Diebold, Francis X., (1988), **Empirical Modelling of Exchange Rate Dynamics**, vol. 303 of Lecture Notes in Economics and Mathematical Systems, Berlin: Springer-Verlag.

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.

Dunsmuir, W. (1983), "A Central Limit Theorem for Estimation in Gaussian Stationary Time Series Observed at Unequally Spaced Times," *Stochastic Processes and Their Applications*, 14, 279-295.

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.

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

Engle, Robert F. and Jeffrey R. Russell, (1995), "Forecasting Transaction Rates The Autoregressive Conditional Duration Model," NBER, working paper 4967, Cambridge (MA).

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

Filardo, Andrew J. (1993), "Business Cycle Phases and Their Transitional Dynamics" Federal Reserve Bank of Kansas City, Working Paper 93-11.

Friedman, Milton and Anna J. Schwartz, (1982), **Monetary Trends in the United States and the United Kingdom Their Relation to Income, Prices, and Interest Rates, 1867-1975**, Chicago University of Chicago Press.

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

Granger, Clive W. J. and Zhuanxin Ding, (1996), "Some Properties of Absolute Return: An Alternative Measure of Risk," *Annales d'Economie et de Statistique*, 0(40), Oct.-Dec., 67-91.

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

Guillaume, Dominique M., Michel M. Dacorogna, Rakhal R. Davé, Ulrich A. Müller, Richard B. Olsen and Olivier V. Pictet, (1995), "From the Bird's Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets," Olsen & Associates, working paper DMG.1994-04-06.

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

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

Hamilton, James D., and Oscar Jordá, (1999), "A Model for the Federal Funds Rate Target," U.C. Davis, working paper 99-07.

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

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

Hendry, David F. and Neil R. Ericsson, (1991), "An Econometric Analysis of UK Money Demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz", *American Economic Review*, 81, 8-38.

Hsieh, David A. (1988), "The Statistical Properties of Daily Foreign Exchange Rates: 1974-1983," *Journal of International Economics*, 129-145.

Jordá, Oscar, (1999), "Random Time Aggregation in Partial Adjustment Models", *Journal of Business Economics and Statistics*, 17(3), 382-395.

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.

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

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

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

Müller, Ulrich A., Michel M. Dacorogna, Richard B. Olsen, Olivier V. Pictet, M. Schwarz, and C. Morgenegg, (1990), "Statistical Study of Foreign Exchange Markets, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis," *Journal of Banking and Finance*, 14, 1189-1208.

O'Hara, Maureen, (1995), **Market Microstructure Theory**, Oxford: Blackwell Publishers.

Robinson, Peter M. (1977), "Estimation of a Time Series Model from Unequally Spaced Data," *Stochastic Processes and Their Applications*, 6, 9-24.

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

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

Sims, Christopher A., James H. Stock, and Mark W. Watson, (1990), "Inference in Linear Time Series Models with Some Unit roots", *Econometrica*, 58, 113-144.

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

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

Taylor, S. J., (1988), **Modelling Financial Time Series**, Cichester: J. Wiley and Sons.

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

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

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

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

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

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

Figure 1

Seasonal Intraday Pattern
Diurnal Dummies

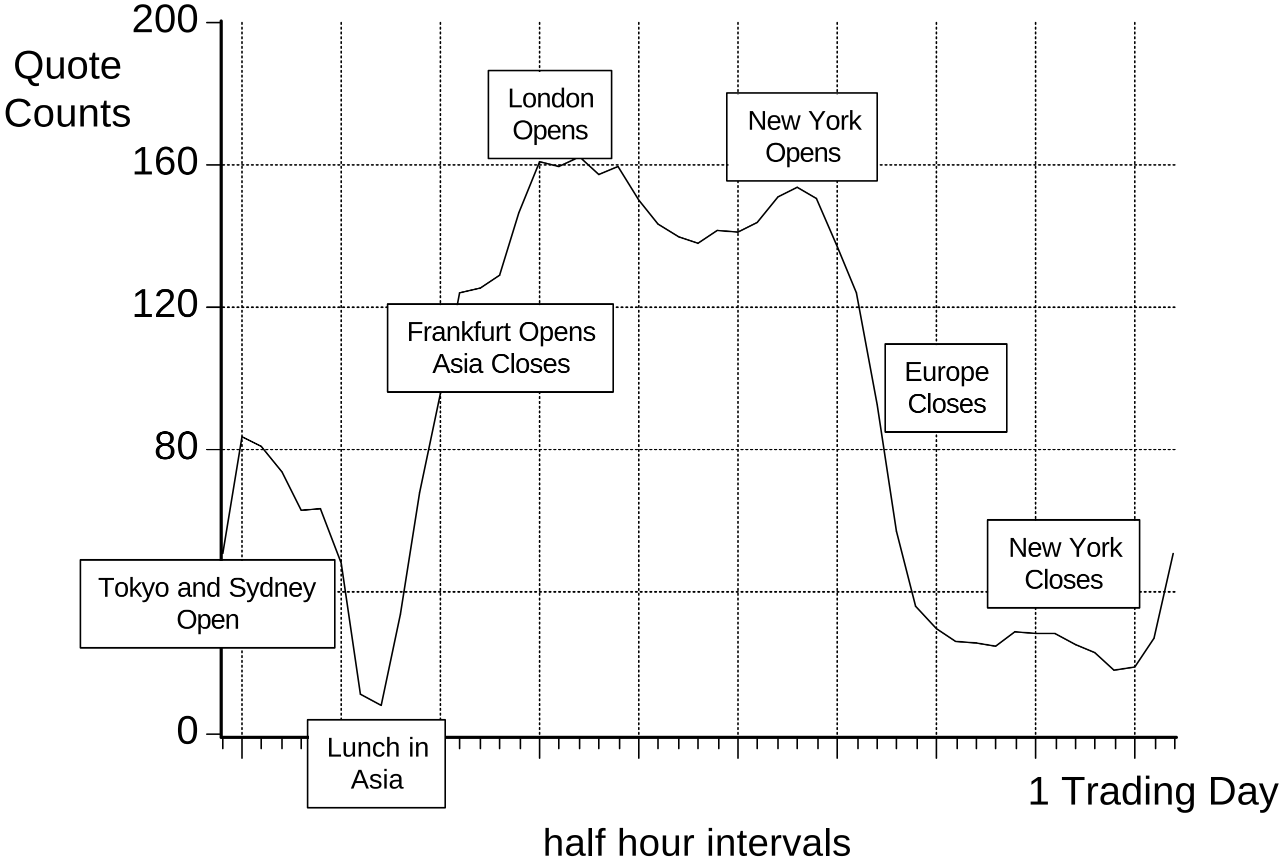

Figure 2

A Typical Week of Trading
Exchange Rate Quotes by 30 Minute Interval-Counts

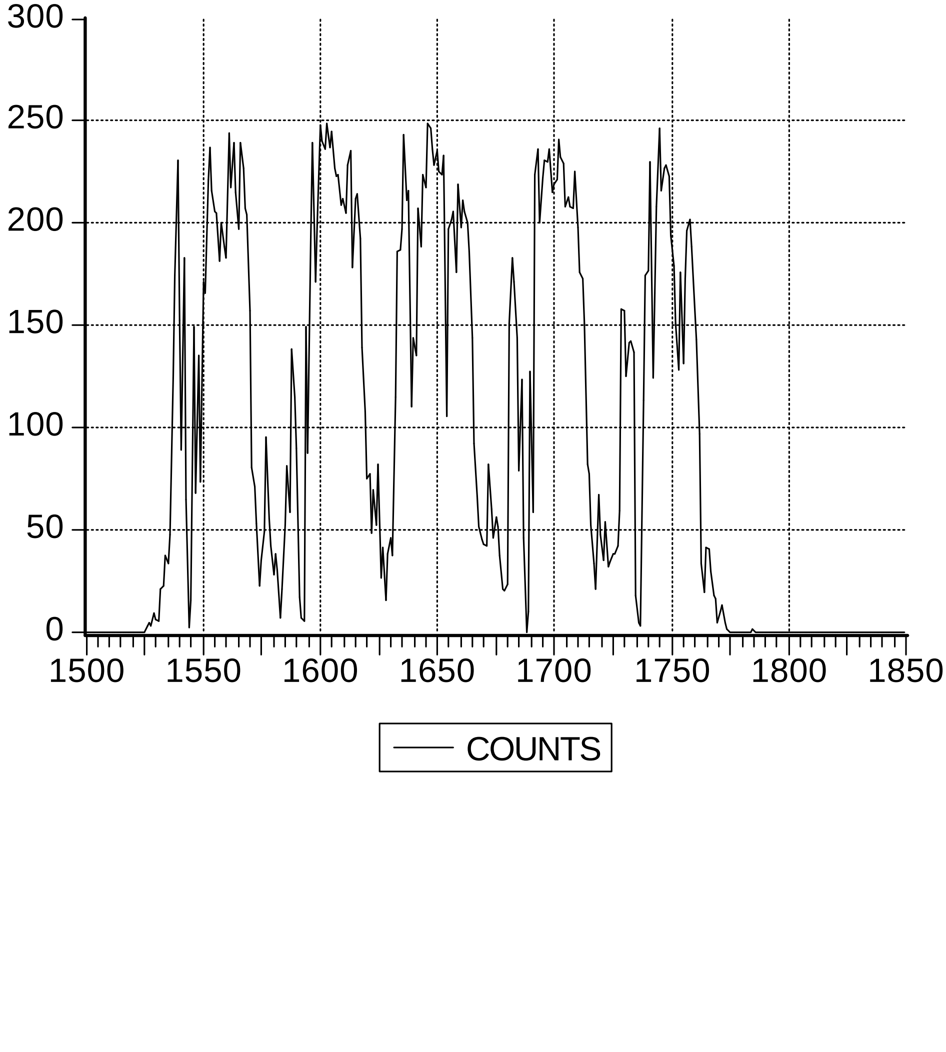

Figure 3

Nonparametric Estimates of the Autoregressive Parameters of the Spread Model as a Function of the Intensity

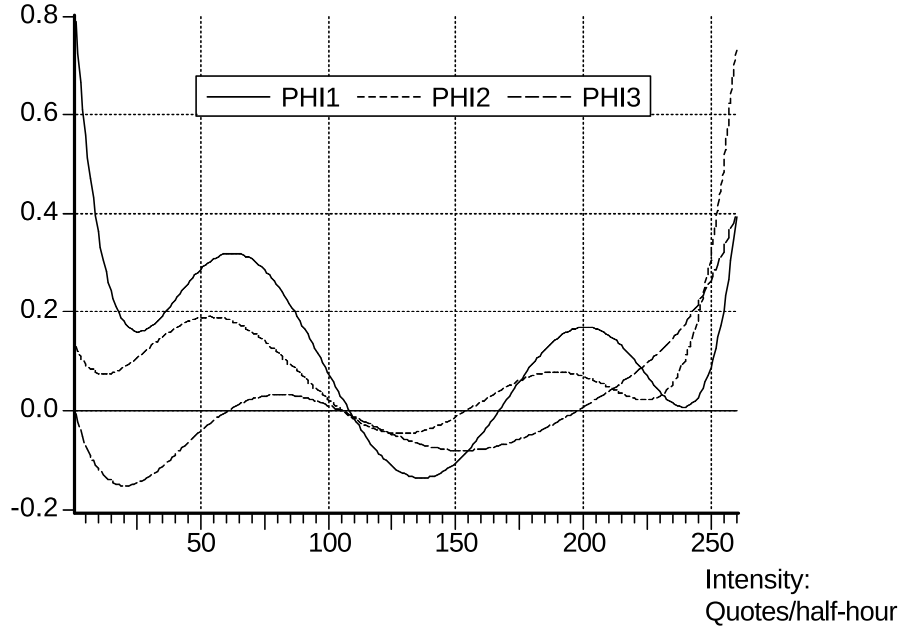

Figure 4

Nonparametric Estimate of the Mean Spread as a Function of the Intensity of Quote Arrivals