RESEARCH REPORT SERIES
(Statistics #2012-09)

Model Estimation, Prediction, and Signal Extraction for Nonstationary Stock and Flow Time Series Observed at Mixed Frequencies

Tucker McElroy
Brian Monsell

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: July 11, 2012

Model Estimation, Prediction, and Signal Extraction for Nonstationary Stock and Flow Time Series Observed at Mixed Frequencies

Tucker McElroy* and Brian Monsell†

* Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

† Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, brian.c.monsell@census.gov

Abstract

An important practical problem for statistical agencies and central banks that publish economic data is the seasonal adjustment of mixed frequency stock and flow time series. This may arise in practice due to changes in funding of a particular survey. Mathematically, the problem can be reduced to the need to compute imputations, forecasts, and backcasts from a given model of the highest available frequency data. The nonstationarity of the economic time series coupled with the alteration of sampling frequency makes the problem of model estimation and imputation challenging. For flow data the analysis cannot be recast as a missing value problem, and our methods are needed. We provide explicit formulas and algorithms that allow one to compute the log Gaussian likelihood of the mixed sample, as well as any imputations and forecasts. Formulas for the relevant mean squared error are also derived. We illustrate the techniques on two economic time series.

1 Introduction

The seasonal adjustment of economic time series is a vast undertaking (involving tens of thousands of time series) at statistical agencies and central banks – such as the U.S. Census Bureau, the Bureau of Labor Statistics, Statistics Canada, the Bank of Spain, the Bank of England, the Bundesbank, the International Labour Office, and many others – and most of production operates on univariate time series observed over a constant sampling frequency, typically either monthly or quarterly. The statistical methods have developed accordingly: X-11-ARIMA (Dagum, 1980) works with a single sampling frequency, applying the fixed X-11 filters to a forecast-extended time series[1]. Model-based approaches, such as in SEATS (Maravall and Caparello, 2004), also proceed by considering a single frequency. However, modifications in survey construction often change fundamental characteristics of an observed time series, even altering the sampling frequency – from higher to lower or lower to higher. How are the seasonal adjustment techniques to be modified in this situation?

The problem of seasonally adjusting mixed frequency data is not an isolated concern. Motivation for the work of this paper stemmed from mixed frequency series being processed at the Bank of England, brought to our attention by Fida Hussain. The problem is mentioned as an explicit concern – related to the question of comparability between successive surveys – by Eurostat; see point number 2 of

`http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/lfs_main/lfs/lfs_comparability.htm`

For developing countries (e.g., South Africa), many of which are now transitioning to more frequent measurement, the problem of mixed frequency data is acute; see Pasteels (2012). The International Labour Office has many mixed frequency series; see `http://laborsta.ilo.org/sti/sti_E.html` for an example of South African unemployment rate, which transitioned from bi-annual to quarterly frequency in 2008.

The topic is somewhat related to the literature on benchmarking and reconciliation, which however has tended to follow a nonparametric approach, with some noteworthy exceptions (such as Durbin and Quenneville (1997)). There is also a substantial literature on time series analysis of mixed frequency data (including Zadrozny (1990) and Chen and Zadrozny (1998)). However, this latter literature focuses on stationary data. What seems to be missing is a treatment of nonstationary time series observed as stock or flow across two or more sampling frequencies, with regard to the following questions: how does one identify a model? How is the model fitted? How does one do forecasting, imputation, and signal extraction? How does one quantify the uncertainty attending these estimates?

Perhaps the most obvious approach is to use the Kalman filtering methodology to address parameter estimation and projection calculations. One simply formulates the underlying process as corresponding to the highest observed frequency, writing down the observation equations accordingly; the mathematics of model fitting and state space smoothing then become equivalent to a missing value problem when the process is a stock. See Durbin and Quenneville (1997) and Durbin and Koopman (2001) for related material.

There are some drawbacks to this approach. First, state space methods require a substantial programming effort, although one may use off-the-shelf routines such as those offered through

[1] This is an approximate view of the procedure – exact details can be found in Ladiray and Quenneville (2001).

Ox (Doornik, 1998), which is not inexpensive. Second, the correct quantification of uncertainty associated with estimation of fixed signals (e.g., seasonally adjusted components defined via the X-11 filters) requires a full knowledge of the projection error covariance matrix. Some expertise and care is needed to produce these quantities from the state space method, which typically will just output the projection error variances, or diagonal entries of the error covariance matrix (see Koopman et al. (1999) and Durbin and Koopman (2001)). Third, numerical devices are typically utilized to evaluate the likelihood, such as setting the variance of a certain initial distribution to be a numerical infinity (i.e., a large float) – this is called the diffuse initialization. Of course, the exact diffuse initialization can be used instead, or better yet the "optimal" initialization of Bell and Hillmer (1991). With any of these state space approaches, there is no explicit formula for the likelihood (let alone forecasts and imputations), this being produced instead as the end product of a series of recursive formulas. Fourthly, the missing value approach alluded to above cannot be used for flows, since a low frequency flow does not correspond to a systematic subsample of the high frequency flow.

However, the advantages are numerous, and are generally felt to outweigh any weaknesses: computational efficiency is foremost, as well as flexibility and the power to handle many types of applications (Durbin and Koopman, 2001). Broadly speaking, these applications mainly fall in the category of calculating Gaussian conditional expectations, or equivalently, minimum mean square error linear estimates of unknown stochastic quantities. Following the language of Brockwell and Davis (1991), we refer to such conditional expectations as projections. Examples of projections include forecasts, missing value imputations, and signal extractions.

The main contribution of this paper is the derivation of certain mathematical results regarding the Gaussian likelihood – facilitating model estimation for mixed frequency data – as well as the derivation of projection matrix formulas for the mixed frequency situation described above. Depending on one's perspective on state space methods, the formulas may be viewed as a precise mathematical foundation for smoothing methods (they correspond to the state space estimates generated under Bell and Hillmer (1991)'s optimal initialization), or as a straight-forward matrix-based method for computing projections when recourse to the Kalman filter is infeasible (e.g., in the case of a flow, or when a long memory model is being utilized) or impractical.

Our results demonstrate the key importance of initial value conditions of the nonstationary stochastic process; without these assumptions, all projection calculations will depend upon nuisance parameters that lie beyond the scope of typical time series models. This observation pertains to both the log Gaussian likelihood – and hence to model fitting – and to projection results proper, as delineated in Sections 2 and 3 respectively. Hence our formulas provide additional insight into projections in time series analysis. In Section 4 we proceed with a motivating case study, an inventory series of food products, available as a mixed sample of quarterly and monthly frequencies. We demonstrate the modified X-11 seasonal adjustment procedure on this case, producing estimates at the full data span at the highest sampling frequency. We also analyze mixed frequency time series from both the Bank of England and the Bundesbank, and summarize our findings. Section 5 presents our conclusions, and mathematical proofs are in the Appendix.

2 Modeling Mixed Frequency Data

2.1 Time Series Data Observed at Multiple Sampling Frequencies

We now discuss our basic working assumptions. We assume that the available data consists of either stock or flow time series observations available at two or more frequencies. This data has the following characteristics by assumption:

- The data occurs with multiple sampling frequencies.
- The data at each sampling frequency is difference-stationary (with known differencing polynomial).
- The observed data is either a stock or a flow.
- It is possible – in the flow case – that some time points may have several observations, one at each sampling frequency.

The fourth point above is rather trivial for stocks, since when passing to different sampling frequencies for a particular epoch, the identical numerical value is obtained. For example, the third monthly and the first quarterly observations for a given year are identical quantities if the series is a stock; but if the series is a flow, it is the aggregation of the first three monthly values that equals the first quarterly number.

So for a stock, the given data may be viewed as belonging to any one of a number of sampling frequencies – in order to pass to lower frequencies, one simply applies systematic temporal sampling to the higher frequency data. For this reason too, one can view any lower frequency series as really a higher frequency series that has systematic missing values. This is not true for flows. For example, a quarterly flow series does not correspond to a sub-sampled monthly flow series. If we know the February and March values of a flow, it does not determine the first quarterly number, since it is equal to the sum of January, February, and March. Hence, we see that for flow time series, a mixed frequency data set cannot be embedded as a missing value high frequency time series, as we can for stock data.

The observed data can generally be written as a column vector. The exact sequencing of observations is not unique for flows, since there may be overlapping observations, but for stocks it is possible to order the data chronologically by intermingling different frequencies as needed. But the ordering of the data vector is not really important, so long as the analyst knows how the data is related to the various frequencies. The key assumption of our method is that all the observations arise as a linear combination of a (potentially partially unobserved) highest frequency data vector, which we will call $Y = (Y_1, Y_2, \cdots, Y_n)'$. This is a feasible assumption, because we assume that our mixed sample consists exclusively of stock data or exclusively of flow data. Lower frequency stock values are obtained from the highest frequency stock value by stratified sampling. But for flows, the lower frequency values are obtained by aggregation of the highest frequency. Hence the observed data takes the form JY, where J is a selection matrix that is described below.

Letting $\{Y_t\}$ denote the sequence of values for the highest frequency time series (as either stock or flow), we suppose that it is difference stationary such that $W_t = \delta(B)Y_t$ is mean zero and covariance stationary, where $\delta(z) = \sum_{j=0}^{d} \delta_j z^j$ is a degree d unit-root polynomial. Let the autocovariance function (ACF) of $\{W_t\}$ be denoted by γ_h. We can conveniently represent Y_t in terms of W_t and d so-called initial values, using results from Bell (1984). Actually, we extend this representation below to consider both backward differencing and forward differencing at the same time.

The highest frequency process $\{Y_t\}$ will also include fixed effects, such as calendar and trading day effects. We proceed by specifying a model for $\{Y_t\}$, noting that models for all other sampling frequencies are implied via the sampling relations, although the resulting models may no longer belong to recognizable model classes.

2.2 Initial Values for Mixed Frequency Time Series

We can write the observed data as $X = JY$ with J given by ones and zeros, where $\{Y_t\}$ is described above. The high-frequency vector Y has length n, but the vector of observed data has length m, so J is $m \times n$. Assuming for now that Y has mean zero – the case of nonzero mean is treated in the next subsection – the log Gaussian likelihood multiplied by -2 is given by

$$X' \left(J \Sigma_Y J' \right)^{-1} X + \log |J \Sigma_Y J'| \qquad (1)$$

up to irrelevant constants. Note that this objective function is derived from the Gaussian joint probability density for the undifferenced data $X = JY$. This contrasts with the single frequency case where one works with differenced data. Actually, the validity of using the likelihood of differenced data is contingent on a factorization of the likelihood function, which in turn depends on the orthogonality of initial values with $\{W_t\}$. But in the multiple frequency case things are not so simple. First, it may not be clear how to do differencing at all – since differencing typically requires d contiguous values in JY of the same frequency. There is no reason to suppose this contiguity in JY exists. Second, the factorization of the likelihood is far from obvious, and need not always occur (see below for a counter-example).

We now proceed to investigate the general case for a difference stationary process, with initial values uncorrelated with $\{W_t\}$. Under what conditions does the likelihood factor in the way needed?

The desired factorization has the following properties:

$$X'\Sigma_X^{-1}X = Q_1(Y_*, \Sigma_*) + Q_2(W, \Sigma_W), \tag{2}$$

where Y_* consists of any d components of Y, and are the initial values (in analogy with the treatment of Bell (1984) and McElroy (2008)), and Σ_* is the covariance matrix of Y_*. Also $W = [W_{d+1}, \cdots, W_n]'$ is the differenced sample, with $n-d$-dimensional covariance matrix Σ_W. The functions Q_1 and Q_2 must be quadratic in their first arguments, but we leave their exact form unspecified for the definition. Note that the initial values Y_* need not be the first d (or last d) values of Y, as is typical in the single frequency literature; cf. the discussion in section 4 of Bell and Hillmer (1991) on different initial values. We provide a formal proof of this assertion in Proposition 1 below.

In order to factorize the Gaussian likelihood in a convenient way, it is necessary that a certain algebraic property holds for (2). To this end, we may think of X, Y_*, and W as fixed non-random vectors, or as realizations of the respective random vectors. We say that the Gaussian likelihood factorizes iff (2) holds, such that Q_1 does not depend on W or Σ_W, and Q_2 does not depend on Y_* or Σ_*. Again, this is purely a property of linear algebra. From this definition it follows that the gradient with respect to W of $Q_1(Y_*, \Sigma_*)$ is zero, and the gradient with respect to Y_* of $Q_2(W, \Sigma_W)$ is zero as well[2].

We first establish a preliminary result that demonstrates that any d (non-consecutive) values of Y may be used as initial values of the process. First we introduce additional notation. Since Y is the highest frequency (unobserved) data vector, it is differenced to stationarity by an $n-d \times n$ dimensional matrix Δ:

$$\begin{bmatrix} \delta_d & \cdots & \delta_1 & \delta_0 & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \cdots & 0 & \delta_d & \cdots & \delta_1 & \delta_0 \end{bmatrix}$$

Then $W = \Delta Y$ is mean zero and has a Toeplitz covariance matrix Σ_W. We claim that the initial values can be drawn anywhere from the unobserved vector Y. This can be described as saying they are equal to the first d components of some n-dimensional permutation matrix P applied to Y. Augment Δ with the first d rows of this P such that

$$\widetilde{\Delta}(P) = \begin{bmatrix} [1_d \; 0]\, P \\ \Delta \end{bmatrix}. \tag{3}$$

(We denote an ℓ-dimensional identity matrix by 1_ℓ, whereas 0 is a rectangular matrix of zeroes

[2] While there is a connection to the factorization problem discussed in Cochran's Theorem (see Cochran (1934) and Scheffe (1959)), that result does not apply here; letting $Z = [Y_*', W']'$, Theorem 1 of Scheffe (1959) is concerned with decomposing $Z'Z$ into $Q_1(Y_*, \Sigma_*)$ and $Q_2(W, \Sigma_W)$, the right hand side of (2). However, the left hand side of (2) is the more complicated form $X'\Sigma_X^{-1}X$, rather than just $Z'Z$.

of dimension dictated by context.) Then applying $\widetilde{\Delta}(P)$ to Y yields a vector with the first d components given by Y_*, followed by W. We now state our first result about initial values.

Proposition 1 *Any difference stationary process $\{Y_t\}$ with differencing polynomial $\delta(B)$ of degree d can be written as*

$$Y_t = \widetilde{A}_t' Y_* + \sum_{j=d+1}^{t} b_j W_j$$

for any d initial values $Y_* = [Y_{t_1}, Y_{t_2}, \cdots, Y_{t_d}]'$. Here $W_t = \delta(B)Y_t$ is stationary, $\widetilde{A}_t$ is a d-vector of time-varying functions, and $\{b_j\}$ are real coefficients. Moreover, there is an invertible mapping between Y_* and any other d initial values.

So we are free to choose initial values at our convenience; as our results below demonstrate, it is advantageous to choose them from among the available observed highest frequency observations in X, because then the likelihood is guaranteed to factor. First, we establish necessary and sufficient conditions for this factorization, given a particular choice of initial values; the following proposition shows that the likelihood factorizes iff there exists some invertible matrix that makes $J\widetilde{\Delta}^{-1}(P)$ appropriately block diagonal. Note that initial values, wherever they are chosen to lie within an actual time series, are typically assumed to be uncorrelated with the differenced process (cf., Bell (1984)); this is a fundamental – and unverifiable – condition used for all sorts of projection results in time series analysis.

Proposition 2 *Assume there is a permutation matrix P such that the initial values $Y_* = [1_d \; 0]PY$ are uncorrelated with the differenced process $\{W_t\}$. Then the Gaussian likelihood factorizes (i.e., equation (2)) iff there exists an invertible m-dimensional matrix R such that*

$$RJ\widetilde{\Delta}^{-1}(P) = \begin{bmatrix} \overline{A} & 0 \\ 0 & \underline{B} \end{bmatrix},$$

for matrices $\overline{A}$ and $\underline{B}$ that are $d \times d$ and $m - d \times n - d$ dimensional, respectively, with $\widetilde{\Delta}(P)$ as in (3).

Since $J\widetilde{\Delta}^{-1}(P)$ can be readily computed for each choice of P, we can verify the condition of Proposition 2 at once by doing a QR decomposition (Golub and Van Loan (1996)). The choice of initial values is equivalent to the choice of P, and may seem to be a conundrum (Proposition 2 makes its assertion for any single given set of initial values, but does not require the orthogonality condition simultaneously for all choices of initial values, which is typically impossible), but in practice for many cases at least d observations of the highest frequency data Y are observed. (If fewer than d of the highest frequency is available, it is really questionable whether there is any merit to including the highest frequency data in the model fitting exercise.) In this case, *any* d of the observed Y values can serve as initial values, as the following theorem demonstrates.

2.3 Maximum Likelihood Estimation of Mixed Frequency Time Series

For simplicity of exposition, we focus on the case that the d initial values actually occur as the first d values of X (this can always be accomplished, without loss of generality, by row permuting the original data vector X into one that has the desired structure). This assumption can be stated mathematically as $[1_d \; 0]J = [1_d \; 0]P$.

Theorem 1 *Let the mixed sample X of size m be written as $X = JY$ for a high frequency vector Y that is a sample of size n from a difference stationary process with degree d differencing polynomial $\delta(B)$. Suppose that the initial values $Y_* = [1_d \; 0]PY$ are uncorrelated with $W = \Delta Y$, and are observed as the first d values of X, i.e., $[1_d \; 0]J = [1_d \; 0]P$. Then the Gaussian likelihood factorizes (viz. Proposition 2), with*

$$R = \begin{bmatrix} 1_d & 0 \\ -\underline{A} & 1_{m-d} \end{bmatrix} \qquad RJ\widetilde{\Delta}^{-1}(P) = \begin{bmatrix} 1_d & 0 \\ 0 & \underline{B} \end{bmatrix},$$

where $\underline{A}$ and $\underline{B}$ are respectively $m - d \times d$-dimensional and $m - d \times n - d$-dimensional matrices, and are the bottom rows of $J\widetilde{\Delta}^{-1}(P)$, i.e.,

$$[\underline{A} \; \underline{B}] = [0 \; 1_{m-d}] \, J\widetilde{\Delta}^{-1}(P). \tag{4}$$

Moreover, $D = [0 \; 1_{m-d}]R = [-\underline{A} \; 1_{m-d}]$ differences X in the sense that

$$DX = \underline{B}W \tag{5}$$

is a linear combination of stationary random variables. Finally, -2 times the log Gaussian likelihood can be written as $Y_*' \Sigma_*^{-1} Y_* + \log|\Sigma_*|$ plus

$$(DX)' \big(\underline{B} \Sigma_W \underline{B}'\big)^{-1} DX + \log|\underline{B}\Sigma_W \underline{B}'|, \tag{6}$$

up to irrelevant constants.

The exact expression (6) for the factorized Gaussian likelihood can be compared to the State Space (SS) approach to the problem. In Durbin and Koopman (2001) a description is given of exact diffuse initialization, which amounts to letting the variances of the initial values tend to infinity in the calculations of conditional variances. It is evident from Theorem 1 that this strategy would produce a value of infinity in the log Gaussian likelihood, but it is unnecessary in any event – since no model parameters are featured in the initial value terms, one can work with (6) alone. Since Result 2 of Bell and Hillmer (1991) – when applied to the problem of computing time series residuals (or innovations) optimally – indicates that their initialization of the Kalman filter produces conditional expectations (for Gaussian data), it follows that their likelihood and our expression in Theorem 1 are identical. Bell and Hillmer (1991) demonstrate that their "optimal" initialization of the Kalman filter (based on the transformation approach of Ansley and Kohn (1985)) can produce different results from a diffuse initialization. However, while the Bell and Hillmer (1991) result furnishes the optimal Gaussian likelihood using a SS approach, our Theorem 1 produces an analytic formula that is easily computable, allowing for treatment of time series models that cannot be embedded in SS (e.g., exponential models or long memory models).

In order to better understand why it is necessary to have d observations available of the highest frequency data, we discuss a counter-example. Consider $\delta(B) = (1-B)^2$ for a flow observed process, with two observations corresponding to a monthly value as well as a quarterly value, such that

$$J = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

The unobserved high frequency vector $Y = [Y_1, Y_2, Y_3]'$ can be differenced to stationarity by left application of the matrix

$$\widetilde{\Delta}(1_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}.$$

In particular $\widetilde{\Delta}(1_3)Y = [Y_1, Y_2, W_3]'$, which is a convenient decomposition into initial values Y_1, Y_2 and differenced data W_3. Typically these initial random variables are assumed to be uncorrelated with the differenced process $\{W_t\}$, which facilitates computation of the Gaussian likelihood. Then it follows that

$$X = JY = J\widetilde{\Delta}^{-1}(1_3) \begin{bmatrix} Y_1 \\ Y_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ W_3 \end{bmatrix}.$$

Proposition 2 indicates that factorization is impossible, because $J\widetilde{\Delta}(1_3)$ does not have the right structure. Let Σ_* denote the 2-dimensional covariance matrix of Y_1, Y_2, with jkth entry denoted σ_{jk} for short. Recalling that $\{\gamma_h\}$ is the ACF of $\{W_t\}$, it follows that

$$\Sigma_X = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \Sigma_* \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \gamma_0 \end{bmatrix}.$$

Hence we can write down $X'\Sigma_X^{-1}X$, and demonstrate that it does not factor into two portions that are each solely dependent on initial values Y_1, Y_2 and W_3 respectively. The quadratic form is

$$X'\Sigma_X^{-1}X = \left(\sigma_{11}[9\sigma_{22} + \gamma_0] - 9\sigma_{12}^2\right)^{-1} \begin{bmatrix} Y_1 \\ Y_2 \\ W_3 \end{bmatrix}' \begin{bmatrix} 9\sigma_{22} + \gamma_0 & -9\sigma_{12} & -3\sigma_{12} \\ -9\sigma_{12} & 9\sigma_{11} & 3\sigma_{11} \\ -3\sigma_{12} & 3\sigma_{11} & \sigma_{11} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ W_3 \end{bmatrix},$$

which cannot factorize in the way desired (unless $\sigma_{11} = 0 = \sigma_{12}$, which makes the expression undefined).

In summary, our expression for the likelihood (6) – which is contingent on having at least d of the highest frequency values observed – is the same as that arising from the Kalman filter with an optimal initialization. Whether it agrees with non-optimal initializations – such as the diffuse – is an empirical matter; these alternative initializations may be approximately correct in many cases. However, for short time series there is no real disadvantage in using the explicit likelihood of Theorem 1; moreover, certain cases (e.g., long-range dependence) preclude the use of SS methods except as an *ad hoc* method. To use our expression for the likelihood, note that – since no parameters enter into Σ_*, it can be ignored – we may focus on minimizing the quadratic form in DX. To summarize, our procedure is:

1. Identify any d observed high-frequency values o f $\{Y_t\}$ and rewrite X so that these occur in the first d time points.
2. Write down J such that $X = JY$, with Y_* the d initial values (and the first d values of X).
3. Determine P corresponding to these initial values, and compute $\widetilde{\Delta}(P)$ via equation (3).
4. Compute $\widetilde{\Delta}^{-1}(P)$.
5. Compute $\underline{A}$ and $\underline{B}$ from $J\widetilde{\Delta}^{-1}(P)$ via (4).
6. Compute D and DX as in (5).
7. Compute Σ_W from the model for $\{W_t\}$ and compute $\left(\underline{B}\Sigma_W\underline{B}'\right)^{-1}$.
8. Evaluate (6).

Only the last two steps need be repeated over different parameter values – all else can be calculated once. Combined with a numerical optimization algorithm, the maximum likelihood estimates (MLEs) can then be obtained. If a Bayesian analysis is desired instead, the same algorithm can be used for likelihood evaluation, finally taking the exponential of $-1/2$ times the computed expression. (The portions involving initial values will be integrated out in any computation of posterior densities.)

2.4 Fitting Models with Regression Effects

We now provide a more nuanced discussion of model fitting for mixed frequency data. Let us here suppose that the aggregation relations between different frequencies hold exactly, as described in section 2.2. In other words, $X = JY$ describes an exact linear relationship between observations and the highest frequency time series. Then it is clear that all covariance structure for lower frequencies is generated by a specified covariance structure at the higher frequencies. In other words, we may consider specifying a model for the highest frequency data, and this automatically determines an implied model for all lower frequencies. Now if the high frequency data follows an ARMA model, the lower frequencies need not follow this same specification. They may not even be ARMA. But we still know how to compute their covariances, and in this sense their model is determined.

Fitting of the model proceeds via maximum likelihood estimation, as described above. We provide a bit more detail below. The fit can be checked by examining the time series residuals – also defined below – for serial correlation. One could also use AIC values based on the likelihoods of competing models, in order to choose between them.

Now a reasonable model will often include regression effects. It is easiest to specify these at the highest frequency, knowing that some of them may not even manifest at lower frequencies. So suppose that the high frequency unobserved data vector is $\mathcal{Y} = \mathbb{X}\beta + Y$, with parameter vector β and design matrix $\mathbb{X}$. Then our observed data follows

$$X = J\mathcal{Y} = J\mathbb{X}\beta + JY.$$

So the new design matrix for the observed mixed data is $J\mathbb{X}$. If a higher frequency effect is not at all present at lower frequencies (e.g., a monthly trading day effect has no impact on annual data), then this absence will be automatically accounted for – the corresponding rows and columns of $J\mathbb{X}$ will then just be zeroed out. It follows from the results above that the Gaussian log likelihood is equal to $-1/2$ times

$$\begin{aligned}
&(X - J\mathbb{X}\beta)'\Sigma_X^{-1}(X - J\mathbb{X}\beta) + \log|\Sigma_X^{-1}| \\
&= (X - J\mathbb{X}\beta)'R'\begin{bmatrix}\Sigma_*^{-1} & 0 \\ 0 & (\underline{B}\Sigma_W\underline{B}')^{-1}\end{bmatrix} R\,(X - J\mathbb{X}\beta) - \log|\Sigma_*| - \log|\underline{B}\Sigma_W\underline{B}'| \\
&= (Y_* - [1_d\,0]P\mathbb{X}\beta)'\Sigma_*^{-1}(Y_* - [1_d\,0]P\mathbb{X}\beta) - \log|\Sigma_*| \\
&+ (DX - \underline{B}\Delta\mathbb{X}\beta)'\big(\underline{B}\Sigma_W\underline{B}'\big)^{-1}(DX - \underline{B}\Delta\mathbb{X}\beta) - \log|\underline{B}\Sigma_W\underline{B}'|.
\end{aligned}$$

This uses (A.2) from the Appendix. Consider the final expression above. The first term is not constant with respect to the model parameters, since β occurs there. Optimizing with respect to β, for any value of the other model parameters, produces the MLE

$$\widehat{\beta} = \left[\mathbb{X}_*'\Sigma_*^{-1}\mathbb{X}_* + \mathbb{X}'\Delta'\underline{B}'\big(\underline{B}\Sigma_W\underline{B}'\big)^{-1}\underline{B}\Delta\mathbb{X}\right]^{-1}\left[\mathbb{X}'\Delta'\underline{B}'\big(\underline{B}\Sigma_W\underline{B}'\big)^{-1}\underline{B}W + \mathbb{X}_*'\Sigma_*^{-1}Y_*\right].$$

Here $\mathbb{X}_* = [1_d\ 0]P\mathbb{X}$. It is problematic that this MLE depends upon the nuisance parameters in Σ_*. This problem afflicts the single frequency case as well, and is typically resolved by ignoring the unknown portions (which is conceptually effected by letting Σ_* diverge to infinity in the formula). Plugging the resulting $\widehat{\beta}$ into the log likelihood produces a concentrated likelihood, which only depends on the parameters for the model of $\{W_t\}$. Or one may directly optimize the likelihood over both the regression and time series model parameters, either jointly or via an iterative algorithm. In this case, one must work with just the final two summands, which together correspond to (6) with DX replaced by $DX - \underline{B}\Delta\mathbb{X}\beta$.

Note that an alternative approach to the treatment of missing values (say for a stock series) is to replace all of them with some large fictitious number, and then insert an Additive Outlier (AO) regressor at each such time point[3]. This produces a different (and incorrect) likelihood, and moreover introduces additional parameters and *ad hoc* user-defined values, and so is not recommended for handling mixed frequency stock data (where the number of missing values is potentially large).

Once the MLEs are determined, we can heuristically assess model fit via the time series residuals defined via

$$\epsilon = \left(\underline{B} \widehat{\Sigma}_W \underline{B}' \right)^{-1/2} \left(DX - \underline{B} \Delta \mathbb{X} \widehat{\beta} \right),$$

where the MLEs for parameters are utilized, hence the notation $\widehat{\Sigma}_W$ and $\widehat{\beta}$. If the model is correctly specified, and all MLEs were to converge to their asymptotic values, the above vector would be *iid* standard normal. Plotting these residuals, along with their sample ACFs, can allow one to assess model fit. If a problem is found, we can try a new model and repeat the whole procedure. An asymptotic theory for such diagnostics is difficult to formulate, since assumptions about future changes in sampling frequency must be accounted for somehow; we won't pursue this further.

2.5 Treatment of Logged Flows

When considering mixed frequency flow time series, it is not possible to embed changing frequency as a missing value problem. This is because any lower frequency flow is not identical with a (regular) sub-sampling of a higher frequency flow. Put another way, there is more than one nonzero entry in each row of J. Hence a SS approach with missing values cannot be used for flows; instead one must consider a more nuanced observation equation, which amounts to transforming underlying states of the highest frequency time series by a matrix. The theory of sections 2.2 and 2.3 provides the general necessary and sufficient conditions for factorization.

However, if the data first undergo a Box-Cox transformation, then flow aggregation relations are no longer linear, and $X = JY$ no longer holds as an exact relation. We illustrate this first with monthly (y_1, y_2, and y_3) and quarterly (x_1) data. In the original scale, the flow property states that $x_1 = y_1 + y_2 + y_3$, when x_1 corresponds to the quarter comprising the months described by y_1, y_2, y_3. Suppose that we wish to model the data in logarithms, this being a sensible variance-stabilizing transform. Denote the log-transformed variables with capital letters. Then $X_1 \neq Y_1 + Y_2 + Y_3$, which is a nuisance. Note that this problem does not arise for stocks (where $x_1 = y_3$ maps to $X_1 = Y_3$ without hindrance).

By Jensen's inequality, we always have $X_1 \geq \log 3 + (Y_1 + Y_2 + Y_3)/3$. More precisely, we can write

$$\log x_1 = \frac{Y_1 + Y_2 + Y_3}{3} + \log \left[\left(\frac{y_1^2}{y_2 y_3} \right)^{1/3} + \left(\frac{y_2^2}{y_1 y_3} \right)^{1/3} + \left(\frac{y_3^2}{y_1 y_2} \right)^{1/3} \right].$$

[3]This method is implemented in the X-12-ARIMA seasonal adjustment program; see Section 7.14 of U. S. Census Bureau (2011) for more details

This latter term is bounded below by log(3), and in practice is close to the lower bound when y_1, y_2, y_3 have values reasonably close to one another. This analysis indicates that we might write $X \approxeq JY + \mu\iota$, where the rows of J that do flow aggregation have values of 1/3 (instead of one in the original data scale), and $\mu = \log(3)$ and ι is a vector of zeroes and ones. The approximation error is hopefully small in practice, such that one proceeds with the algorithm as if it were exact (more research is needed to ascertain the extent to which, empirically speaking, this approximation is harmless on real flow series). Furthermore, if any quarter appears alongside with all its constituent months, we may strike out that quarter's row since it really imparts no additional information over the three months. Then each quarter that appears contains at most two months that are available elsewhere in the sample, so that at least one month that makes up the quarter is opaque (i.e., unobserved).

As for $\mu\iota$, this will be subsumed in the regression effects $J\mathbb{X}$; we have $X \approxeq JY + J\mathbb{X}\beta + \mu\iota$, and the latter two terms can be combined into a signal regression expression. However, μ is not a parameter to be estimated, but rather a known quantity that forms an offset to the data X. More generally, suppose that several lower frequencies are available, such that each is the aggregation of m_k highest frequency values, for various k indexing the lower frequency portions. The matrix J for the logged data should have each unit value replaced by $1/m_k$, if that particular row corresponds to data at the kth lowest frequency. (Above, we had $m_1 = 3$.) The compensating mean vector is now written $\mathbb{Z}\mu$ for a vector of known values μ and $\mathbb{Z}$ consisting of the corresponding regressors. These μ values are equal to $\log(m_k)$ in a row corresponding to the kth lowest frequency data. Then $X \approxeq JY + [J\mathbb{X}\ \mathbb{Z}][\beta', \mu']'$, and we treat the approximation as an exact equality.

3 Projection: Forecasting and Signal Extraction

Many applications can also be handled using the results of Section 2. If we are interested in optimal estimates of past or future values, i.e., backcasts and forecasts, then this can be solved through the theory of projections. Missing data, i.e., omissions in the observed data JY at any frequency, is handled with the same theory. More generally, any linear function of the high frequency vector Y can be optimally estimated. The general theory is well-known, going back to Parzen (1961); its application to the particular case of mixed frequency nonstationary time series data is given below.

Consider a vector of "target" quantities Z, written as a column vector of length r, which can be expressed as a linear combination of the highest frequency data series $\{Y_t\}$. This means there is a $r \times n$ matrix I such that $Z = IY$ represents the target. This target is a linear combination of high frequency variables that we wish to optimally estimate. Different choices of I allow for backcasting, forecasting, imputation (missing values), and signal extraction. Moreover, this can be considered at any frequency or combination of frequencies (so long as they are lower frequency than the $\{Y_t\}$ process). When we speak of signal extraction, here we refer to a signal defined as a fixed filter of the data rather than a stochastic component. For example, the X-11 filter (i.e., the final X-11 filter applied to the extreme-value adjusted series, as described in Ladiray and Quenneville (2001)) defines a target signal with I given by a matrix with rows given by the coefficients of the moving average filter, appropriately shifted.

We give further illustrations of projections – forecasting, imputation, signal extraction, etc. – following our main theorem. All these problems have in common that one seeks to estimate IY optimally from JY. Parzen (1961) provides a general formula for this problem, and we can get a particular solution in our context that is computable without knowledge of nuisance values. In particular, suppose the same two assumptions utilized for Theorem 1 hold, namely that at least d of the highest frequency values are observed and are uncorrelated with $\{W_t\}$.

Theorem 2 *Let the mixed sample X of size m be written as $X = JY$ for a high frequency vector Y that is a sample of size n from a difference stationary process with degree d differencing polynomial $\delta(B)$. Suppose that the initial values $Y_* = [1_d \; 0]PY$ are uncorrelated with $W = \Delta Y$, and are observed as the first d values of X, i.e., $[1_d \; 0]J = [1_d \; 0]P$. Then the formula for the optimal estimate of IY from JY is*

$$\widehat{IY} = I\Sigma_Y J' \Sigma_X^{-1} X = I\widetilde{\Delta}^{-1} \begin{bmatrix} X_* \\ \Sigma_W \underline{B}' (\underline{B} \Sigma_W \underline{B}')^{-1} DX \end{bmatrix}, \tag{7}$$

where $X_* = [1_d \; 0]X$ is the first d values of X. The covariance matrix of the error is

$$I\widetilde{\Delta}^{-1} \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_W - \Sigma_W \underline{B}' (\underline{B} \Sigma_W \underline{B}')^{-1} \underline{B} \Sigma_W \end{bmatrix} \widetilde{\Delta}^{\dagger} I',$$

where $\dagger$ denotes inverse transpose.

Remark 1 The first formula in (7) is the general expression from Parzen (1961), but the second formula is practicable for implementation. The estimate $\widehat{IY}$ is computable from quantities appearing in the algorithm of Section 2, and hence are readily available. Since the error covariance matrix contains no nuisance values, it is readily computed from I. Its diagonal gives the mean squared error.

Remark 2 In the case that I is an identity matrix, we obtain a vector of forecasts, backcasts, and imputations for Y from X. Also in the special case that $I = J$, we immediately recover X, as is seen from the first formula of (7).

Note that in producing estimates of IY, we do not treat the classical signal extraction problem of estimating a latent process S when $Y = S + N$ and N is noise (cf., Bell (1984)). The nature of the signals S and IY are quite different, as the latter is a linear function of the data, while the former is not. Extending classical signal extraction results to the mixed frequency case requires different formulas and assumptions, and is not treated here.

We now proceed to work through some applications. The first thing to note is that in practice I is determined first – which determines the exact length needed for Y – which in turn will determine J. We construct I and J such that the same vector Y is featured in both Z and X. That is, if I is $r \times n$ and requires a certain span of the $\{Y_t\}$ series for its definition, some of these Y_t variables may not be featured anywhere in the available data matrix X. In that case, the corresponding column of J will have all zeroes. For example, suppose we have only a single frequency and the data is $X = [Y_1, Y_2]'$, but our target is $Z = Y_0$. Then $Y = [Y_0, Y_1, Y_2]'$ and $I = [1, 0, 0]$, whereas

$$J = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This is a backcasting problem, which is only solvable when $d \leq 1$. More generally, I and J will be constructed by first building the Y vector out of the collection of all $\{Y_t\}$ variables featured in the definitions of Z and X.

The main application we have interest in occurs where we wish to compute a signal of the form $\sum_j \psi_j Y_{t-j}$ for a finite string of filter coefficients $\{\psi_j\}$. Suppose that these form a two-sided symmetric filter of total length $2q + 1$. This occurs, for example, with X-11 seasonal adjustment filters (applied as a single linear filter with full forecast extension). If we desire a total of r time points of the seasonally adjusted series, we construct an $r \times r + 2q$ dimensional matrix with row entries given by the filter coefficients:

$$\Psi = \begin{bmatrix} \psi_{-q} & \cdots & \psi_0 & \cdots & \psi_q & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & \psi_{-q} & \cdots & \psi_0 & \cdots & \psi_q \end{bmatrix}$$

Then we have $Z = \Psi Y$ where $n = r + 2q$. In terms of Theorem 2, $I = \Psi$, and we obtain our estimate and its MSE by plugging into the stated formulas.

More precisely, suppose that our target is to produce filtered values of the highest available frequency, for every such time point that occurs in our sample X. Since $X = JY$ and J is $m \times n$, this means that $r = n$, and Ψ has n rows and $n + 2q$ columns. We must extend Y to apply the Theorem 2. So let $\widetilde{Y} = [Y_{-q+1}, \cdots, Y_0, Y, Y_{n+1}, \cdots, Y_{n+q}]'$, where the middle portion is just our original $Y = [Y_1, \cdots, Y_n]'$. We must modify J accordingly, by appending columns of zeroes to its front and back:

$$\widetilde{J} = [0 \; J \; 0],$$

where the number of zero columns is q fore and aft. Then $X = \widetilde{J}\widetilde{Y}$, and we run the method on this new J and extended Y. This will change some of the formulas (n gets updated to $n + 2q$, etc.). Then the application of Theorem 2 is immediate.

This approach produces filtered values at the highest frequency. It might also be desirable to produce filtered estimates of some lower frequency. That is, suppose the target is now ΨKY, where we have used KY to denote an entire sweep of some lower frequency series. For example, if Y were monthly and we wanted quarterly values, we could write down K by selecting or aggregating (for stock and flow cases respectively) components of Y. Then with Ψ designed appropriately for that particular frequency (one must avoid using a monthly filter on quarterly data!), our target is ΨKY.

Say you have n high frequency values and the number of lower frequency values produced is r, which is less than n in general. In fact, $r = pn$ for a fraction p that is the ratio of the respectively sampling frequencies, i.e., the reciprocal of the number of time units of the highest frequency featured in one time unit of the lower frequency. For example, $p = 1/3$ for the relation of monthly to quarterly frequency. Then K is $r \times n$, with entries depending on the stock or flow cases respectively:

$$K = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \end{bmatrix} \qquad K = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & \cdots \end{bmatrix}.$$

Now the number of rows of Ψ is r, so it has $r + 2q$ columns, and hence K and Y need to be extended. In fact, we need to extend Y to $\widetilde{Y}$ by adding q/p backcast values and q/p forecast values, fore and aft. Then K is modified to $\widetilde{K}$ in the obvious fashion, such that $\widetilde{K}\widetilde{Y}$ has q low frequency values appended fore and aft to KY. Once we have determined all these things, we have $I = \Psi\widetilde{K}$ in Theorem 2, and obtain $\widetilde{J}$ as described above such that $X = \widetilde{J}\widetilde{Y}$.

In this manner, one might obtain estimates of filtered quantities at all frequencies desired. For example, we might have seasonal adjustments at quarterly (using a quarterly Ψ) and monthly (using a monthly Ψ) frequencies. At this point, we are free to splice the estimates however we desire – this can be done to mimic the mingled-frequency structure of X, if desired. For example, suppose that on January 2005 monthly data became available, the sample being purely quarterly beforehand. We can run both procedures and obtain quarterly and monthly seasonal adjustments. Then we could plot the quarterly seasonal adjustment up through 4th quarter 2004, and the monthly seasonal adjustment from January 2005 onwards. If the filters are coherent and the model is decent, the results should splice.

4 Empirical Applications

We are interested in applying these techniques to time series under production at official agencies. We first examine a U.S. Census Bureau time series for which all monthly values are available, but for which we artificially construct a quarterly segment, so that we can assess our method. Secondly, we examine a stock time series from the Bank of England, where monthly frequencies were added to the previous quarterly survey, in the middle of 2009. Thirdly, we examine a flow time series from the Bundesbank, where an originally monthly series became quarterly after 1996. All applications were computed by an implementation of the above methods in R (R Development Core Team, 2011).

4.1 Application to Series U11SFI

Our first example is a stock time series available at a monthly frequency. The series is "Food Product Finished Goods Inventory," denoted by *U11SFI*, for the dates January, 1992 through December, 2005. Now this series is not actually mixed frequency: we generated a mixed quarterly-monthly version of it, and analyzed the result. Projections from this mixed sample can be compared to truth, and the mixed frequency seasonal adjustments can also be compared to straight X-11 applied to the original monthly series. We worked with a regression-adjusted version of the series, using X-12-ARIMA initially to remove an additive outlier (no other fixed effects, such as trading day or holiday, were found to exist). Although the regression estimation approach of section 2.3 could be utilized, this would introduce extra error at the model fitting stage – to isolate the X-11 filtering aspect and facilitate comparisons with truth, we presume that all fixed effects have been satisfactorily removed.

In order to produce the mixed frequency sample, we will suppose that the last five full years of monthly observations are available, but prior to this all data is quarterly[4]. Then X consists of the last 60 monthly values, followed by all the quarterly values in order, for all but the last 5 years; see Figure 1. In the following plots, the following colors are used consistently for this example: blue represents the true full monthly time series, whereas black is what we pretend is only available to us for analysis. Projections are in green, whereas the mixed sample X-11 seasonal adjustment is in red. We can also compute the full sample X-11 seasonal adjustment based on the (blue) monthly time series, which is depicted in purple. These plots are further discussed below.

Now although a Box-Jenkins airline model (Box and Jenkins, 1976) – deemed best by X-12-ARIMA – fitted to the logged monthly data produced nonseasonal and seasonal moving average parameter MLEs of $-.08$ and $.61$ respectively, we don't know *a priori* that this model will work well with the mixed data. After trying various specifications, it became clear that an $I(2)$ model for the monthly data was really necessary (see below), and the airline model performed well (recall that we have no formal technique for testing goodness-of-fit or doing model comparisons). That is, the low frequency data seems to inherit the $I(2)$ structure of the airline model specified for the higher frequencies. The nonseasonal and seasonal moving average parameters were estimated to be $-.13$ and $.66$ respectively, reflecting only modest departures from the model fitted to the full data span. The time series residuals display little serial structure, as is evident from the ACF plot in Figure 1.

[4] We also did an analysis with only two full years of monthly data, but the quality of the results deteriorated somewhat; cf. further discussion below.

Note that because the series is a stock, the model fitting exercise to the mixed data is equivalent to fitting a monthly model to a series with a regular pattern of missing observations, i.e., prior to 2001 two out of every three observations is missing. The fact that the estimated parameters in this case are close to those obtained from the full monthly data confirms that our methodology is sound.

Using these MLEs, we then calculated the quantities in Theorem 2 to obtain the X-11 seasonal adjustment[5]. We are primarily interested in the values for the monthly segment, but values for the quarterly portion of the mixed sample are produced by our algorithm as well. In Figure 2 we have the mixed sample plotted with the projections, i.e., the forecasts, backcasts, and imputations. Also overlaid are the monthly seasonal adjustments. It is evident that the forecasts and backcasts are heavily based on the pattern present in the five years of monthly data; for the sweep of years corresponding to quarterly data, the projections effect an interpolation utilizing the inferred monthly pattern (see the right panel of Figure 3).

In order to assess whether the seasonally adjusted values for the monthly portion – i.e., January 2001 through December 2005 – it is informative to apply the same X-11 filter to the complete monthly data-set (purple). For this exercise we utilize the model that was fitted to the mixed frequency data, although one could utilize the original model instead. We do it this way to isolate the effect of the mixed sampling, since otherwise the discrepancies between the two seasonal adjustments would be due to both differences in model parameters as well as effects of the mixed sample. Figure 3 displays both X-11 seasonal adjustments together; we note there is little discrepancy in the quarterly portion, where much of the monthly pattern is inferred, and even better agreement on the monthly portion. This agreement increases as one progresses from 2001 to 2005, since the impact of the older data, where the two methods diverge, has less impact. The right panel of Figure 3 reveals that the good results are due to the high accuracy of the projection results in the quarterly span.

MSEs can also be easily produced. We first display the projection MSEs for the monthly series in the left panel of Figure 4; note that the central portion has the value zero, along with every third value in the quarterly span of the mixed sample – of course these values are known and the error is zero. Uncertainty rises predictably at the boundaries of the sample, which are marked by the dotted blue lines. Finally, the right panel of Figure 4 gives the MSEs for the X-11 seasonal adjustment estimates, the asymmetric structure following the pattern for the projections. The oscillatory and step-function characteristics of these plots are a familiar feature of finite-sample MSEs for time series projections (cf. McElroy (2008)).

We also examined the same series with a monthly span of two years, rather than five years (we do not present results here). As with the longer span, the airline model was still an adequate fit,

[5]The specification in this case was a 3×3 and 3×5 for the seasonal moving averages, and a 9-term Henderson, all for the monthly frequency. See Ladiray and Quenneville (2001) for definitions.

with parameters $-.08$ and $.59$. The projections did a remarkable job of tracking the real monthly movements in the quarterly portion, although the error was increased over the former results. Accordingly, the difference between the mixed and full monthly X-11 seasonal adjustments was still small. This encouraging result suggests that when only $24 - 13 = 11$ monthly observations are available (after differencing), the method is still viable because the quarterly data can be sensibly put to use.

4.2 Application to Sterling Series

Our second example is a stock time series available at a quarterly frequency up until July 2009, when it also became available at a monthly frequency. The original data set (which we obtained from Fida Hussain of the Bank of England) ran through May, 2011, although new values have since been added to the series. The title is "Sterling net lending to construction companies" – or Sterling for short – and is published in the "Analysis of Monetary Financial Institutions' deposits from and lending to UK residents" statistical release of the Bank of England[6]. It covers the dates December, 1986 through May, 2011, with the last 23 observations being monthly in addition to quarterly. The series was deemed seasonal according to standard diagnostics of X-12-ARIMA, including the stable F test, and this data has traditionally been seasonally adjusted by Bank of England. Note that this short span of monthly observations indicates, in light of our experience with U11SFI, that the inferred monthly patterns may be somewhat unreliable. This is just less than two years of data, and it is quite difficult to see a seasonal monthly pattern here.

A salient feature in the data is an enormous level shift occurring in January 2011, due to a reclassification. No other fixed effects were deemed significant in the quarterly data, so we proceeded only with the level shift regressor, which was estimated in an initial run of X-12-ARIMA on the quarterly span. Adopting a similar modeling approach as with the U11SFI series, our final model for logged monthly Sterling – after testing some inferior competitors – was a SARIMA (1,1,1)(0,1,1) given by

$$(1 - .96B)(1 - B^{12}) \left[X_t - .844 \, LS_t \right] = (1 - .79B)(1 - .71B^{12}) \epsilon_t.$$

The fit produced residuals with little apparent serial structure. Some other SARIMA models had "adequate" residuals as well, but maximized likelihoods that were quite a bit lower.

Then we computed all projections and applied an X-11 monthly seasonal adjustment filter (same specification as with U11SFI). The left panel of Figure 5 displays the available data together with all projections and the seasonal adjustments, while the right panel focuses on just the portion of July 2009 through May 2011. All projection results are based on regression-adjusted data, and so in the left panel the level shift has been removed from them. The right panel displays a more

[6]see http://www.bankofengland.co.uk/statistics/abl/current/index.htm.

focused picture with the projections removed. For production purposes, the level shift would be re-incorporated into the published seasonal adjustment.

The result of the analysis is difficult to assess, due to the limited scope of the 23 monthly observations, but no egregious problems with the adjustment are visible at this point. Figure 6 displays the MSE curves for both the projections and the seasonal adjustments. The left panel has zeroes at the correct times, and a pattern similar to that found with the U11SFI series. The right panel has the actual MSE curves for the seasonal adjustment.

4.3 Application to Construction Series

It is of interest to consider the opposite case to the Sterling series, where the data becomes less frequent over time. Our third example is a flow series where the frequency of observation decreases, which is the opposite of the previous example. Available as a monthly flow from 1977 onwards, the series was only published at a quarterly interval starting in 1996. The title is "Total turnover of installation and building completion work," or Construction for short. The source of the data is the German Federal Statistical Office, and former values in German marks were converted to Euros via the official conversion rate. Our sample is taken up through December 2008, so the series has 228 monthly observations followed by 52 quarterly observations.

The series presents obvious seasonality, and also has a significant trading day pattern. Since both the monthly and quarterly spans contain a fair amount of data, we can expect the method to work reliably, so long as a decent model is identified. We used a trading day model, estimated using X-12-ARIMA on the monthly portion, that resulted in estimated coefficients of $-2.63, 13.98, -4.32, 32.78, -17.37, -.79, -21.66$ for Monday through Sunday. Since the model was adequate without a transformation, we chose not to utilize a logarithmic transformation in order to avoid the disruption of flow aggregation (see section 2.5). The final model, after some analysis of competing SARIMA specifications, was again an airline model with nonseasonal and seasonal moving average parameters of .64 and .52. The fit produced residuals with little apparent serial correlation.

For such a series – which was monthly and has become quarterly – one can simply aggregate past monthly data to form a historical quarterly series, and seasonally adjust it. Such a procedure can be accomplished using standard software. A more challenging question is: can we continue to produce *monthly* seasonal adjustments on the basis of the historical monthly data together with the current quarterly data? We proceed to apply our methodology to this problem.

Visualization of mixed frequency data is harder for flows than stocks, because lower frequency data has much higher values (since it is obtained by summing higher frequency values). Figure 7 displays the Construction series: its monthly portion is in blue, and the higher quarterly portion is given with black dots. We also converted the monthly flows to quarterly flows by aggregation, and plotted in red these numbers alongside the latter quarterly flows. Projection results in Figure 7 show forecasts and backcasts, but the later quarterly data has been left out of the picture for easier viewing. Note that the projections (in green) take into account regression pre-adjustment (for trading day), and hence in the center the projections differ very slightly from the monthly data – without trading day adjustment, these portions would be identical. Also, the forecasts of the monthly portion will aggregate exactly to the known quarterly values for the years 1996 through 2008, since this is a built-in constraint. The monthly seasonal adjustment looks eminently reasonable, and has the right statistical behavior (there is some slight negative autocorrelation at lag 12, but this is fairly typical).

Figure 8 displays the MSE curves for both the projections and the seasonal adjustments, with results similar to Sterling and U11SFI. The three vertical lines mark the beginning and end of the sample, and also the transition time from monthly to quarterly flow. Unlike with the MSEs of the previous stock examples, the error for the projections during the quarterly period are all nonzero, which is because none of the quarterly flow values correspond to monthly flow values, and hence all must be estimated (imputed). Beyond the first and third vertical lines we fall into the territory of backcasts and forecasts respectively, and the uncertainty increases at a greater rate over time. The mesa-like structure of the X-11 seasonal adjustment MSE again is due to this transition from monthly to quarterly, with additional uncertainty due to the flow structure relative to stock series.

5 Conclusion

The research of this paper stems naturally from an applied problem in time series analysis with potentially wide ramifications. The issue of changing sampling frequency can be viewed from the broader perspective of the task with which many statisticians are faced, whereby it is necessary to integrate several disparate sources of information – potentially arising from different surveys, constructed under disparate assumptions – into a coherent whole. In this particular case, the problem can be tackled fairly easily by reducing all observations to a linear function of the highest available frequency. Although the application of projection theory is classical at this stage, the practical formulas for nonstationary processes are subtle and require some care.

This paper's main contribution is a presentation of exact formulas for likelihoods and projections in the context of such mixed frequency data, which fills an important theoretical gap in the literature. Some authors may prefer to utilize the formulas herein, rather than encode a state space approach to the problem. We believe there are some advantages to having the exact matrix formulas for conditional expectations, and the corresponding error covariance matrices. In particular, the case of mixed flow data cannot be handled with SS methodology, and recourse to our method is required. Practical issues, such as dealing with logarithmic transformations of flow variables, as well as regressors, are addressed as well.

The method applied to produce X-11 seasonal adjustments performs quite reasonably once a decent model is identified for the mixed sample. We emphasize that this aspect is crucial, since our own earlier efforts based on faulty models produced grotesque results. In particular, we have found that $I(1)$ models for the highest frequency generate projections that attempt to maintain a steady level while at the same time remaining faithful to constraints from the lower frequencies. For example, with U11SFI the projections from an $I(1)$ model extrapolate the monthly seasonal pattern backwards in time, at the same level, but with ever-increasing amplitude of the seasonal factors to ensure that the projections coincide with every third monthly historical value (i.e., the quarterly values). In order to ensure that high frequency structure can be appropriately imposed on low frequency observations with changing level, an $I(2)$ model is necessary. Once the differencing polynomial $\delta(B)$ is correctly ascertained, the other aspects of the model (such as SARMA specification and parameter values) had a much narrower impact. Of course, this sort of discussion is well-known for single frequency time series, but our point here is that the repercussions of mis-specifying the unit root have a much greater impact on projections in the case of mingled frequency data.

In actual practice, earlier seasonal adjustments from a single frequency may be historically available to the analyst, and any viable model should produce new seasonal adjustments that are largely in agreement with the past. Our methodology (and examples) suggests that an available model for the highest frequency data (in the past) may well serve as a good model for $\{Y_t\}$ when fitting to the mingled data as well; but this strategy is only available when the data is becoming less frequent over time, as with the Construction series. The practitioner naturally wishes the new seasonal adjustments, at whatever desired frequency that is considered, to be in agreement with past adjustments. This aesthetic can be used to guide model selection, i.e., models that produce seasonal adjustments radically disparate from previous single frequency seasonal adjustments should be questioned and revised. Adopting this principle has allowed us – in the examples of section 4 – to obtain superior models to those that we had initially conceived.

In addition, assessing the model fit through acf plots of the time series residuals seems to be a good screening technique as well. An extensive exposition of model selection in the context of multiple frequency data would be welcome, but is beyond the scope of this article. We also reiterate that it is important to have a suitably long stretch of highest frequency data, since the inferred patterns that are imputed to the lower frequency stretches principally arise from the observed high frequencies. Our particular implementation of the likelihoods and projections were written in R (R Development Core Team, 2011), and all programs and scripts are available from the authors upon request. (This methodology has not yet been incorporated into X-12-ARIMA.)

Acknowledgements: The authors are grateful to Fida Hussain (Bank of England) for suggesting the topic and for providing the Sterling series. We also thank Jens Mehrhoff (Bundesbank) for providing the Construction series, along with discussions about alternative approaches to the problem; we thank David Findley for criticisms.

Appendix

We begin with a lemma that is used in the proof of Proposition 1 and Theorem 1.

Lemma 1 *For any permutation matrix P, the matrix $\widetilde{\Delta}(P)$ defined via (3) is invertible.*

Proof of Lemma 1. Suppose that an n-vector v is annihilated by $\widetilde{\Delta}(P)$. Since $\Delta v = 0$, we can write $v = \{\sum_{i=1}^{d} \beta_i z_i(t_j)\}_{j=1}^{n}$ for basis functions $z_i(t)$ and times t_j. (The action of Δ on each basis function annihilates it.) But $0 = [1_d \, 0] P v$ implies that there are d times $t_{k_1}, \cdots, t_{k_d}$ such that $0 = \sum_{i=1}^{d} \beta_i z_i(t_{k_j})$ for $j = 1, 2, \cdots, d$, which in turn implies that $\beta_1 = \beta_2 = \cdots = \beta_d = 0$ by the linear independence of the basis functions. Hence $v = 0$. $\square$

Proof of Proposition 1. We begin with the representation result of Bell (1984), which can be written as

$$Y_t = A_t' \underline{Y}_0 + \sum_{j=d+1}^{t} \xi_{t-j} W_j, \tag{A.1}$$

where the kth component of A_t is $\xi_{t-k} - \xi_{t-k-1}\delta_1 - \cdots - \xi_{t-d}\delta_{d-k}$ for $k = 1, 2, \cdots, d$ and the $\{\xi_j\}$ are the coefficients of $\delta^{-1}(B)$. Here $\underline{Y}_0 = [Y_1, Y_2, \cdots, Y_d]'$, which is a "canonical" choice of initial values. When P is the identity matrix, we write equation (3) as $\widetilde{\Delta}(1_\ell)$ for given dimension ℓ. Then by Lemma 1

$$Y = \widetilde{\Delta}^{-1}(1_\ell) \begin{bmatrix} \underline{Y}_0 \\ W \end{bmatrix} = \widetilde{\Delta}^{-1}(P) \begin{bmatrix} Y_* \\ W \end{bmatrix},$$

where $W = [W_{d+1}, \cdots, W_\ell]'$. This shows that

$$\underline{Y}_0 = [1_d \; 0] \widetilde{\Delta}^{-1}(P) [1_d \; 0] Y_* + [1_d \; 0] \widetilde{\Delta}^{-1}(P) [0 \; 1_{\ell-d}] W.$$

Plugging this into (A.1) with $\ell = t$ then yields the representation discussed in the statement of the proposition, after collecting terms. Moreover, the above argument shows that $Y_* = C\underline{Y}_0 + BW$ with C a $d \times d$ matrix and B a matrix with d rows; cf. the representation in equation (4.1) of Bell and Hillmer (1991). This relates Y_* to $\underline{Y}_0$, but we can also relate Y_* to any other initial values in this way. $\square$

Proof of Proposition 2. We begin by calculating $X' \Sigma_X^{-1} X$, and then consider the direct and converse statements. A general block decomposition of $RJ\widetilde{\Delta}^{-1}(P)$ for any invertible matrix R is

$$RJ\widetilde{\Delta}^{-1}(P) = \begin{bmatrix} \overline{A} & \overline{B} \\ \underline{A} & \underline{B} \end{bmatrix},$$

partitioned into d and $m-d$ rows, and d and $n-d$ columns. Then $RX = RJY$ is given by

$$RX = \begin{bmatrix} \overline{A} & \overline{B} \\ \underline{A} & \underline{B} \end{bmatrix} \begin{bmatrix} Y_* \\ W \end{bmatrix}.$$

Then, using the assumption that Y_* and W are uncorrelated, along with the Schur decomposition (Axelsson, 1996), we obtain:

$$\begin{aligned}
\Sigma_X &= \begin{bmatrix} \overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}' & \overline{A}\Sigma_*\underline{A}' + \overline{B}\Sigma_W\underline{B}' \\ \underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}' & \underline{A}\Sigma_*\underline{A}' + \underline{B}\Sigma_W\underline{B}' \end{bmatrix} \\
S &= \underline{A}\Sigma_*\underline{A}' + \underline{B}\Sigma_W\underline{B}' - \left(\underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}'\right)\left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}\left(\overline{A}\Sigma_*\underline{A}' + \overline{B}\Sigma_W\underline{B}'\right) \\
X'\Sigma_X^{-1}X &= [\overline{A}Y_* + \overline{B}W]'\left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}[\overline{A}Y_* + \overline{B}W] \\
&\quad + \left[\underline{A}Y_* + \underline{B}W - \left(\underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}'\right)\left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}\left(\overline{A}Y_* + \overline{B}W\right)\right]' \\
&\quad \cdot S^{-1}\left[\underline{A}Y_* + \underline{B}W - \left(\underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}'\right)\left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}\left(\overline{A}Y_* + \overline{B}W\right)\right]
\end{aligned}$$

Note that R does not appear in the final quadratic form, since it is invertible. If we first assume that a block-diagonalizing R exists, then $\overline{B} = 0$ and $\underline{A} = 0$, and the quadratic form reduces to the sum of $[\overline{A}Y_*]'\left(\overline{A}\Sigma_*\overline{A}'\right)^{-1}[\overline{A}Y_*]$ and $[\underline{B}W]'(\underline{B}\Sigma_W\underline{B}')^{-1}[\underline{B}W]$, so factorization is immediate. Conversely, suppose that the likelihood factorizes. Then if we apply ∇_W, the gradient with respect to W, to the quadratic form, the resulting expression will depend upon Σ_W unless $\overline{B} = 0$. Similarly, if we apply ∇_{Y_*} to $X'\Sigma_X^{-1}X$ we find that $\underline{A} = 0$ must hold (otherwise S will depend on Σ_*). This shows that $RJ\widetilde{\Delta}^{-1}(P)$ is block diagonal as desired. □

Proof of Theorem 1. The invertibility of $\widetilde{\Delta}(P)$ is shown in Lemma 1. Also $[1_d\ 0]J\widetilde{\Delta}^{-1}(P) = [1_d\ 0]$ since $[1_d\ 0]J = [1_d; 0]\widetilde{\Delta}(P)$. With the definition of $\underline{A}$ and $\underline{B}$ in (4), the block diagonal form of $RJ\widetilde{\Delta}^{-1}(P)$ follows at once from the definition of R, which is clearly lower triangular with unit diagonal (and hence invertible). Also

$$X = JY = J\widetilde{\Delta}^{-1}(P)\begin{bmatrix} Y_* \\ W \end{bmatrix} = R^{-1}\begin{bmatrix} 1_d & 0 \\ 0 & \underline{B} \end{bmatrix}\begin{bmatrix} Y_* \\ W \end{bmatrix} = R^{-1}\begin{bmatrix} Y_* \\ \underline{B}W \end{bmatrix}, \tag{A.2}$$

from which follows (5) upon left multiplication of (A.2) by R. In addition it follows from (A.2) that particular,

$$\begin{aligned}
\Sigma_X &= R^{-1}\begin{bmatrix} \Sigma_* & 0 \\ 0 & \underline{B}\Sigma_W\underline{B}' \end{bmatrix} R^{\dagger} \\
\Sigma_X^{-1} &= R'\begin{bmatrix} \Sigma_*^{-1} & 0 \\ 0 & (\underline{B}\Sigma_W\underline{B}')^{-1} \end{bmatrix} R,
\end{aligned}$$

where all the stated inverses indeed exist. The dagger notation stands for inverse transpose. Now utilizing this in (1) together with (A.2) and the fact that R has unit determinant, we obtain (6). $\square$

Proof of Theorem 2. The first line of $\widehat{IY}$ follows directly from Parzen (1961). Note that

$$\Sigma_Y = \widetilde{\Delta}^{-1}(P) \begin{bmatrix} \Sigma_* & 0 \\ 0 & \Sigma_W \end{bmatrix} \widetilde{\Delta}^{\dagger}(P),$$

so that by (A.2) and the other calculations in the proof of Theorem 1 we obtain

$$\widehat{IY} = I\widetilde{\Delta}^{-1}(P) \begin{bmatrix} \Sigma_* & 0 \\ 0 & \Sigma_W \end{bmatrix} \begin{bmatrix} 1_d & 0 \\ 0 & \underline{B}' \end{bmatrix} \begin{bmatrix} \Sigma_*^{-1} & 0 \\ 0 & (\underline{B}\Sigma_W \underline{B}')^{-1} \end{bmatrix} RX,$$

which simplifies to the stated formula. This depends on no nuisance values. The error process is

$$\widehat{IY} - IY = -I \left(1_n - \Sigma_Y J' \Sigma_X^{-1} J \right) Y,$$

which has covariance matrix

$$\begin{aligned} & I \left(1_n - \Sigma_Y J' \Sigma_X^{-1} J \right) \Sigma_Y \left(1_n - J' \Sigma_X^{-1} J \Sigma_Y \right) \\ & = I\widetilde{\Delta}^{-1}(P) \begin{bmatrix} \Sigma_* & 0 \\ 0 & \Sigma_W \end{bmatrix} \widetilde{\Delta}^{\dagger}(P) I' \\ & - I\widetilde{\Delta}^{-1}(P) \begin{bmatrix} \Sigma_* & 0 \\ 0 & \Sigma_W \underline{B}' (\underline{B}\Sigma_W \underline{B}')^{-1} \underline{B}\Sigma_W \end{bmatrix} \widetilde{\Delta}^{\dagger}(P) I', \end{aligned}$$

which simplifies to the stated formula. $\square$

References

[1] Ansley, C. and Kohn, R. (1985) Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *Ann. Statist.* **13**, 1286–1316.

[2] Axelsson, O. (1996) *Iterative Solution Methods.* Cambridge: Cambridge University Press.

[3] Bell, W. (1984) Signal extraction for nonstationary time series. *The Annals of Statistics* **12**, 646–664.

[4] Bell, W. and Hillmer, S. (1991) Initializing the Kalman filter for nonstationary time series models. *Journal of Time Series Analysis* **12**, 283–300.

[5] Box, G. and Jenkins, G. (1976) *Time Series Analysis,* San Francisco, California: Holden-Day.

[6] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods.* New York: Springer.

[7] Chen, B. and Zadrozny, P. (1998) An extended Yule-Walker method for estimating a vector autoregressive model with mixed-frequency data. *Advances in Econometrics* **13**, 47–73.

[8] Cochran, W. (1934) The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Proc. Cambridge Philos. Soc.* **30**, 178–191.

[9] Dagum, E. (1980) *The X-11-ARIMA Seasonal Adjustment Method.* Ottawa: Statistics Canada.

[10] Doornik, J. (1998) *Object-oriented Matrix Programming using Ox 2.0.* London: Timberlake Consultants Press.

[11] Durbin, J. and Koopman, S. (2001) *Time Series Analysis by State Space Methods*. Oxford University Press.

[12] Durbin, J. and Quenneville, B. (1997) Benchmarking by state space methods. *International Statistical Review* **65**, 23–48.

[13] Golub, G. and Van Loan, C. (1996) *Matrix Computations.* Baltimore and London: The Johns Hopkins University Press.

[14] Koopman, S., Shephard, N., and Doornik, J. (1999) Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* **2**, 113–166.

[15] Ladiray, D. and Quenneville, B. (2001) *Seasonal adjustment with the X-11 method.* New York: Springer-Verlag, Vol. 158.

[16] Maravall, A. and Caparello, G. (2004) Program TSW: Revised Reference Manual. Working paper 2004, Research Department, Bank of Spain. http://www.bde.es.

[17] McElroy, T. (2008) Matrix Formulas for Nonstationary ARIMA Signal Extraction. *Econometric Theory* **24**, 1-22.

[18] Parzen, E. (1961) An approach to time series analysis. *The Annals of Mathematical Statistics* **32**, 951–989.

[19] Pasteels, J-M. (2012) Seasonal adjustment of data derived from labour force surveys: some specific issues. *2012 Workshop on methodological issues in seasonal adjustment*, Luxembourg. `http://www.cros-portal.eu/page/2012-workshop-methodological-issues-seasonal-adjustment`

[20] R Development Core Team (2011) R Foundation for Statistical Computing.

[21] Scheffe, H. (1959) *The Analysis of Variance*. New York: John Wiley and Sons.

[22] U.S. Census Bureau (2011) *X-12-ARIMA Reference Manual.* `http://www.census.gov/ts/x12a/v03/x12adocV03.pdf`

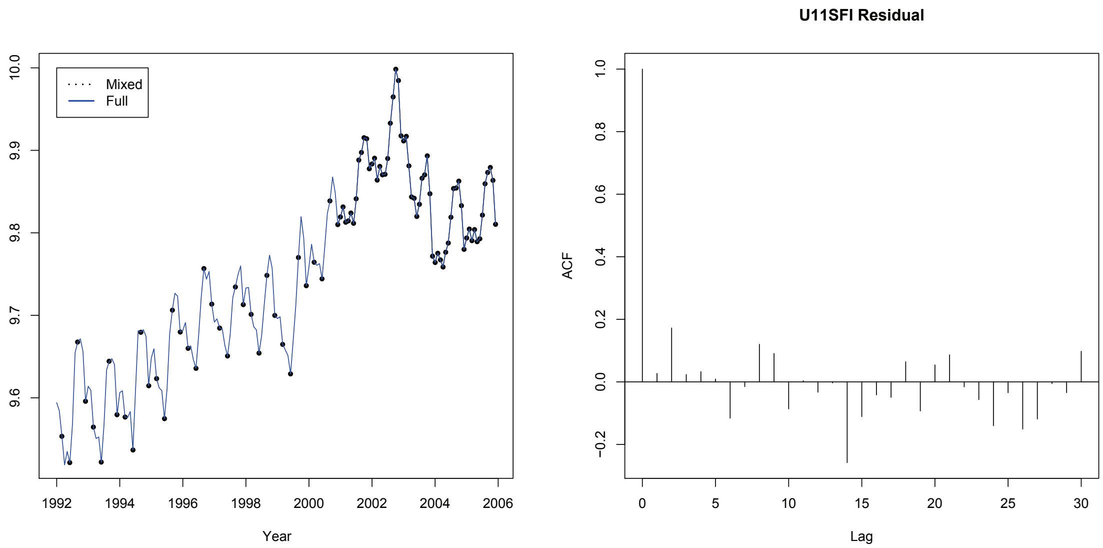

Figure 1: The left panel displays the mixed frequency sample along with the original monthly sample of *U11SFI*, whereas the right panel displays the ACF of the time series residuals from the fitted mixed frequency airline model. The spacing of the black dots in the left panel changes in 2001, which reflects the transition from quarterly to monthly frequencies.

[23] Zadrozny, P. (1990) Estimating a multivariate ARMA model with mixed-frequency data: an application to forecasting U.S. GNP at monthly intervals. Research paper, Research Department, Federal Reserve Bank of Atlanta.

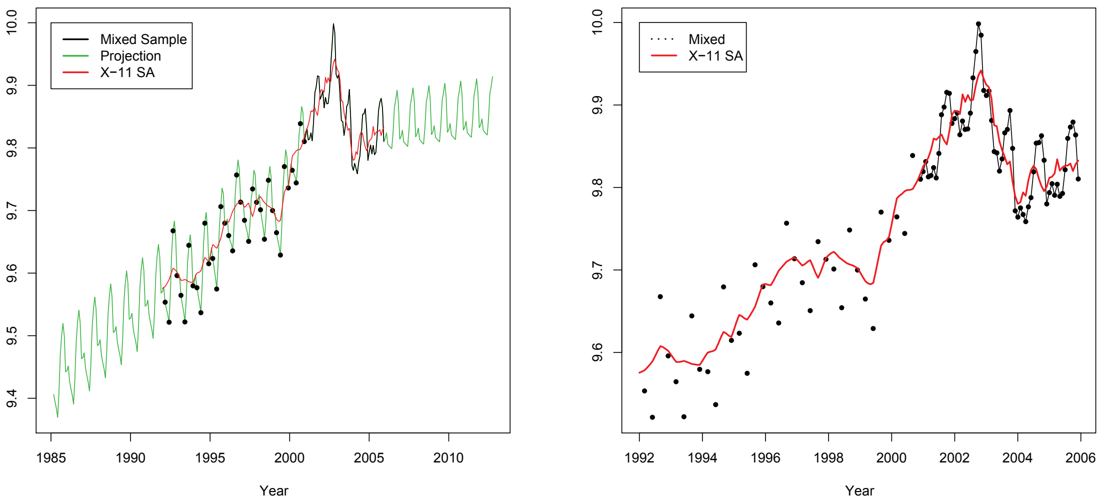

Figure 2: The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for *U11SFI*. The right panel displays the same information without the projections. The spacing of the black dots in these panels changes in 2001, which reflects the transition from quarterly to monthly frequencies.

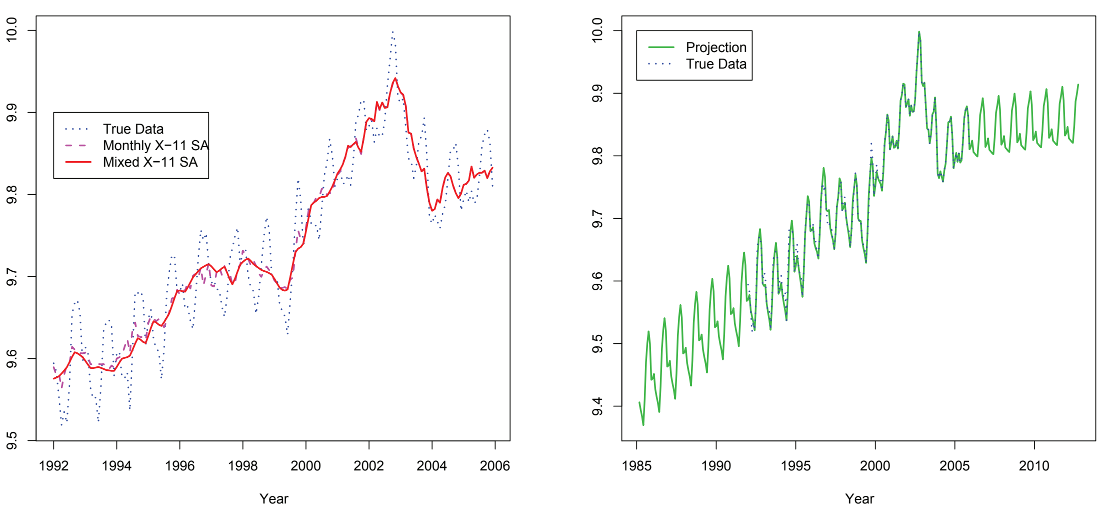

Figure 3: The left panel displays the seasonal adjustments for *U11SFI* utilizing both the mixed sample and the entire monthly data span. The right panel compares the projections directly to the actual partially unobserved true monthly data.

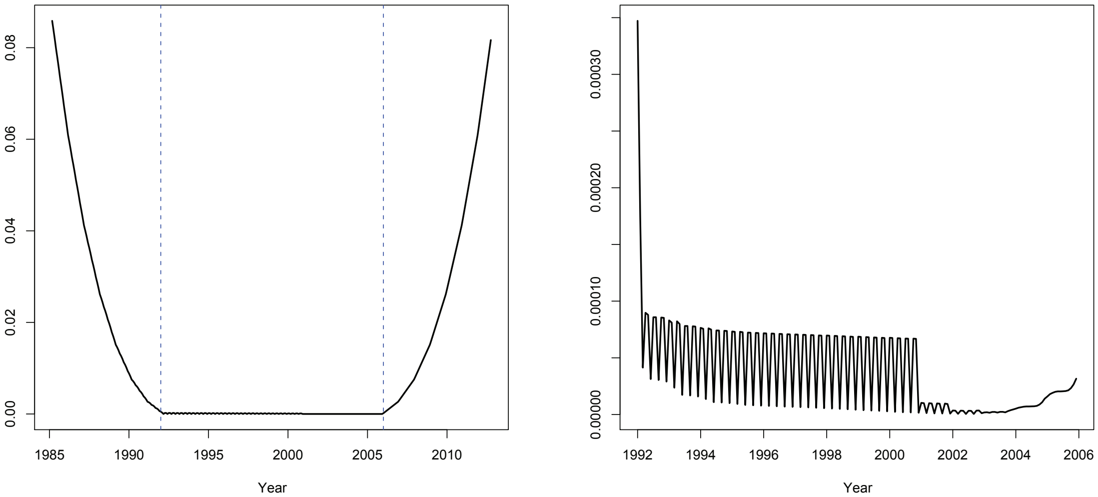

Figure 4: The left panel displays the MSEs for the full projections for *U11SFI*, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.

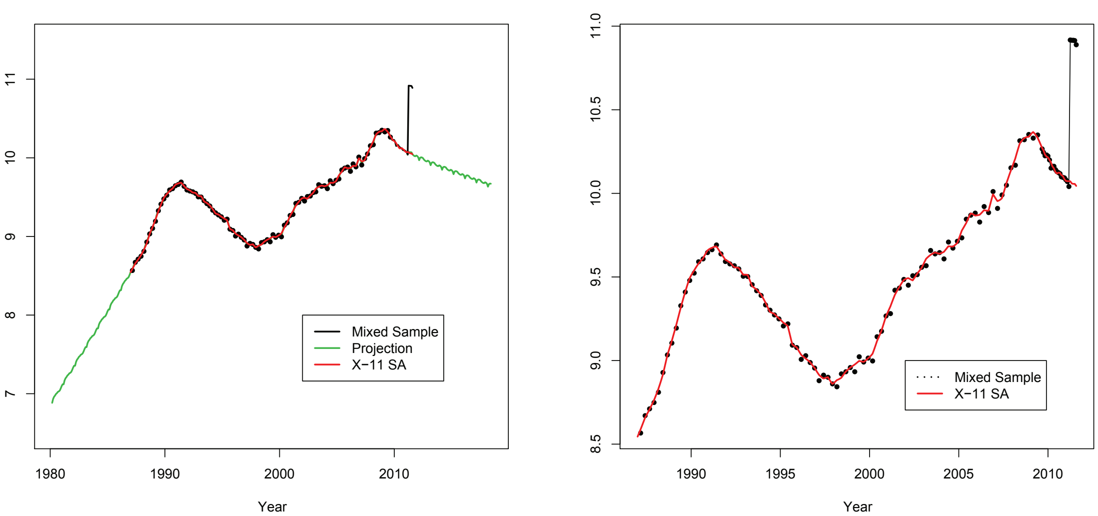

Figure 5: The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for the *Sterling* series. The right panel displays the same information without the projections. The spacing of the black dots in the left panel changes in 2009, which reflects the transition from quarterly to monthly frequencies.

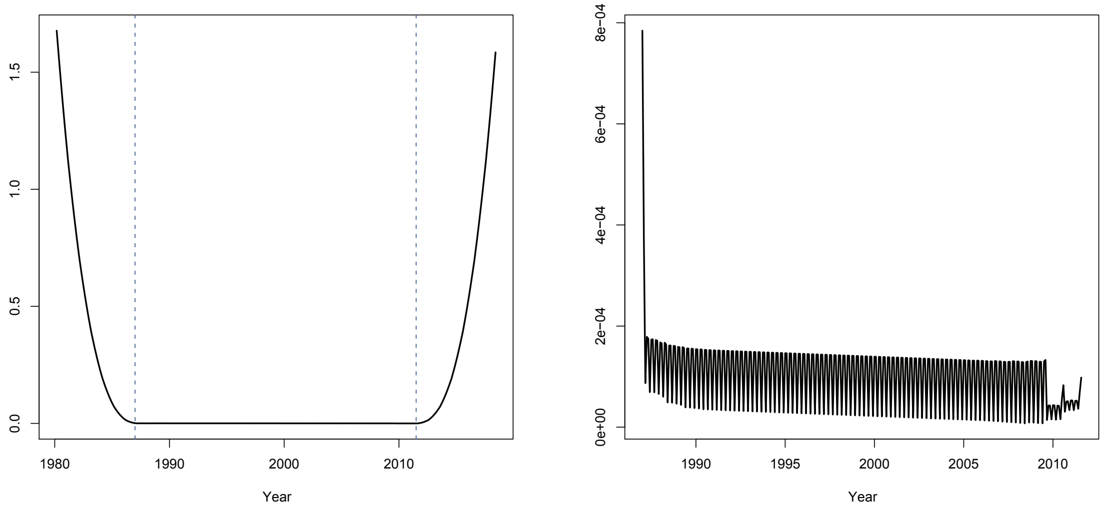

Figure 6: The left panel displays the MSEs for the full projections for the *Sterling* series, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.

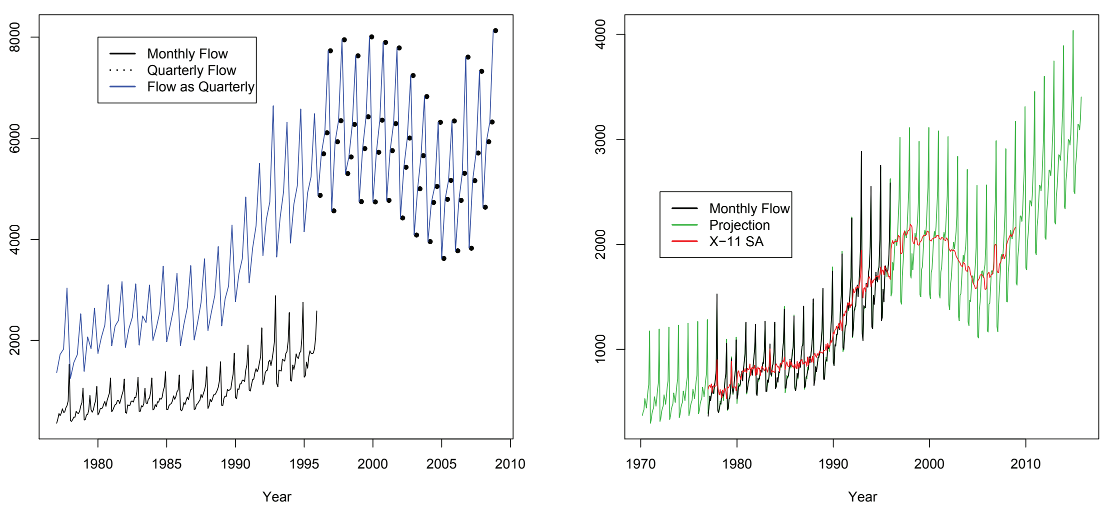

Figure 7: The left panel displays the mixed frequency sample for the *Construction* series, together with the monthly flow aggregated to a quarterly frequency (in red). The right panel displays the monthly data, full projections, and monthly seasonally adjusted data for the *Construction* series.

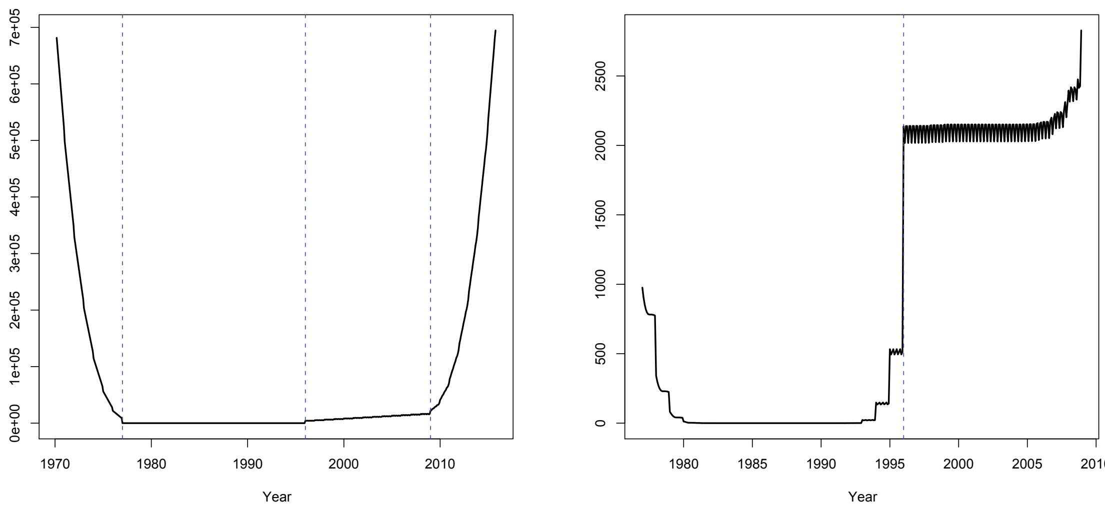

Figure 8: The left panel displays the MSEs for the full projections for the *Construction* series, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.