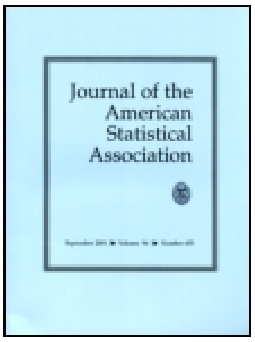

Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

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

Tucker McElroy & Brian Monsell

To cite this article: Tucker McElroy & Brian Monsell (2015) Model Estimation, Prediction, and Signal Extraction for Nonstationary Stock and Flow Time Series Observed at Mixed Frequencies, Journal of the American Statistical Association, 110:511, 1284-1303, DOI: 10.1080/01621459.2014.978452

To link to this article: http://dx.doi.org/10.1080/01621459.2014.978452

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

Tucker McElroy and Brian Monsell

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, so that time series imputation methods are ineffective. 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 evaluate the methodology through simulations, and illustrate the techniques on some economic time series.

KEY WORDS: Imputation; Missing data; Sampling frequency; Seasonal adjustment

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. (This is an approximate view of the procedure—exact details can be found in Ladiray and Quenneville (2001).) 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 article 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 challenges associated with this approach. First, state–space methods require a substantial programming effort, although one may use off-the-shelf routines such as those offered through 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

Tucker McElroy (E-mail: *tucker.s.mcelroy@census.gov*) and Brian Monsell (E-mail: *brian.c.monsell@census.gov*), Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100.

This article is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

In the Public Domain
Journal of the American Statistical Association
September 2015, Vol. 110, No. 511, Theory and Methods
DOI: 10.1080/01621459.2014.978452

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, Shephard, and Doornik 1999; Durbin and Koopman 2001). Third, numerical devices are typically used 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. Fourth, 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; a different state-space approach is needed instead, which requires some care to implement.

However, the advantages of state-space methods 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 article 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's (1991) optimal initialization), or as a straightforward matrix-based method for computing projections when recourse to the Kalman filter is infeasible (e.g., when a long memory model is being used) 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. Section 4 of the article explores model fitting and seasonal adjustment through a simulation study, while in Section 5 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, as well as on an industrial production series, 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 6 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. These data have the following characteristics by assumption:

- The data occur with multiple sampling frequencies.
- The data at each sampling frequency are difference-stationary (with known differencing polynomial).
- The observed data are 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.

For a stock, the given data may be viewed as belonging to any one of a number of sampling frequencies—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 subsampled monthly flow series. The first quarterly value equals the sum of the January, February, and March values; hence knowing the quarterly value does not allow us to deduce any of the three monthly values, and therefore quarterly flow series cannot be rewritten as a monthly flow series (with missing values). For a stock series, in contrast, the first quarterly value equals the March value, so that the quarterly stock series can be reexpressed as a monthly stock series (with missing values).

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 are 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 $X = (X_1, X_2, \dots, X_n)'$. 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 regular subsampling (i.e., sampling by skipping a certain number of observations). But for flows, the lower frequency values are obtained by aggregation of the highest frequency. Hence, the observed data take the form $J X$, where J is a selection matrix that is described below.

Letting $\{X_t\}$ 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 $W_t = \delta(B)X_t$ is mean zero and covariance stationary, where B is the backshift operator and $\delta(z) = \sum_{j=0}^{d} \delta_j z^j$ is a degree d unit-root polynomial. Let the autocovariance function (ACF) of $\{W_t\}$ be denoted by $\{\gamma_h\}$. We can conveniently represent $\{X_t\}$ 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 $\{X_t\}$ will also include fixed effects, such as calendar and trading day effects. We proceed by specifying a model for $\{X_t\}$, 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

In the basic framework, suppose that we can write the observed data as $Y = J\,X$ with J typically having entries of ones and zeroes, where $\{X_t\}$ is described above. The high-frequency vector X has length n, but the vector of observed data has length m, so that J is $m \times n$. Assuming for now that X has mean zero—the case of nonzero mean is treated later—the log Gaussian likelihood multiplied by -2 is given by

$$Y' \left(J\,\Sigma_X\,J'\right)^{-1} Y + \log |J\,\Sigma_X\,J'| \tag{1}$$

up to irrelevant constants. Note that this objective function is derived from the Gaussian joint probability density for the undifferenced data $Y = J\,X$. 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 $J\,X$ of the same frequency. There is no reason to suppose this contiguity in $J\,X$ 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:

$$Y'\Sigma_Y^{-1}Y = Q_1(X_*, \Sigma_*) + Q_2(W, \Sigma_W), \tag{2}$$

where X_* consists of some d components of X that are admissible (see below) initial values (in analogy with the treatment of Bell 1984 and McElroy 2008), and Σ_* is the covariance matrix of X_*. Also $W = [W_{d+1}, \ldots, W_n]'$ 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 X_* need not be the first d (or last d) values of X, as is typical in the single frequency literature; see Bell and Hillmer (1991, sec. 4) on different initial values. The term "admissible" means that the vector X can be reconstructed from X_* and W; not all choices of X_* have this property (see Proposition 1).

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 Y, X_*, and W as fixed nonrandom 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 X_* or Σ_*. Again, this is purely a property of linear algebra. From this definition, it follows that the gradient with respect to W of $Q_1(X_*, \Sigma_*)$ is zero, and the gradient with respect to X_* of $Q_2(W, \Sigma_W)$ is zero as well. (While there is a connection to the factorization problem discussed in Cochran's Theorem (see Cochran 1934; Scheffé 1959), that result does not apply here; letting $Z = [X_*', W']'$, Theorem 1 of Scheffé (1959) is concerned with decomposing $Z'Z$ into $Q_1(X_*, \Sigma_*)$ and $Q_2(W, \Sigma_W)$, the right-hand side of (2). However, the left-hand side of (2) is the more complicated form $Y'\Sigma_Y^{-1}Y$, rather than just $Z'Z$.)

We first establish a preliminary result about initial values of the process. Since X is the highest frequency (unobserved) data vector, it is differenced to stationarity by an $n - d \times n$-dimensional matrix Δ:

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

Then $W = \Delta X$ is mean zero and has a Toeplitz covariance matrix Σ_W. Suppose that we consider any d values of X (denoted X_*) as initial values, and seek to know whether they are admissible. These values are equal to the first d components of some n-dimensional permutation matrix P applied to X. 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 of dimension dictated by context.) Then applying $\widetilde{\Delta}(P)$ to X yields a vector with the first d components given by X_*, followed by W. Therefore, X is equal to the inverse of $\widetilde{\Delta}(P)$, when it exists, applied to $[X_*', W']'$; hence X_* is admissible if and only if $\widetilde{\Delta}(P)$ is invertible—this can be considered as the definition of admissibility. We now state our first result about initial values.

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

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

for d admissible initial values $X_* = [X_{t_1}, X_{t_2}, \ldots, X_{t_d}]'$. Here, $W_t = \delta(B)X_t$ 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 X_* and any other d admissible initial values.

For example, consider $\delta(B) = 1 - B^2$ and $n = 3$, with $X_*' = [X_1, X_3]$; then

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \widetilde{\Delta}(P) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix},$$

and the latter matrix has null vector [0, 1, 0]. So this is inadmissible (there is no way to obtain X_2 from X_1, X_3, and W_3), but in contrast $X'_* = [X_1, X_2]$ and $X_* = [X_2, X_3]$ are both admissible.

So we are free to choose admissible 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 Y, because then the likelihood is guaranteed to factor. First, we establish necessary and sufficient conditions for this factorization, given a particular choice of admissible 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 admissible initial values, wherever they are chosen to lie within an actual time series, are typically assumed to be uncorrelated with the differenced process (see 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 admissible initial values $X_* = [1_d\ 0]PX$ are uncorrelated with the differenced process $\{W_t\}$. Then the Gaussian likelihood factorizes (i.e., Equation (2) holds) 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 $\widetilde{\Delta}(P)$ defined as in (3).

Since $J\widetilde{\Delta}^{-1}(P)$ can be readily computed for each choice of admissible initial values, we can verify the condition of Proposition 2 at once by doing row reduction (Golub and Van Loan 1996, p. 102). The choice of admissible initial values may seem to be a conundrum (Proposition 2 makes its assertion for any single given set of admissible initial values, but does not require the orthogonality condition simultaneously for all such choices, which is typically impossible). However, whenever d initial values are contiguous (i.e., the elements of X_* have indices $t_1, \ldots, t_d \in \{t_* + 1, \ldots, t_* + d\}$ for some integer t_*), then they are admissible, as the following result describes.

Lemma 1. Initial values are contiguous if and only if there exists a $d \times d$ permutation matrix Π such that $[1_d\ 0]P = [0\ \Pi\ 0]$, where the first t_* columns of this matrix are zero. Furthermore, contiguous initial values are admissible.

Note that admissible values need not be contiguous, but contiguity is often a convenient sufficient condition for admissibility; in many practical situations, at least d contiguous values of the high-frequency data X are observed—then any d-length block of such can be used as admissible initial values.

2.3 Maximum Likelihood Estimation of Mixed Frequency Time Series

We now assume that d contiguous values of X are observed (i.e., appear in Y), and choose these to be the initial values—Lemma 1 guarantees that they are admissible. We also suppose—row-permuting Y if necessary—that these d initial values actually occur as the first d values of Y. This assumption enforces a structure on J, which can be stated mathematically as $[1_d\ 0]J = [1_d\ 0]P$.

Theorem 1. Let the mixed sample Y of size m be written as $Y = JX$ for a high-frequency vector X that is a sample of size n from a difference stationary process with degree d differencing polynomial $\delta(B)$. Suppose that the initial values $X_* = [1_d\ 0]PX$ are contiguous and uncorrelated with $W = \Delta X$, and are observed as the first d values of Y, that is, $[1_d\ 0]J = [1_d\ 0]P$. Then the Gaussian likelihood factorizes (see 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 $J\widetilde{\Delta}^{-1}(P)$, that is,

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

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

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

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

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

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 (but $\log|\Sigma_*|$ is handled differently). It is evident from Theorem 1 that this strategy is unnecessary, because no model parameters are featured in the initial value terms and 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) demonstrated 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 an 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 an SS framework with finite-length state vector.

For example, although long memory models can be embedded in an SS framework with infinite-length state vector (Chan and Palma 1998), for applications a truncation is needed, which can substantially degrade the model's description of long-range dependence when persistence is high; see McElroy and Holan (2012) for computational challenges associated with long-range-dependent processes. Another example is furnished by the exponential models of Bloomfield (1973), which are convenient for Bayesian modeling (Holan, McElroy, and Chakraborty 2009). These models have an infinite moving average representation, and cannot be embedded in a finite SS framework. Both these examples—as well as any process with a well-defined autocovariance function computable from model parameters—serve as applications of Theorem 1: to evaluate the Gaussian likelihood of a long memory, exponential, or general time series model, it is required to compute γ_h and form Σ_W. McElroy and Holan (2012) provided such algorithms for the case of long memory and exponential models, whereas conventional models (such as ARMA) are handled in classic texts (Brockwell and Davis 1991). For auto regressive moving average (ARMA) models, one may indeed embed them in an SS framework and use the Kalman filter to evaluate the Gaussian likelihood—this is generally speaking more computationally advantageous over a direct application of Theorem 1, so long as the initialization associated with nonstationary effects is correctly handled.

In summary, our expression for the likelihood (6)—which is contingent on having at least d of the highest frequency values contiguously observed—is the same as that arising from the Kalman filter with an optimal initialization. Whether it agrees with nonoptimal 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, for certain cases (e.g., long-range dependence) the use of SS methods involves approximations. 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 DY. To summarize, our procedure is:

1. Identify any d contiguous observed high-frequency values of $\{X_t\}$ and rewrite Y so that these occur in the first d time points.
2. Write down J such that $Y = J\,X$, with X_* the d initial values (and the first d values of Y).
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 DY 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 to 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, $Y = J\,X$ 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 follow 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 MLE, 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 Akaike information criterion (AIC) values based on the likelihoods of competing models, 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 $\mathcal{X} = \mathbb{X}\beta + X$, with parameter vector β and design matrix $\mathbb{X}$. Then our observed data follow

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

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

$$\begin{aligned}&(Y - J\,\mathbb{X}\beta)'\Sigma_Y^{-1}(Y - J\,\mathbb{X}\beta) + \log|\Sigma_Y^{-1}|\\ &= (X_* - [1_d\;0]P\mathbb{X}\beta)'\Sigma_*^{-1}(X_* - [1_d\;0]P\mathbb{X}\beta) - \log|\Sigma_*|\\ &\quad+\left(DY - \underline{B}\Delta\mathbb{X}\beta\right)'\left(\underline{B}\Sigma_W\underline{B}'\right)^{-1}\left(DY - \underline{B}\Delta\mathbb{X}\beta\right)\\ &\quad-\log|\underline{B}\Sigma_W\underline{B}'|.\end{aligned}$$

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

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

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 DY replaced by $DY - \underline{B}\Delta\mathbb{X}\beta$.

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. (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 procedure introduces additional parameters, 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 as

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

where the MLEs for parameters are used—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. Two conventional approaches, used for single frequency time series, were described by Brockwell and Davis (1991): sums of squared residual autocorrelations can be compared to the χ^2 distribution (i.e., either a Box–Pierce or Ljung–Box statistic). One can also use nonparametric methods, such as the "runs" test and the "difference-sign" test. In our simulation and empirical work, we use both the Ljung–Box statistic and the difference-sign test.

2.5 Treatment of Logged Flows and Approximation Error

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) subsampling of a higher frequency flow. Put another way, there is more than one nonzero entry in each row of J. Hence, an 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 $Y = J\,X$ no longer holds as an exact relation. We illustrate this first with monthly (x_1, x_2, and x_3) and quarterly (y_1) data. In the original scale, the flow property states that $y_1 = x_1 + x_2 + x_3$, when y_1 corresponds to the quarter comprising the months described by x_1, x_2, x_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 $Y_1 \neq X_1 + X_2 + X_3$, which is a nuisance. Note that this problem does not arise for stocks (where $y_1 = x_3$ maps to $Y_1 = X_3$ without hindrance).

By Jensen's inequality, we always have $Y_1 \geq \log 3 + (X_1 + X_2 + X_3)/3$. More precisely, we can write

$$\log y_1 = \frac{\log x_1 + \log x_2 + \log x_3}{3} + \log\left[\left(\frac{x_1^2}{x_2x_3}\right)^{1/3} + \left(\frac{x_2^2}{x_1x_3}\right)^{1/3} + \left(\frac{x_3^2}{x_1x_2}\right)^{1/3}\right]. \quad (7)$$

This latter term is bounded below by log(3), and in practice is close to the lower bound when x_1, x_2, x_3 have values reasonably close to one another. This analysis indicates that we might write $Y \approx J\,X + \mu\iota$, 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 may be small in certain cases, such that one proceeds with the algorithm as if it were exact; we explore this issue further through simulations in Section 4. Furthermore, if any quarter is observed together 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 1 month that makes up the quarter is opaque (i.e., unobserved).

As for $\mu\iota$, this can be subtracted directly from the relevant components of Y, so that $Y - \mu\iota \approx J\,X$. Then $Y - \mu\iota$ is modeled in lieu of the original Y. 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; f_k denotes the sampling period relative to the highest frequency. The matrix J for the logged data should have each unit value replaced by $1/f_k$, if that particular row corresponds to data at the kth lowest frequency. (Above, we had $f_1 = 3$.) The compensating mean vector is now written $\mu\iota$ for a vector of known values μ and ι consisting of the corresponding zeroes and ones. These μ values are equal to $\log(f_k)$ in a row corresponding to the kth lowest frequency data. Then $Y - \mu\iota \approx J\,X$, and we treat the approximation as an exact equality.

Thus $Y - \mu\iota$ is the observed data, which approximately equals $J\,X$. It is more convenient to make any adjustments $\mu\iota$ needed to the data, and call the result Y; then $Y = J\,X + E$, where E measures the approximation error between Y and $J\,X$. This error is unobserved, and might be considered to be stochastic, say with mean zero and covariance matrix Σ_E. We assume here that the error E is uncorrelated with X, mainly for simplicity. We next derive an expression for the likelihood that involves Σ_E, and we discuss the ramifications of ignoring the presence of Σ_E.

First, consider the assumptions of Theorem 1 on J and X, with the same notation, and let C be the block diagonal matrix consisting of the blocks Σ_* and $\underline{B}\Sigma_W\underline{B}'$. Also let $\varepsilon = RE$ (the matrix R is given in Theorem 1), with conformable decomposition $\varepsilon' = [\varepsilon_*', \underline{\varepsilon}']$, say. Thus $\Sigma_\varepsilon = R\Sigma_E R'$. Finally, define $U' = [U_*', \underline{U}']$, with $U_* = X_* + \varepsilon_*$ and $\underline{U} = \underline{B}W + \underline{\varepsilon}$. Then we can state the following result.

Theorem 2. Let the mixed sample Y of size m be written as $Y = J\,X + E$ for a high-frequency vector X that is a sample of size n from a difference stationary process with degree d differencing polynomial $\delta(B)$, and with E mean zero, uncorrelated with X, and having covariance matrix Σ_E. Suppose that the initial values $X_* = [1_d \; 0]PY$ are contiguous and uncorrelated with $W = \Delta X$, and appear as the first d values of $Y - E$, that is, $[1_d \; 0]J = [1_d \; 0]P$. Let

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

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 $J\widetilde{\Delta}^{-1}(P)$—see (4). Then -2 times the log Gaussian likelihood can be written as

$$U'C^{-1}U + \log|C| - U'C^{-1}\Sigma_E[1_m + C^{-1}\Sigma_\varepsilon]^{-1}C^{-1}U \\ - \log|1_m + C^{-1}\Sigma_\varepsilon|,$$

where the first two terms correspond to the likelihood in Theorem 1, and together factorize. But the latter two terms do not factorize in general.

Note that if $\Sigma_E = 0$, we recover the likelihood of Theorem 1. Because Σ_ε need not have a block structure, the Gaussian likelihood will not factorize when Σ_E is nonzero. Omitting the log determinantal terms, the scaled log-likelihood is lowered by nontrivial Σ_ε; when the noise-to-signal ratio matrix $C^{-1}\Sigma_\varepsilon$ is sufficiently close to the zero matrix, there is little change from the likelihood of Theorem 1. Likewise, for the determinantal terms we have $\log|\Sigma_Y| = \log|1_m + C^{-1}\Sigma_\varepsilon| + \log|C|$, which is approximately $\log|C|$ when the noise-to-signal ratio is small. This suggests that we can measure the noise-to-signal ratio via $\log|1_m + C^{-1}\Sigma_\varepsilon|$, with values close to zero indicating that Σ_E is relatively negligible.

The difficulty with the more general likelihood given in Theorem 2 is that the additional terms (due to nonzero Σ_ε) prevent factorization, which is inconvenient for applications. So in practice we might wish to use the likelihood (6) of Theorem 1 instead—where $DY = \underline{B}W + \underline{\varepsilon}$—whenever it is believed that the signal-to-noise ratio is high. Otherwise, the approximation of Y by JX will be poor and other techniques (such as a nonlinear method) should be used instead.

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, that is, backcasts and forecasts, then this can be solved through the theory of projections. Missing data, that is, omissions in the observed data JX at any frequency, are handled with the same theory. More generally, any linear function of the high-frequency vector X 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 $\{X_t\}$. This means that there is an $r \times n$ matrix I such that $Z = IX$ 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 $\{X_t\}$ 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 IX optimally from JX. Parzen (1961) provided 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 used for Theorem 1 hold, namely, that at least d of the highest frequency values are contiguously observed and are uncorrelated with $\{W_t\}$. We also assume that X has mean zero—if the data have mean regression effects $\mathbb{X}\beta$, we just apply the following Theorem to $X - \mathbb{X}\beta$ instead of X, and then add back $I\mathbb{X}\beta$ to the estimator of $\widehat{IX}$.

Theorem 3. Let the mixed sample Y of size m be written as $Y = JX$ for a high-frequency vector X that is a sample of size n from a difference stationary process with degree d differencing polynomial $\delta(B)$. Suppose that the initial values $X_* = [1_d \; 0]PX$ are contiguous and uncorrelated with $W = \Delta X$, and are observed as the first d values of Y, that is, $[1_d \; 0]J = [1_d \; 0]P$. Then the formula for the optimal estimate of IX from JX is

$$\widehat{IX} = I\,\Sigma_X\,J'\,\Sigma_Y^{-1}\,Y = I\,\widetilde{\Delta}^{-1}(P)\begin{bmatrix} Y_* \\ \Sigma_W\underline{B}'\left(\underline{B}\Sigma_W\underline{B}'\right)^{-1}DY \end{bmatrix}, \quad (8)$$

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

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

where $\dagger$ denotes inverse transpose.

Remark 1. The first formula in (8) is the general expression from Parzen (1961), but the second formula is practicable for implementation. The estimate $\widehat{IX}$ 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 given the matrix I. Its diagonal gives the mean squared error, and $1 - \alpha$ prediction intervals are given by

$$\widehat{IX} \pm q_{1-\alpha/2}\,\sigma,$$

where σ is the vector of square roots of the mean squared errors. Also $q_{1-\alpha/2}$ is the upper right quantile of the high-frequency marginal distribution; it is common to set $q_{1-\alpha/2} = 2$ for approximately Gaussian 95% prediction intervals. To guarantee that $\Sigma_W - \Sigma_W\underline{B}'(\underline{B}\Sigma_W\underline{B}')^{-1}\underline{B}\Sigma_W$ is positive definite (this can fail due to numerical rounding errors), it may be useful to use the numerical approximation $(\Sigma_W^{-1} + \kappa\underline{B}'\underline{B})^{-1}$ for large κ, which is more likely to be positive definite. This approximation follows from the Sherman–Morrison–Woodbury formula:

$$\left(\Sigma_W^{-1} + \kappa \underline{B}'\underline{B}\right)^{-1} = \Sigma_W - \Sigma_W \underline{B}'\left(\kappa^{-1} 1_{m-d} + \underline{B}\Sigma_W \underline{B}'\right)^{-1} \underline{B}\Sigma_W.$$

Remark 2. In the case that I has the structure [1 0], we obtain a vector of forecasts, backcasts, and imputations for X from Y. Also in the special case that $I = J$, we immediately recover Y, as is seen from the first formula of (8).

Note that in producing estimates of $I\,X$, we do not treat the classical signal extraction problem of estimating a latent process S when $X = S + N$ and N is noise (see Bell 1984). The nature of the signals S and $I\,X$ 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 X—which in turn will determine J. We construct I and J such that the same vector X is featured in both Z and Y. That is, if I is $r \times n$ and requires a certain span of the $\{X_t\}$ series for its definition, some of these X_t variables may not be featured anywhere in the available data matrix Y. In that case, the corresponding column of J will have all zeroes. For example, suppose we have only a single frequency and the data are $Y = [X_1, X_2]'$, but our target is $Z = X_0$. Then $X = [X_0, X_1, X_2]'$ 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 X vector out of the collection of all $\{X_t\}$ variables featured in the definitions of Z and Y.

The main application we have interest in occurs where we wish to compute a signal of the form $\sum_j \psi_j X_{t-j}$ 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:

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

Then we have $Z = \Psi\,X$, where $n = r + 2q$. In terms of Theorem 3, $I = \Psi$, 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 Y. Since $Y = J\,X$ and J is $m \times n$, this means that $r = n$, and Ψ has n rows and $n + 2q$ columns. We must extend X to apply the Theorem 3. So let $\widetilde{X} = [X_{-q+1}, \ldots, X_0, X', X_{n+1}, \ldots, X_{n+q}]'$, where the middle portion is just our original $X = [X_1, \ldots, X_n]'$. 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 $Y = \widetilde{J}\,\widetilde{X}$, and we run the method on this new J and extended X. This will change some of the formulas (n gets updated to $n + 2q$, etc.). Then the application of Theorem 3 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 $\Psi\,K\,X$, where we have used $K\,X$ to denote an entire sweep of some lower frequency series. For example, if X were monthly and we wanted quarterly values, we could write down K by selecting or aggregating (for stock and flow cases, respectively) components of X. Then with Ψ designed appropriately for that particular frequency (one must avoid using a monthly filter on quarterly data!), our target is $\Psi\,K\,X$.

Say one has 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, that is, 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:

$$K = \begin{bmatrix} 0\,0\,1\,0\,0\,0\,0 \ldots \\ 0\,0\,0\,0\,0\,1\,0 \ldots \end{bmatrix} \qquad K = \begin{bmatrix} 1\,1\,1\,0\,0\,0\,0 \ldots \\ 0\,0\,0\,1\,1\,1\,0 \ldots \end{bmatrix}.$$

Now the number of rows of Ψ is r, so it has $r + 2q$ columns, and hence K and X need to be extended. In fact, we need to extend Y to $\widetilde{X}$ 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 $\widetilde{K}\,\widetilde{X}$ has q low-frequency values appended fore and aft to $K\,X$. Once we have determined all these things, we have $I = \Psi\,\widetilde{K}$ in Theorem 3, and obtain $\widetilde{J}$ as described above such that $Y = \widetilde{J}\,\widetilde{X}$.

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 Y, 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 onward. If the filters are coherent and the model is decent, the results should splice.

4. SIMULATION STUDIES

In this section, we seek to investigate the performance of our methodology by addressing two aspects: (i) if the relative proportion—in terms of how much of the whole sample of data they each occupy—of two frequencies is unbalanced, how are results affected? (ii) in the case of flow requiring a log transformation, how are results sensitive to the ratio of sampling frequencies? We are interested in evaluating these questions through criteria associated with model fitting, but also through the quality of seasonal adjustment—this being a key application of interest in the article. It will be important to assess how sample size and the underlying dynamics of the process exert an impact on both model fitting and seasonal adjustment. One limitation of our study is the exclusion of the model misspecification question.

To these ends, we consider simulations of monthly Gaussian time series $\{X_t\}$ drawn from an airline process

$$(1 - B)(1 - B^{12})X_t = (1 - \theta B)(1 - \Theta B^{12})\epsilon_t,$$

with $\{\epsilon_t\}$ a white-noise process of variance one. Note that W_t is the moving average process $(1 - \theta B)(1 - \Theta B^{12})\epsilon_t$. We consider four specifications, which allow for some variation in both the trend and seasonal dynamics:

1. Process 1: $\theta = 0.3$, $\Theta = 0.3$
2. Process 2: $\theta = 0.3$, $\Theta = 0.6$
3. Process 3: $\theta = 0.3$, $\Theta = 0.9$
4. Process 4: $\theta = 0.6$, $\Theta = 0.6$

Note that Process 1 has a fairly chaotic (rapidly changing) seasonal pattern, while Process 3 has more stable seasonality. For background on such processes and the implications for seasonal adjustment, see Hillmer and Tiao (1982) and Bell and Hillmer (1991). For these processes, the lag one autocorrelation is equal to $-\theta/(1 + \theta^2)$, whereas the lag 12 autocorrelation is $-\Theta/(1 + \Theta^2)$; the lag 11 and lag 13 autocorrelations are both equal to the product of the lag 1 and 12 autocorrelations. Hence, increasing θ (e.g., Process 2 compared to Process 4) generates a more substantial negative correlation at lag one, while comparing the first three processes provides a contrast in the lag 12 correlation.

To investigate the impact of sample size, we suppose the underlying monthly series to be of length $T = 60, 120, 240$, which corresponds, respectively, to series that are "too short," "moderate," and "lengthy" for purposes of model-based seasonal adjustment. We also consider three ratios $\rho = 0.3, 0.5, 0.7$ corresponding to the proportion of monthly observations that are only available at a quarterly sampling frequency. For example, with $T = 120$ we consider scenarios in which 36, 60, or 84 of the monthly observations actually correspond to Q quarterly observations (respectively, either 12, 20, or 28 quarters). These quarterly observations are produced in one of three ways: if the series is a stock, we simply subsample systematically; if the series is a flow, we aggregate the subsamples appropriately; if the series is a flow that is modeled with a log transformation, we exponentiate the monthly simulation, flow aggregate, and then take logarithms to produce the mixed sample.

Each simulation is then fitted with a correctly specified airline model for the monthly frequency—note that the quarterly portion is assigned to the latter end of the time series. In the most meager scenario with $T = 60$ and $\rho = 0.7$, there are only 18 monthly observations, followed by 14 quarterly observations; since the 18 contiguous high-frequency observations exceed the total order of differencing ($d = 13$), we can apply our methodology. The model fit is assessed through analysis of the time series residuals ϵ (discussed in Section 2.4) via the lag 24 Ljung–Box (LB) statistic and the difference-sign (DS) test, each at a nominal $\alpha = 0.05$ Type I error rate. The proportion of rejections across 1000 simulations is recorded in Table 1. The MLEs for θ and Θ were also obtained, but tended not to be clustered about the true parameters—estimating parameters through the mixed frequency framework may well interfere with asymptotic consistency. Being less easy to interpret, and ultimately less important than the time series residuals for the purpose of assessing model fit, we have omitted a summary of the MLEs.

Each simulation is also seasonally adjusted by first computing an appropriate number of forecasts, backcasts, and imputations, to which an X-11 filter was applied. To match our subsequent empirical investigations, we used the same specification for the X-11 filter: we used 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). The quality of seasonal adjustment can be assessed through the presence or absence of seasonal spectral peaks (Soukup and Findley 1999). In this case, we apply the nonparametric spectral peak test of McElroy and Holan (2009) to each trend-differenced seasonal adjustment; recall that the output seasonal adjustments are available at a monthly frequency for the entire monthly data length T. There are five spectral peak tests, one for each monthly seasonal frequency $\pi/6$, $2\pi/6$, $3\pi/6$, $4\pi/6$, and $5\pi/6$, and the maximum of the normalized test statistics is used at the appropriate $\alpha = 0.05$ quantile (the test statistics for various peaks are asymptotically independent and normally distributed). The null hypothesis in this case is a locally flat spectrum, so even moderate departures toward residual seasonal peaks can produce a significant test statistic (we use an upper one-sided alternative, wishing to test against residual peaks, but being content with a residual spectral trough). The seasonal adjustments were very rarely flagged as being inadequate according to this criterion, as shown in Tables 1–3.

We note that the rejection rate for the LB statistic tended to exceed the nominal in the majority of cases, with little improvement due to sample size (when $T = 60$, $Q = 42$, there are too few observations to compute a lag 24 autocovariance, so these cells are reported as an NA). The DS statistics tended to be closer to nominal, with worse size in the cases where $\rho = 0.5$ and sample size is small. In the stock case (Table 1), there was little appreciable impact due to changing autocorrelation in the underlying process, and the story is similar for flows (Tables 2 and 3). Seasonal adjustments were rarely inadequate at sample size $T = 240$, and never inadequate with smaller sample sizes. Overall, the method is able to handle different sampling ratios at small to moderate sample sizes quite well, at least for the purpose of seasonal adjustment.

These studies address the first question (i) posed in this section, namely, how the ratio ρ—as well as sample size T, series type (stock versus flow), and level of autocorrelation—affects model fitting and seasonal adjustment quality. The second question (ii) requires us to investigate differing sampling frequencies. Recall from Section 2.5 the notation f_k for the sampling period, relative to the highest frequency. We consider monthly ($k = 1$), quarterly ($k = 2$), bi-annual ($k = 3$), and annual data ($k = 4$), so that $f_1 = 1$, $f_2 = 3$, $f_3 = 6$, and $f_4 = 12$, respectively. When we consider a flow that is modeled in its logs, we offset the logged simulation by $\log f_k$ for observations of the kth lowest frequency. By studying monthly to quarterly, monthly to bi-annual, and monthly to annual comparisons in sequence, we

Table 1. Simulation results for stock series

Airline process diagnostic	(0.3, 0.3)			(0.3, 0.6)			(0.3, 0.9)			(0.6, 0.6)		
	LB	DS	Spec	LB	DS	Spec	LB	DS	Spec	LB	DS	Spec
$T = 60, Q = 18$	0.085	0.035	0	0.086	0.045	0	0.096	0.038	0	0.089	0.043	0
$T = 60, Q = 30$	0.095	0.097	0	0.084	0.110	0	0.087	0.098	0	0.089	0.102	0
$T = 60, Q = 42$	NA	0.055	0	NA	0.043	0	NA	0.057	0	NA	0.052	0
$T = 120, Q = 36$	0.063	0.026	0	0.097	0.038	0	0.084	0.041	0	0.098	0.034	0
$T = 120, Q = 60$	0.077	0.056	0	0.094	0.047	0	0.092	0.053	0	0.102	0.058	0
$T = 120, Q = 84$	0.080	0.030	0	0.085	0.021	0	0.075	0.019	0	0.085	0.028	0
$T = 240, Q = 72$	0.064	0.057	0	0.065	0.050	0	0.075	0.044	0	0.103	0.047	0
$T = 240, Q = 120$	0.060	0.072	0	0.086	0.058	0	0.084	0.067	0	0.106	0.069	0
$T = 240, Q = 168$	0.075	0.034	0.004	0.087	0.031	0	0.107	0.035	0	0.121	0.035	0

NOTES: Simulated airline processes with $T - Q$ monthly stock observations followed by $Q/3$ quarterly stock observations. Four airline processes are considered, with nonseasonal and seasonal moving average parameters varying by column. Each cell reports the proportion of significant lag 24 Ljung–Box (LB) statistics, the proportion of significant difference-sign (DS) test statistics, or the proportion of significant seasonal spectral peak (Spec) test statistics, out of 1000 Monte Carlo simulations, at $\alpha = 0.05$.

can look for deterioration in performance. To focus our study somewhat, for this part we set $T = 120$, but still consider various choices of ρ.

The results are presented in Table 4. Again, changes in the process' autocorrelations are adequately handled by the model, but as the respective sampling period f_k increases—from quarterly to bi-annual to annual—the LB statistic becomes more missized. In the case of the last row, consisting of 7 annual observations ($A = 84$) and 36 monthly observations, the monthly pattern must be imputed for 92 % of the observations in each of the 7 years, based on only 36 monthly observations. Nevertheless, the seasonal adjustments continue to be adequate. At least for this particular set of processes, and assuming the model has been correctly specified, the mixed frequency methodology appears to perform reasonably well.

5. 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. Second, we perform a similar exercise with the flow series of industrial production, modeled with a seasonal long memory process, and illustrate the impact of the proportion of months to quarters. Third, 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. Fourth, 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 2014).

5.1 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 used, this would introduce extra error at the model fitting stage—to isolate the X-11

Table 2. Simulation results for untransformed flow series

Airline processes Diagnostic	(0.3,0.3)			(0.3,0.6)			(0.3,0.9)			(0.6,0.6)		
	LB	DS	Spec	LB	DS	Spec	LB	DS	Spec	LB	DS	Spec
$T = 60, Q = 18$	0.075	0.047	0	0.077	0.047	0	0.078	0.027	0	0.099	0.033	0
$T = 60, Q = 30$	0.115	0.065	0	0.100	0.096	0	0.100	0.110	0	0.095	0.083	0
$T = 60, Q = 42$	NA	0.047	0	NA	0.049	0	NA	.046	0	NA	0.045	0
$T = 120, Q = 36$	0.071	0.034	0	0.108	0.035	0	0.104	0.044	0	0.115	0.036	0
$T = 120, Q = 60$	0.083	0.054	0	0.085	0.059	0	0.085	0.058	0	0.079	0.046	0
$T = 120, Q = 84$	0.101	0.029	0	0.104	0.037	0	0.099	0.022	0	0.085	0.040	0
$T = 240, Q = 72$	0.069	0.057	0	0.097	0.051	0	0.089	0.056	0	0.105	0.047	0
$T = 240, Q = 120$	0.075	0.062	0.004	0.073	0.077	0	0.112	0.063	0	0.132	0.076	0
$T = 240, Q = 168$	0.073	0.038	0.015	0.066	0.034	0	0.092	0.043	0.001	0.076	0.034	0.001

NOTES: Simulated airline processes with $T - Q$ monthly flow observations followed by $Q/3$ quarterly flow observations. Four airline processes are considered, with nonseasonal and seasonal moving average parameters varying by column. Each cell reports the proportion of significant lag 24 Ljung–Box (LB) statistics, the proportion of significant difference-sign (DS) test statistics, or the proportion of significant seasonal spectral peak (Spec) test statistics, out of 1000 Monte Carlo simulations, at $\alpha = 0.05$.

Table 3. Simulation results for log flow series

Airline processes Diagnostic	(0.3,0.3) LB	DS	Spec	(0.3,0.6) LB	DS	Spec	(0.3,0.9) LB	DS	Spec	(0.6,0.6) LB	DS	Spec
$T = 60$, $Q = 18$	0.101	0.042	0	0.100	0.052	0	0.075	0.051	0	0.096	0.035	0
$T = 60$, $Q = 30$	0.100	0.107	0	0.088	0.100	0	0.098	0.080	0	0.079	0.100	0
$T = 60$, $Q = 42$	NA	0.044	0	NA	0.045	0	NA	0.044	0	NA	0.049	0
$T = 120$, $Q = 36$	0.073	0.033	0	0.081	0.044	0	0.104	0.042	0	0.073	0.032	0
$T = 120$, $Q = 60$	0.085	0.047	0	0.072	0.077	0	0.100	0.047	0	0.089	0.055	0
$T = 120$, $Q = 84$	0.092	0.016	0	0.081	0.029	0	0.097	0.020	0	0.098	0.026	0
$T = 240$, $Q = 72$	0.058	0.056	0.001	0.061	0.052	0	0.090	0.045	0	0.094	0.046	0
$T = 240$, $Q = 120$	0.066	0.061	0.004	0.082	0.071	0	0.079	0.065	0	0.098	0.071	0
$T = 240$, $Q = 168$	0.067	0.039	0.011	0.078	0.035	0	0.085	0.029	0	0.083	0.022	0

NOTES: Simulated airline processes with $T - Q$ monthly flow observations followed by $Q/3$ quarterly flow observations, where modeling is done in logarithms. Four airline processes are considered, with nonseasonal and seasonal moving average parameters varying by column. Each cell reports the proportion of significant lag 24 Ljung–Box (LB) statistics, the proportion of significant difference-sign (DS) test statistics, or the proportion of significant seasonal spectral peak (Spec) test statistics, out of 1000 Monte Carlo simulations, at $\alpha = 0.05$.

filtering aspect and facilitate comparisons with truth, we presume that all fixed effects have been satisfactorily removed.

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. (We also did an analysis with only two full years of monthly data, but the quality of the results deteriorated somewhat; see further discussion below.) Then Y consists of the last 60 monthly values, followed by all the quarterly values in order, for all but the last five years; see Figure 1. In the following plots, the following colors are used consistently for this example: green represents the true full monthly time series, whereas black is what we pretend is only available to us for analysis. Projections are in blue, 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 (green) monthly time series, which is depicted in purple. Shaded intervals depict confidence intervals using two standard errors. 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 -0.08 and 0.61, respectively, we do not 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; the lag 24 Ljung–Box statistic was 20.904, and the difference-sign test statistic was -0.756, neither being significant. So 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 -0.13 and 0.66, 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. 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, that is, 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 3 to obtain the X-11 seasonal adjustment. (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.) We are

Table 4. Simulation results for changing frequency series

Airline processes Diagnostic	(0.3,0.3) LB	DS	Spec	(0.3,0.6) LB	DS	Spec	(0.3,0.9) LB	DS	Spec	(0.6,0.6) LB	DS	Spec
$T = 120$, $Q = 36$	0.073	0.033	0	0.081	0.044	0	0.104	0.042	0	0.073	0.032	0
$T = 120$, $Q = 60$	0.085	0.047	0	0.072	0.077	0	0.100	0.047	0	0.089	0.055	0
$T = 120$, $Q = 84$	0.092	0.016	0	0.081	0.029	0	0.097	0.020	0	0.098	0.026	0
$T = 120$, $B = 36$	0.084	0.071	0	0.075	0.088	0	0.080	0.062	0	0.100	0.071	0
$T = 120$, $B = 60$	0.106	0.033	0	0.105	0.031	0	0.107	0.026	0	0.086	0.037	0
$T = 120$, $B = 84$	0.255	0.049	0	0.189	0.054	0	0.128	0.042	0	0.115	0.044	0
$T = 120$, $A = 36$	0.076	0.041	0	0.082	0.043	0	0.078	0.055	0	0.095	0.051	0
$T = 120$, $A = 60$	0.078	0.049	0	0.095	0.060	0	0.103	0.045	0	0.088	0.070	0
$T = 120$, $A = 84$	0.132	0.068	0	0.138	0.062	0	0.111	0.060	0	0.120	0.062	0

NOTES: Simulated airline processes with differing sampling frequencies, for a flow process modeled in the logarithms. First, for a monthly-to-quarterly study, $T - Q$ monthly flow observations are followed by $Q/3$ quarterly flow observations; for a monthly-to-biannual study, $T - B$ monthly flow observations are followed by $B/6$ biannual flow observations; finally for a monthly-to-annual study, $T - A$ monthly flow observations are followed by $A/12$ annual flow observations. Four airline processes are considered, with nonseasonal and seasonal moving average parameters varying by column. Each cell reports the proportion of significant lag 24 Ljung–Box (LB) statistics, the proportion of significant difference-sign (DS) test statistics, or the proportion of significant seasonal spectral peak (Spec) test statistics, out of 1000 Monte Carlo simulations, at $\alpha = 0.05$.

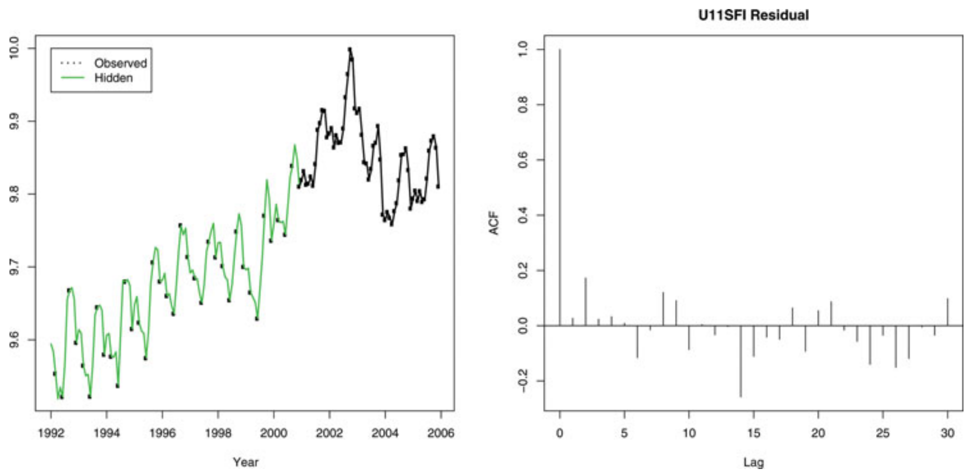

Figure 1. The left panel displays the mixed frequency sample along with the original monthly sample of *U11SFI*, whereas the right panel displays the ACF of the time series residuals from the fitted mixed frequency airline model. The spacing of the black dots in the left panel changes in 2001, which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

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, that is, the forecasts, backcasts, and imputations. Also overlaid are the monthly seasonal adjustments. We depict with shading the confidence intervals associated with two standard errors, where a standard error is the square root of the projection MSE (see Remark 1). 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 using the inferred monthly pattern (see the right panel of Figure 3).

To assess whether the seasonally adjusted values for the monthly portion—that is, January 2001 through December 2005—it is informative to apply the same X-11 filter to the complete monthly dataset (purple). For this exercise, we use the model that was fitted to the mixed frequency data, although one could use 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 directly examined (see Remark 1 for their calculation). 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

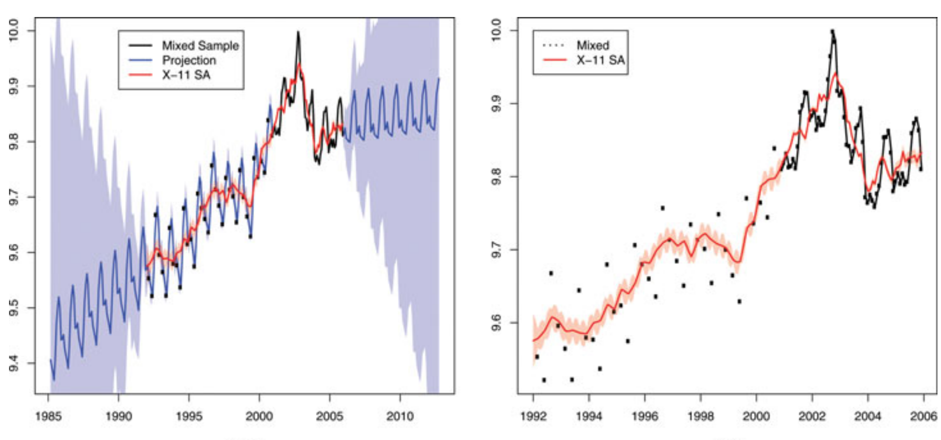

Figure 2. The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for *U11SFI*. The right panel displays the same information without the projections. The spacing of the black dots in these panels changes in 2001, which reflects the transition from quarterly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

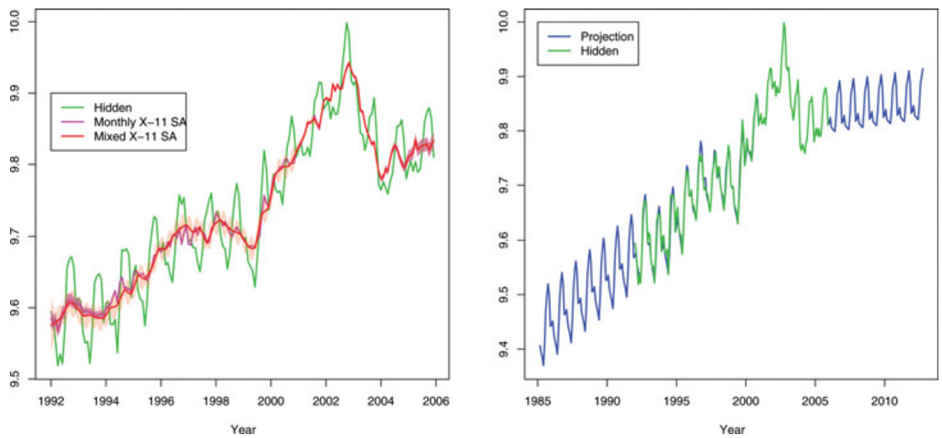

Figure 3. The left panel displays the seasonal adjustments for *U11SFI* using both the mixed sample and the entire monthly data span. The right panel compares the projections directly to the actual partially unobserved true monthly data.

of finite-sample MSEs for time series projections (see 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, with parameters -0.08 and 0.59. 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. This agrees with our simulation results for the case that ρ is small.

5.2 Application to Industrial Production

We extend the previous subsection's exercise to a monthly flow series, namely, industrial production (Source: Federal Reserve Board 1959–2007), where we fictitiously generate quarterly flows by averaging 3 monthly values at a time, from the beginning of the sample. The span (in years) of monthly data, at the end of the series, is taken to be 10, 20, 30, or 40, yielding a progressively greater proportion of monthly observations. Note that we do not use a log transformation for this example, and that we define the quarterly flows by averaging, rather than summing, because this allows for easier visualization.

We are interested in the performance of long memory models in this context, and therefore consider a 7-GEXP (generalized EXPonential) model with seven spectral peaks, discussed in McElroy and Holan (2012). This model is written in the frequency domain, with spectral density f for the trend-differenced monthly high-frequency process given by

$$f(\lambda) = (2 - 2\cos\lambda)^{-a}\,(2 + 2\cos\lambda)^{-b} \times \prod_{j=1}^{k} \left|1 - 2\cos(\pi j/6)e^{-i\lambda} + e^{-i2\lambda}\right|^{2c_j} g(\lambda)$$

$$g(\lambda) = \exp\left\{\sum_{j=1}^{3} g_j \cos(\lambda j)\right\}\sigma^2.$$

The short memory part of the model is provided by g, and the parameters g_1, g_2, g_3 are real numbers known as cepstral coefficients (Bloomfield 1973), with σ^2 being the innovation variance. The long memory portion of the model—including

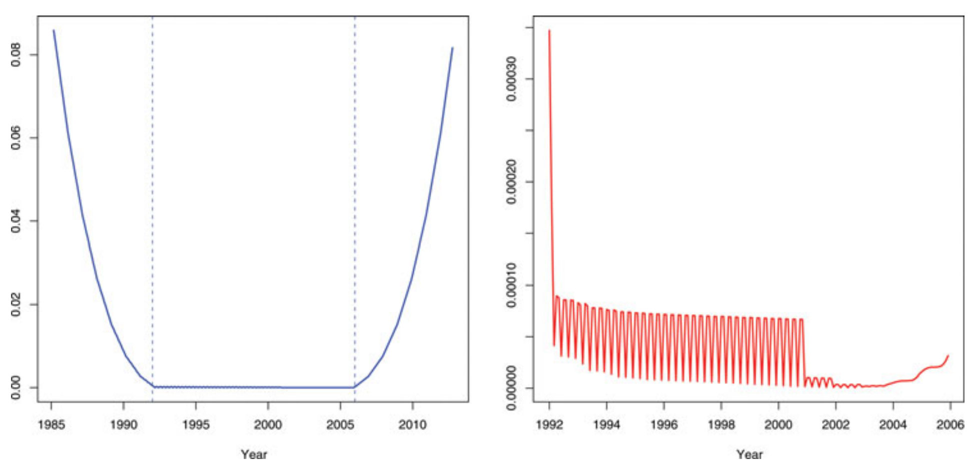

Figure 4. The left panel displays the MSEs for the full projections for *U11SFI*, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.

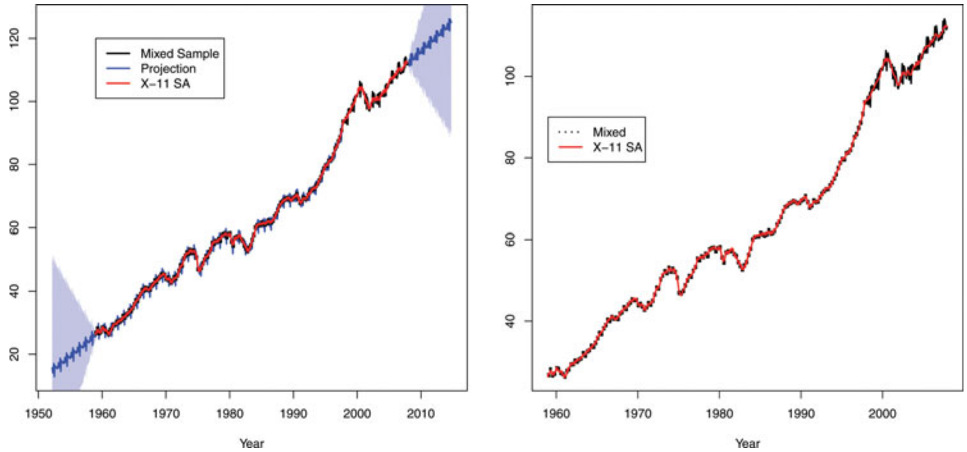

Figure 5. The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for industrial production. The right panel displays the same information without the projections. This case uses 10 years of monthly and 39 years of quarterly observations.

seasonal long memory and possible anti-persistence—is parameterized through $a, b, c_1, \ldots, c_5$, which correspond to spectral behavior at frequencies $0, \pi, \pi/6, \ldots, 5\pi/6$, respectively. Positive exponents correspond to long memory (at the corresponding frequency), whereas negative exponents correspond to negative memory, or anti-persistence (McElroy and Politis 2014 provided an overview).

We expect that as the proportion of monthly observations decreases, the model will be more difficult to fit. This is because in the extreme case that the data are completely quarterly, only three spectral peaks—instead of seven—will be present, so that some of the memory parameters will not be identified. Therefore, we are interested in the modeling performance as the monthly span is adjusted, and we examine both model fit and seasonal adjustment. We are not aware of any other methods (there seems to be no literature on embedding seasonal long memory processes into an SS framework, although it may be possible to extend Chan and Palma (1998) to the mutiple-spectral pole case) that can address this type of model structure. Whereas a SARIMA model, such as is used in the U11SFI, is reasonable for the industrial production, there seems to be some evidence that a seasonal long memory approach, as provided by the 7-GEXP model, can ably produce a better frequency domain description of the seasonality.

Using the spans of 10, 20, 30, and 40, we obtain fitted models in each case after several minutes of computing (the sample size is much larger here than for U11SFI), with sign test p-values of 0.42, 0.36, 0.12, and 0.08, respectively. This indicates the residuals have a bit more serial correlation when more monthly data are present. The maximum likelihood estimates of the parameters show a quite high degree of long memory for all seven frequencies, albeit the first seasonal $\pi/6$ is much more moderate than the others—this behavior persists for all four spans.

We apply the same seasonal adjustment filters as for the U11SFI example, and display the working data, the fitted spectral density (for the underlying trend-differenced high-frequency model), and the seasonal adjustment; see Figures 5–8. We also compared these adjustments to those obtained by having the full monthly data, but there was no visual discrepancy, and hence these figures are omitted. The standard errors for projections and adjustments were quite small, being hardly visible (excepting for forecasts and backcasts). Overall, the adjustments were

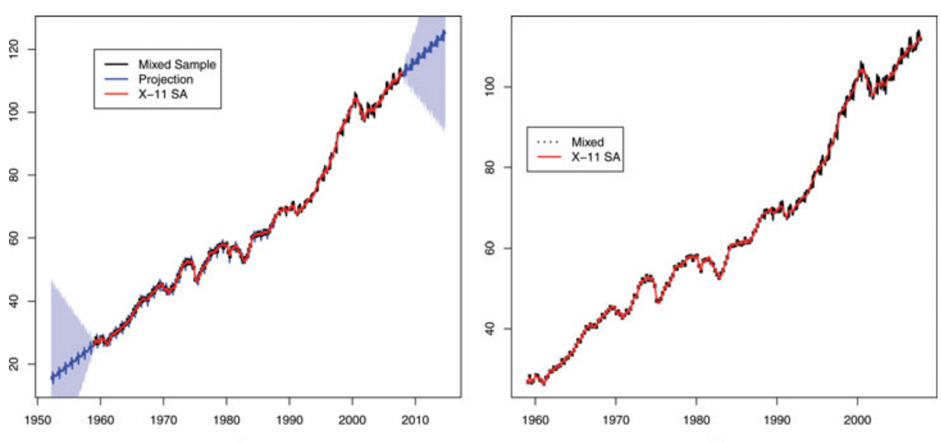

Figure 6. The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for industrial production. The right panel displays the same information without the projections. This case uses 20 years of monthly and 29 years of quarterly observations.

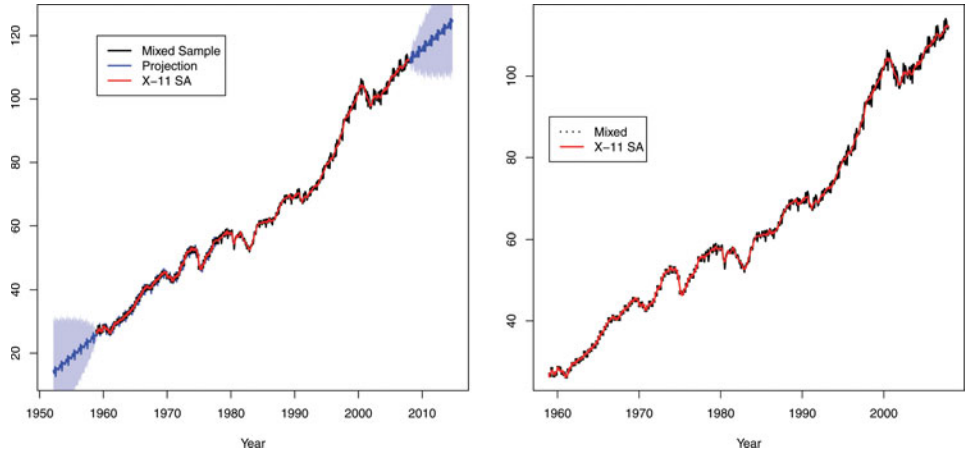

Figure 7. The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for industrial production. The right panel displays the same information without the projections. This case uses 30 years of monthly and 19 years of quarterly observations.

quite adequate, and the methodology worked quite satisfactorily for all four spans.

5.3 Application to Sterling Series

Our third 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 dataset (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 (see *http://www.bankofengland.co.uk/statistics/abl/current/index.htm*.) 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 these data have 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

$$(1-0.96B)(1-B)(1-B^{12})\,[X_t - 0.844\; LS_t] \\ = (1-0.79B)(1-0.71B^{12})\epsilon_t.$$

The fit produced residuals with little apparent serial structure; the lag 24 Ljung–Box statistic was 19.453, and the difference-sign test statistic was 1.715 (significantly different from zero with p-value 0.043).

Then we computed all projections and applied an X-11 monthly seasonal adjustment filter (same specification as with U11SFI). The left panel of Figure 9 displays the available data together with all projections and the seasonal adjustments (with two standard error confidence intervals), while the right panel

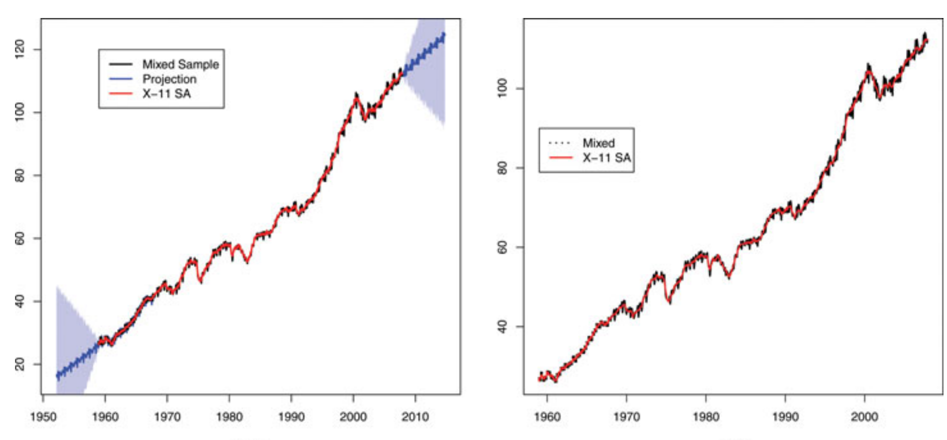

Figure 8. The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for industrial production. The right panel displays the same information without the projections. This case uses 40 years of monthly and 9 years of quarterly observations.

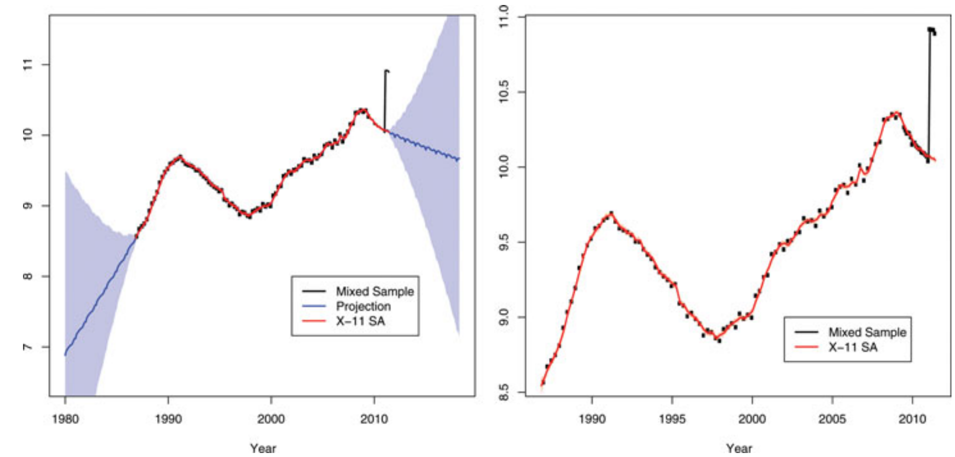

Figure 9. The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data for the *Sterling* series. The right panel displays the same information without the projections. The spacing of the black dots in the left panel changes in 2009, which reflects the transition from quarterly to monthly frequencies.

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 focused picture with the projections removed. For production purposes, the level shift would be reincorporated 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 10 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.

5.4 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 fourth example is a flow series where the frequency of observation decreases. Available as a monthly flow from 1977 onward, 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, -0.79, -21.66 for Monday through Sunday. Since the model was adequate without a transformation, we chose not to use a logarithmic transformation. The final model, after some analysis of competing SARIMA specifications, was again an airline model with nonseasonal and seasonal moving average parameters of 0.64 and

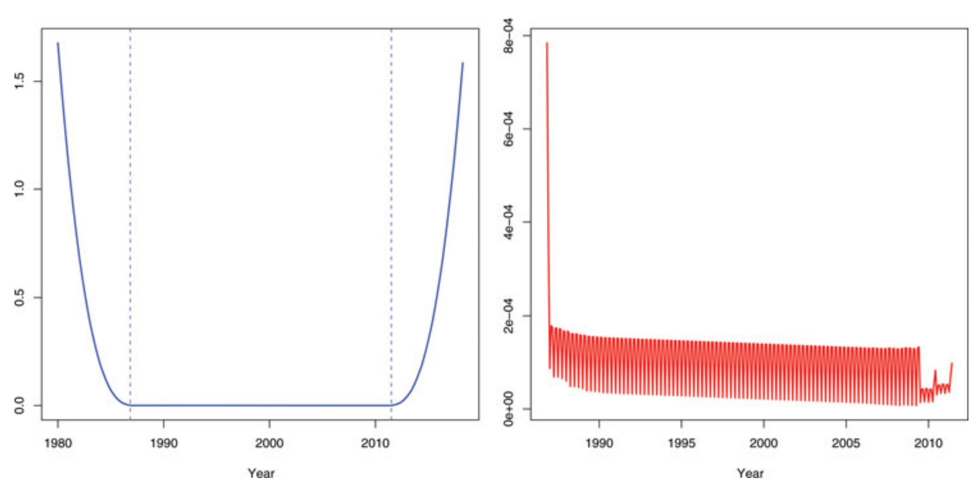

Figure 10. The left panel displays the MSEs for the full projections for the *Sterling* series, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.

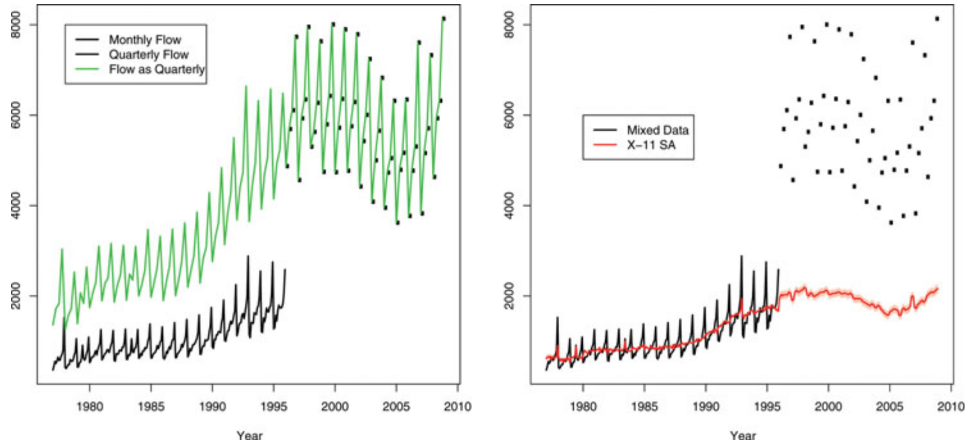

Figure 11. The left panel displays the mixed frequency sample for the *Construction* series, together with the monthly flow aggregated to a quarterly frequency (in red). The right panel displays the monthly data, full projections, and monthly seasonally adjusted data for the *Construction* series.

0.52. The fit produced residuals with little apparent serial correlation; the lag 24 Ljung–Box statistic was 19.842, while DS was -1.481.

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 have much higher values (since it is obtained by summing higher frequency values). Figure 11 displays the Construction series: its monthly portion is in black, and the higher quarterly portion is given with black dots. We also converted the monthly flows to quarterly flows by aggregation, and plotted in green these numbers alongside the latter quarterly flows. Projection results in Figure 11 show forecasts and backcasts, but the later quarterly data has been left out of the picture for easier viewing. Note that the projections take into account regression preadjustment (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 12 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,

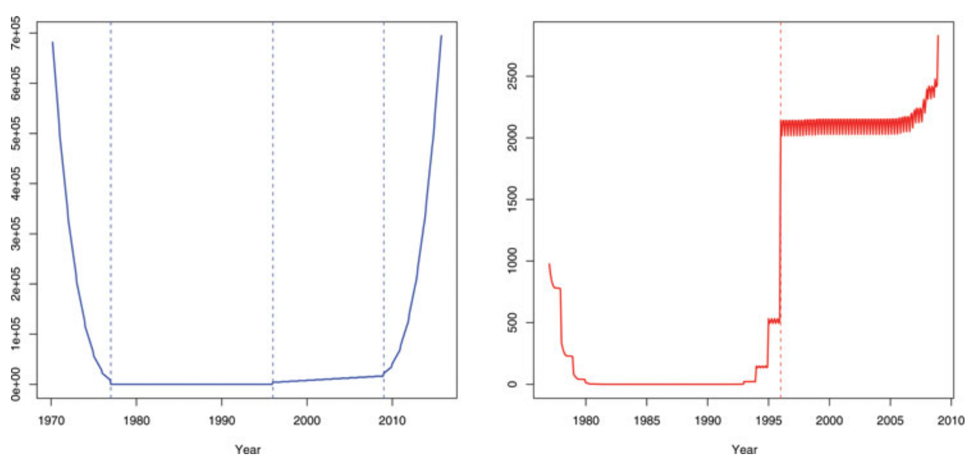

Figure 12. The left panel displays the MSEs for the full projections for the *Construction* series, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.

with additional uncertainty due to the flow structure relative to stock series.

6. CONCLUSION

The research of this article 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 article'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 use 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 cases of long memory models and exponential models cannot be given an exact treatment with SS methodology, and recourse to a direct method is recommended. Practical issues, such as dealing with logarithmic transformations of flow variables, as well as regressors, are addressed herein 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 backward 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). To ensure that high-frequency structure can be appropriately imposed on low-frequency observations with changing level, an $I(2)$ model was 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 $\{X_t\}$ when fitting to the mingled data as well; but this strategy is only available when the data are 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, that is, 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 5—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—as supported by our simulation studies of Section 4—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 2014), and all programs and scripts are available from the authors upon request. (This methodology has not yet been incorporated into X-12-ARIMA or X-13 ARIMA-SEATS.)

APPENDIX: PROOFS OF RESULTS

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

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

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 $k = 1, 2, \ldots, d$ and the $\{\xi_j\}$ are the coefficients of $\delta^{-1}(B)$. Here $\underline{X}_0 = [X_1, X_2, \ldots, X_d]'$, which is a "canonical" choice of initial values; in this case the corresponding P is the identity matrix, and the admissibility of X_0 follows at once from the fact that $\widetilde{\Delta}(1_\ell)$ is unit lower triangular (hence invertible) for any given ℓ. Then for any admissible X_*, we have

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

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

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

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 $X_* = C\underline{X}_0 + BW$ with C a $d \times d$ matrix and B a matrix with d rows; see the representation in eq. (4.1) of Bell and Hillmer (1991). This relates X_* to $\underline{X}_0$, but we can also relate X_* to any other admissible initial values in this way. □

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

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

partitioned into d and $m-d$ rows, and d and $n-d$ columns. Then $R\,Y = R\,J\,X$ is given by

$$R\,Y = \begin{bmatrix} \overline{A} & \overline{B} \\ \underline{A} & \underline{B} \end{bmatrix} \begin{bmatrix} X_* \\ W \end{bmatrix}.$$

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

$$\begin{aligned}
R\,\Sigma_Y\,R' &= \begin{bmatrix} \overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}' & \overline{A}\Sigma_*\underline{A}' + \overline{B}\Sigma_W\underline{B}' \\ \underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}' & \underline{A}\Sigma_*\underline{A}' + \underline{B}\Sigma_W\underline{B}' \end{bmatrix} \\
S &= \underline{A}\Sigma_*\underline{A}' + \underline{B}\Sigma_W\underline{B}' - \left(\underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}'\right) \\
&\quad \left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}\left(\overline{A}\Sigma_*\underline{A}' + \overline{B}\Sigma_W\underline{B}'\right) \\
Y'\Sigma_Y^{-1}Y &= \left[\overline{A}X_* + \overline{B}W\right]'\left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1} \\
&\quad \left[\overline{A}X_* + \overline{B}W\right] \\
&\quad + \Big[\underline{A}X_* + \underline{B}W - \left(\underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}'\right) \\
&\quad \left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}\left(\overline{A}X_* + \overline{B}W\right)\Big]' \\
&\quad \cdot\; S^{-1}\Big[\underline{A}X_* + \underline{B}W - \left(\underline{A}\Sigma_*\overline{A}' + \underline{B}\Sigma_W\overline{B}'\right) \\
&\quad \left(\overline{A}\Sigma_*\overline{A}' + \overline{B}\Sigma_W\overline{B}'\right)^{-1}\left(\overline{A}X_* + \overline{B}W\right)\Big].
\end{aligned}$$

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 $[\overline{A}X_*]'(\overline{A}\Sigma_*\overline{A}')^{-1}[\overline{A}X_*]$ 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 ∇_{X_*} to $Y'\Sigma_Y^{-1}Y$, we find that $\underline{A} = 0$ must hold (otherwise S will depend on Σ_*). This shows that $R\,J\,\widetilde{\Delta}^{-1}(P)$ is block diagonal as desired. □

Proof of Lemma 1. Contiguity means that the indices $\{t_1, \ldots, t_d\}$ of X_* are in a block $\{t_* + 1, \ldots, t_* + d\}$, and Π is the mapping of this set to $\{1, \ldots, d\}$. The matrix form of this is the first statement of the lemma. For the second assertion, suppose that $\widetilde{\Delta}(P)v = 0$, so that $\Delta v = 0$. Therefore, 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.) This can be written $v = Z\beta$ with $Z_{ji} = z_i(t_j)$. So $0 = [0\ \Pi\ 0]v = [\Pi\ 0]\underline{Z}\beta$, where $\underline{Z}$ consists of removing the first t_* rows of Z. Consider column operations M (with M invertible) such that $\underline{Z}M$ has first d rows equal to 1_d, which we can accomplish because the columns of Z are linearly independent. Thus,

$$0 = [\Pi\ 0]\underline{Z}\beta = \Pi M^{-1}\beta,$$

which implies that $\beta = 0$, and hence that $v = 0$. This shows that $\widetilde{\Delta}(P)$ is invertible. □

Proof of Theorem 1. The invertibility of $\widetilde{\Delta}(P)$ follows from contiguity and 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 $R\,J\,\widetilde{\Delta}^{-1}(P)$ follows at once from the definition of R, which is clearly lower triangular with unit diagonal (and hence invertible). Also

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

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

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

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

Proof of Theorem 2. First, we can write $J\Sigma_X J' = R^{-1}CR^{\dagger}$, and thus $(J\Sigma_X J')^{-1} = R'C^{-1}R$. Then $\Sigma_Y = R^{-1}CR^{\dagger} + \Sigma_E = R^{-1}[C + \Sigma_\varepsilon]R^{\dagger}$, where $\Sigma_\varepsilon = R\Sigma_E R'$. Using the Sherman–Morrison–Woodbury identity (Golub and Van Loan 1996),

$$\Sigma_Y^{-1} = R'(C^{-1} - C^{-1}\Sigma_\varepsilon[1_m + C^{-1}\Sigma_\varepsilon]^{-1}C^{-1})R.$$

Also we have

$$Y = R^{-1}\begin{bmatrix} X_* \\ \underline{B}W \end{bmatrix} + E = R^{-1}U.$$

As a result, we obtain

$$Y'\Sigma_Y^{-1}Y = U'C^{-1}U - U'C^{-1}\Sigma_\varepsilon[1_m + C^{-1}\Sigma_\varepsilon]^{-1}C^{-1}U,$$

from which the result of the Theorem now follows. □

Proof of Theorem 3. The first line of $\widehat{IX}$ follows directly from Parzen (1961). Note that

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

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

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

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

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

which has covariance matrix

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

which simplifies to the stated formula. □

[Received June 2012. Revised September 2014.]

REFERENCES

Ansley, C., and Kohn, R. (1985), "Estimation, Filtering, and Smoothing in State Space Models With Incompletely Specified Initial Conditions," *The Annals of Statistics*, 13, 1286–1316. [1287]

Axelsson, O. (1996), *Iterative Solution Methods*, Cambridge, UK: Cambridge University Press. [1302]

Bell, W. (1984), "Signal Extraction for Nonstationary Time Series," *The Annals of Statistics*, 12, 646–664. [1286,1287,1291,1301]

Bell, W., and Hillmer, S. (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models," *Journal of Time Series Analysis*, 12, 283–300. [1285,1286,1287,1292,1301]

Bloomfield, P. (1973), "An Exponential Model for the Spectrum of a Scalar Time Series," *Biometrika*, 60, 217–226. [1287,1296]

Box, G., and Jenkins, G. (1976), *Time Series Analysis*, San Francisco, CA: Holden-Day. [1294]

Brockwell, P., and Davis, R. (1991), *Time Series: Theory and Methods*, New York: Springer. [1285,1288,1289]

Chan, N. H., and Palma, W. (1998), "State Space Modeling of Long-Memory Processes," *The Annals of Statistics*, 26, 719–740. [1287,1297]

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. [1284]

Cochran, W. (1934), "The Distribution of Quadratic Forms in a Normal System, With Applications to the Analysis of Covariance," *Proceedings of Cambridge Philosophical Society*, 30, 178–191. [1286]

Dagum, E. (1980), *The X-11-ARIMA Seasonal Adjustment Method*, Ottawa: Statistics Canada. [1284]

Doornik, J. (1998), *Object-Oriented Matrix Programming Using Ox 2.0*, London: Timberlake Consultants Press. [1284]

Durbin, J., and Koopman, S. (2001), *Time Series Analysis by State Space Methods*, Oxford, UK: Oxford University Press. [1284,1285,1287]

Durbin, J., and Quenneville, B. (1997), "Benchmarking by State Space Methods," *International Statistical Review*, 65, 23–48. [1284]

Golub, G., and Van Loan, C. (1996), *Matrix Computations*, Baltimore and London: The Johns Hopkins University Press. [1287,1302]

Hillmer, S., and Tiao, G. (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment," *Journal of the American Statistical Association*, 77, 63–70. [1292]

Holan, S., McElroy, T., and Chakraborty, S. (2009), "A Bayesian Approach to Estimating the Long Memory Parameter," *Bayesian Analysis*, 4, 159–190. [1288]

Koopman, S., Shephard, N., and Doornik, J. (1999), "Statistical Algorithms for Models in State Space Using SsfPack 2.2," *Econometrics Journal*, 2, 113–166. [1285]

Ladiray, D., and Quenneville, B. (2001), *Seasonal Adjustment With the X-11 Method* (Vol. 158), New York: Springer-Verlag. [1284,1290,1292,1294]

Maravall, A., and Caparello, G. (2004), Program TSW: Revised Reference Manual. Working Paper 2004, Research Department, Bank of Spain. Available at http://www.bde.es. [1284]

McElroy, T. (2008), "Matrix Formulas for Nonstationary ARIMA Signal Extraction," *Econometric Theory*, 24, 1–22. [1286,1296]

McElroy, T., and Holan, S. (2009), "Using Spectral Peaks to Detect Seasonality," *2009 Proceedings of the Federal Conference on Statistical Methodology*. [1292]

McElroy, T., and Holan, S. (2012), "On the Estimation of Autocovariances for Generalized Gegenbauer Processes," *Statistica Sinica*, 22, 1661–1687. [1287,1296]

McElroy, T., and Politis, D. (2014), "Spectral Density and Spectral Distribution Inference for Long Memory Time Series via Fixed-b Asymptotics," *Journal of Econometrics*, 182, 211–225. [1297]

Parzen, E. (1961), "An Approach to Time Series Analysis," *The Annals of Mathematical Statistics*, 32, 951–989. [1290,1302]

Pasteels, J.-M. (2012), "Seasonal Adjustment of Data Derived From Labour Force Surveys: Some Specific Issues," in *2012 Workshop on Methodological Issues in Seasonal Adjustment*, Luxembourg. Available at http://www.cros-portal.eu/page/2012-workshop-methodological-issues-seasonal-adjustment [1284]

R Development Core Team (2014), *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at *http://www.R-project.org/*. [1293,1301]

Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley. [1286]

Soukup, R., and Findley, D. (1999), "On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects After Modeling or Adjustment," in *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, pp. 144–149. [1292]

U. S. Census Bureau. (2011), "X-12-ARIMA Reference Manual," available at *http://www.census.gov/ts/x12a/v03/x12adocV03.pdf* [1289]

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, pp. 90–96. [1284]