Extension of X-12-ARIMA to Seasonally Adjust Mixed Frequency Stock and Flow Time Series

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

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

1 Introduction

Although the seasonal adjustment of economic time series is a vast undertaking at statistical agencies – such as the U.S. Census Bureau, Bureau of Labor Statistics, Statistics Canada, Bank of Spain, and many others – most of production operates on univariate time series observed over a constant sampling frequency, typically either monthly or quarterly. The statistical methods have developed accordingly: X-11-ARIMA (Dagum, 1980) works with a single sampling frequency, applying the fixed X-11 filters to a forecast-extended time series[1]. The advent of model-based approaches, such as in SEATS (Maravall and Caparello, 2004), also proceed by considering a single frequency. However, changes 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 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 (Zadrozny (1990) and Chen and Zadrozny (1998)). However, this 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. See Durbin and Quenneville (1997) and Durbin and Koopman (2001) for related material.

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

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

1998), but there are costs (both in terms of money and time spent learning the language) here as well. Second, the correct quantification of uncertainty associated with estimation of fixed signals (e.g., seasonally adjusted components defined via the X-11 filters) requires a full knowledge of the projection error covariance matrix. Some expertise and care is needed to produce these quantities from the state space method, which typically will just output the projection error variances, or diagonal entries of the error covariance matrix (see Koopman et al. (1999) and Durbin and Koopman (2001)). Third, numerical tricks are typically utilized, 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. This clouds the simplicity of the actual optimal solution to the projection problem, and introduces potential numerical inaccuracy.

There may be other disadvantages, but we are not aware of any. The advantages are numerous, and are generally felt to outweigh any weaknesses: computational efficiency is foremost, as well as flexibility and the power to handle many types of applications (Durbin and Koopman, 2001). Broadly speaking, these applications mainly fall in the category of calculating Gaussian conditional expectations, or equivalently, minimum mean square error linear estimates of unknown stochastic quantities. Following the language of Brockwell and Davis (1991), we refer to such conditional expectations as projections. Examples of projections include forecasts, missing value imputations, and signal extractions. The mathematical results of this paper include projection formulae for the mixed frequency situation described above. Depending on one's appetency regarding state space methods, the formulas may be viewed as a precise mathematical foundation for smoothing methods, or as a straight-forward matrix-based method for computing projections when recourse to the Kalman filter is infeasible.

Our results demonstrate the key importance of ubiquitous – but commonly unmentioned – initial value conditions on 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. In Section 4 we proceed with a motivating case study, an inventory series of food products, available as a mixed sample of quarterly and monthly frequencies. We demonstrate the modified X-11 seasonal adjustment procedure on this case, producing estimates at the full data span at the highest sampling frequency, and present our conclusions.

2 Modeling Mixed Frequency Data

2.1 Time Series Data at Multiple Sampling Frequencies

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

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

The observed data can then be written as a column vector. The exact sequencing of observations is not unique in the case that there are multiple observations at a given time. For example, suppose that a flow series is observed at quarterly and monthly frequencies for ten years. The March time point will have a quarterly number – corresponding to the months of January, February, and March – as well as a monthly number. Note that this illustration is meaningless for a stock series, since the March quarterly value is identical with the March monthly value.

The key assumption of our method is that all the observations arise as a linear combination of an unobserved highest frequency data vector, which we will call $Y = (Y_1, Y_2, \cdots, Y_n)'$. This is a feasible assumption, because we assume that our mixed sample is either a stock or a flow. Lower frequency stock values are obtained from the highest frequency stock value by an identity mapping. But for flows, the lower frequency values are obtained by aggregation of the highest frequency. Hence the observed data takes the form JY, where J is a selection matrix that is described below.

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

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

2.2 Model Estimation

So we can write the observed data as $X = JY$ with J given by ones and zeros, where $\{Y_t\}$ is described above. The high-frequency vector Y has length n, but the observed data has length m, so J is $m \times n$. For our method, we must have at least d observations available on the high-frequency portion of the data, which may come at arbitrary points in time. Let us write our observed data into a column vector X such that any d of these high-frequency observations occur as the first d components of X. It is immediate that the first d rows of J have the form $[1_d \, 0] \, P$ for some $n \times n$ dimensional permutation matrix P. This matrix maps the corresponding components of Y into the first d slots of X. Assuming for now that Y has mean zero – the case of nonzero mean is treated in the next subsection – the log Gaussian likelihood multiplied by -2 is given by

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

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

However, we may guess that a differencing matrix does exist that reduces X to stationarity, by the following reasoning. Nonstationary effects include deterministic (e.g., polynomial functions) as well as stochastic phenomena; the former will be annihilated by differencing, whereas the latter are reduced to a mean zero covariance stationary process. Hence if we can identify a generalized differencing matrix that reduces the deterministic portion to zero, we may suppose that it will also reduce the stochastic portion to stationarity. If we have n observations – possibly at irregular, arbitrary time points – of d linearly independent time-varying functions, there exists a $n-d$ by n dimensional matrix that annihilates all of the corresponding column vectors. This matrix will be the generalized differencing matrix; we next proceed to describe its calculation.

Now since Y is the highest frequency (unobserved) data vector, it is differenced to stationarity by a $n-d \times n$ dimensional matrix Δ:

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

Then $W = \Delta Y$ is mean zero and has a Toeplitz covariance matrix, since $\{W_t\}$ is covariance stationary. Augment Δ with the first d rows of the permutation matrix P, and call this $\widetilde{\Delta}$:

$$(2) \quad \widetilde{\Delta} = \begin{bmatrix} [1_d \, 0] \, P \\ \Delta \end{bmatrix}.$$

Then applying $\widetilde{\Delta}$ to Y yields a vector with the first d components given by Y_*, followed by W, where $Y_* = [1_d \, 0]PY$. These will be called the initial values, by analogy with the treatment of Bell (1984) and McElroy (2008).

The following result establishes invertibility of $\widetilde{\Delta}$, and describes the construction of a differencing matrix D, that when applied to X yields linear combinations of $\{W_t\}$, such that DX has an invertible covariance matrix.

Theorem 1 *Suppose that the mixed sample X can be written as $X = JY$ for a high frequency sample Y, which is difference stationary with degree d differencing polynomial $\delta(B)$. Also let P be defined such that $[1_d \, 0]P$ is the first d rows of J. Then the matrix $\widetilde{\Delta}$ defined in (2) is invertible, and*

$$(3) \quad [\underline{A} \; \underline{B}] = [0 \; 1_{m-d}] \, J\widetilde{\Delta}^{-1}$$

is well-defined. It also follows that $D = [-\underline{A} \; 1_{m-d}]$ differences X in the sense that

$$(4) \quad DX = [-\underline{A} \; 1_{m-d}]X = \underline{B}W.$$

If Y_* is uncorrelated with W and Σ_* is the covariance matrix of Y_*, then -2 times the log Gaussian likelihood can be written as

$$(5) \quad \begin{aligned} & Y_*' \Sigma_*^{-1} Y_* + \log |\Sigma_*| \\ & + (DX)' \left(\underline{B} \Sigma_W \underline{B}'\right)^{-1} DX + \log |\underline{B} \Sigma_W \underline{B}'|, \end{aligned}$$

up to irrelevant constants.

This result gives us a practical ability to compute the log likelihood. Since no parameters enter into Σ_*, it can be ignored and we focus on minimizing the quadratic form in DX. In this way the maximum likelihood estimates (mles) can be obtained.

2.3 Model Specification, Fitting, and Testing

We now provide a more nuanced discussion of model fitting for mixed frequency data. Let us first suppose that the aggregation relations between different frequencies hold exactly, as described above. It is clear that all covariance structure for lower frequencies is generated by a specified covariance structure at the higher frequencies. In other words, we may consider specifying a model for the highest frequency data, and this automatically determines an implied model for all lower frequencies.

Now if the high frequency data follows an ARMA model, the lower frequencies need not follow this same specification. They may not even be ARMA. But we still know how to compute their covariances, and in this sense their model is determined.

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

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

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

So the new design matrix for the observed mixed data is $J\mathbb{X}$. This will automatically account for elimination of unobserved effects at lower frequencies (the corresponding rows and columns of $J\mathbb{X}$ will then just be zeroed out), and the likelihood can be adjusted accordingly.

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

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

where the mles for parameters are utilized, hence the notation $\widehat{\Sigma}_W$ and $\widehat{\beta}$. If the model is correctly specified, and all mles were to converge to their asymptotic values, the above vector would be *iid* standard normal. Plotting these residuals, along with their sample ACFs, can allow one to assess model fit. If a problem is found, we can try a new model and repeat the whole procedure.

Next, we must address the problem of flow aggregation and the logarithmic transformation. We describe the problem with monthly (y_1, y_2, and y_3) and quarterly (x_1) data. In the original scale, the flow property states that $x_1 = y_1 + y_2 + y_3$, when x_1 corresponds to the quarter comprising the months described by y_1, y_2, y_3. Suppose that we wish to model the data in logarithms, this being a sensible variance-stabilizing transform. Denote the log-transformed variables with capital letters. Then $X_1 \neq Y_1 + Y_2 + Y_3$, which is a bother. Note that this problem does not arise for stocks (where $x_1 = y_3$ maps to $X_1 = Y_3$ without hindrance).

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

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

This latter term is bounded below by log(3), and in practice is close to the lower bound when y_1, y_2, y_3 have values reasonably close to one another. This analysis indicates that we might write $X \approxeq JY + \mu\iota$, where the rows of J that do flow aggregation have values of $1/3$ (instead of one in the original data scale), and $\mu = \log(3)$ and ι is a vector of zeroes and ones. The approximation is hopefully small in practice, such that one proceeds with the algorithm as if it were exact. Furthermore, if any quarter appears alongside with all its constituent months, we may strike out that quarter's row since it really imparts no additional information over the three months. Then each quarter that appears contains at most two months that are available elsewhere in the sample, so that at least one month that makes up the quarter is opaque (i.e., unknown). We might then imagine altering this one value slightly, such that $X = JY + \mu\iota$ holds exactly. Statistically, such a slightly mollified Y will follow the same model as the original Y, since there is little variation.

As for $\mu\iota$, this will be subsumed in the regression effects $J\mathbb{X}$ if a constant is already included – otherwise, it can be added in. When the regression parameter for mean effect is determined, it is equal to the real mean effect plus log(3), i.e., we should subtract out log(3). In this way we can handle flow data in logs.

3 Projection: Forecasting and Signal Extraction

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

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

We give further illustrations of projections – forecasting, imputation, signal extraction, etc. – following our main theorem. All these problems have in common that one seeks to estimate IY optimally from JY. Parzen (1961) provides a general formula for this problem, and we can get a particular solution in our context that is computable without knowledge of nuisance values.

Theorem 2 *Under the assumption that Y_* and W are uncorrelated, the formula for the optimal estimate of IY from JY is*

$$(6) \quad \widehat{IY} = I\Sigma_Y J' \Sigma_X^{-1} X = I\widetilde{\Delta}^{-1} \begin{bmatrix} X_* \\ \Sigma_W \underline{B}' (\underline{B}\Sigma_W \underline{B}')^{-1} DX \end{bmatrix},$$

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

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

Remark 1 The first formula in (6) is the general expression from Parzen (1961), but the second formula is practicable for implementation. The estimate $\widehat{IY}$ is computable from quantities appearing in the algorithm of Section 2, and hence are readily available. Since the error covariance matrix contains no nuisance values, it is readily computed from I. Its diagonal gives the mean squared error.

Remark 2 In the case that I is an identity matrix, we obtain a vector of forecasts, backcasts, and imputations for Y from X. Also in the special case that $I = J$, we immediately recover X, as is seen from the first formula of (6).

We now proceed to work through some applications. The first thing to note is that in practice I is determined first – which determines the exact length needed for Y – which in turn will determine J.

We construct I and J such that the same vector Y is featured in both Z and X. That is, if I is $r \times n$ and requires a certain span of the $\{Y_t\}$ series for its definition, some of these Y_t variables may not be featured anywhere in the available data matrix X. In that case, the corresponding column of J will have all zeroes. For example, suppose we have only a single frequency and the data is $X = [Y_1, Y_2]'$, but our target is $Z = Y_0$. Then $Y = [Y_0, Y_1, Y_2]'$ and $I = [1, 0, 0]$, whereas

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

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

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

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

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

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

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

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

4 Application to Series U11SFI

We are interested in applying these techniques to time series under production at the U.S. Census Bureau. Our 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 will generate a mixed quarterly-monthly version of it, and analyze the result. Projections from this mixed sample can then 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 will also work with a regression-adjusted version of the series, using X-12-ARIMA initially to remove an additive outlier (no other fixed effects, such as trading day or holiday, were found to exist). Although the regression estimation approach of section 2.3 could be utilized,

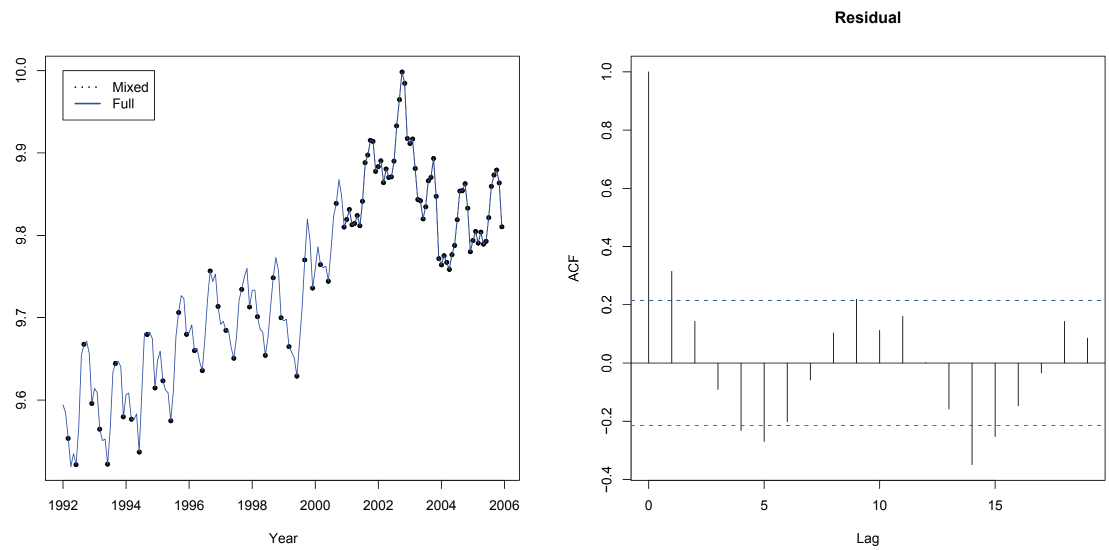

Figure 1: The left panel displays the mixed frequency sample along with the original monthly sample of *U11SFI*, whereas the right panel display the ACF of the time series residuals from the fitted mixed frequency airline model.

this would introduce extra error at the model fitting stage – to isolate the X-11 filtering aspect and facilitate comparisons with truth, we presume that all fixed effects have been satisfactorily removed.

In order to produce the mixed frequency sample, we will suppose that the last five full years of monthly observations are available, but prior to this all data is quarterly[2]. Then X consists of the last 60 monthly values, followed by all the quarterly values in order, for all but the last 5 years; see Figure 1. Now although a Box-Jenkins airline model (Box and Jenkins, 1976) – deemed best by X-12-ARIMA – fitted to the logged monthly data produced nonseasonal and seasonal moving average parameter mles of $-.08$ and .61 respectively, the same model fitted to the mixed frequency sample yielded .77 and .78 respectively. Note that less information is present in the mixed frequency sample, and moreover the sampled airline model is not optimal; future research must examine the behavior of mles via simulation studies. The time series residuals display some serial structure, as is evident from the ACF plot in Figure 1. Although the original mles for the monthly series yield time series residuals with less structure, in practice these parameters are unavailable, and one must utilize the mles derived from the mixed frequency sample.

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

A curious "stationary seasonal" pattern is evident in the non-seasonal component: there is a quasi-periodicity of 3 months, or every quarter, which does not persist over long stretches of time. A spectral estimate for this component shows residual power at the quarterly frequency, manifested

[2]We also did an analysis with only two full years of monthly data, but this seemed to be an insufficient span to generate reasonable results; cf. further discussion below.

[3]The specification in this case was a 3×3 and 3×5 for the seasonal moving averages, and a Henderson 9, all for the monthly frequency.

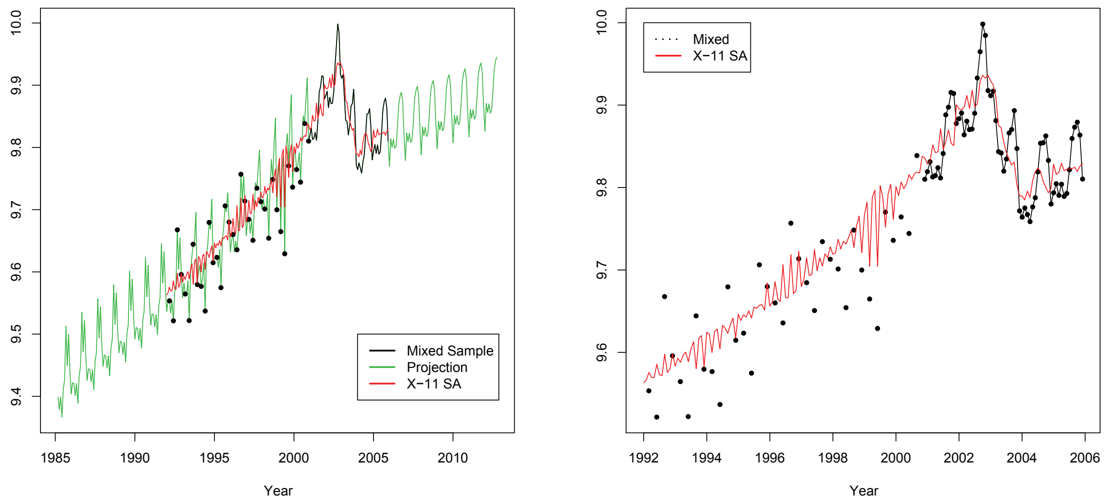

Figure 2: The left panel displays the mixed frequency sample, full projections, and seasonally adjusted data. The right panel displays the same information without the projections.

as an inverted "W"-shape – this is not uncommon, corresponding to insufficient smoothing for the quarterly frequency. Modifying the smoothing patterns of the X-11 filter did not ameliorate the situation, although better modeling of the quarterly aspect of the mixed frequency sample – such that the cyclical correlation in the right panel of Figure 1 is removed – would ensure that the projections in Theorem 2 more properly account for the data's structure, thereby wiping out the quarterly cycle in the seasonal adjustment. One can see this assertion is true by examining the segment of the seasonal adjustment corresponding to monthly values: there the quarterly cyclicality is much reduced.

MSEs can also be easily produced. We first display the projection MSEs for the monthly series in the left panel of Figure 3; 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. Finally, the right panel of Figure 3 gives the MSEs for the X-11 seasonal adjustment estimates, the asymmetric structure following the pattern for the projections. The oscillatory and step-function characteristics of these plots are a familiar feature of finite-sample MSEs for time series projections (cf. McElroy (2008)).

We also examined the same series with a monthly span of two years, rather than five years. The quarterly cycle phenomenon was much more pronounced in this case (we do not present results here). Observe that only $24 - 13 = 11$ monthly observations are available in this case, after differencing, to determine the stationary monthly time series structure – this is not much data to fit an ARIMA model. With five years we have an effective sample of 47, and the problem is greatly diminished. This suggests that our methodology is likely to produce unsatisfactory results when the span of monthly data is short, even when a substantial quantity of quarterly data is available.

REFERENCES (RÉFERENCES)

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

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

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

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.

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

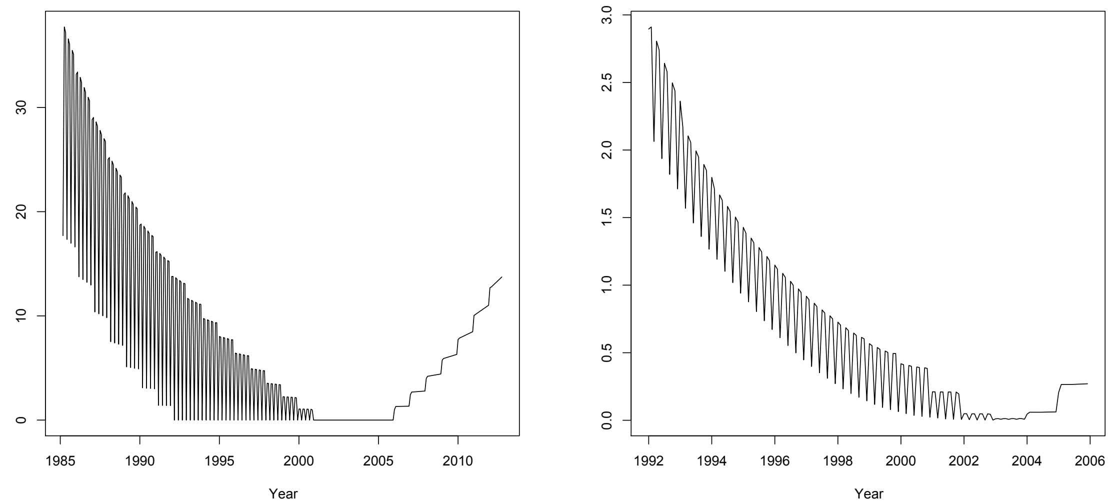

Figure 3: The left panel displays the MSEs for the full projections, whereas the right panel displays the MSEs for the mixed frequency seasonal adjustment.

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

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

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

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

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

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

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

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

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

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.

RÉSUMÉ (ABSTRACT)

An important practical problem for statistical agencies 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; for example, a series published at a quarterly frequency in the past might transition to a monthly publication schedule. We demonstrate that the X-12-ARIMA method can be applied to such mixed frequency data, by deriving formulas for conditional expectations that can be applied to forecasting problems. Moreover, standard errors are easily produced for this method as a by-product of the calculations. The method is illustrated on a mixed frequency time series.