Issues Related to the Modeling and Adjustment of High Frequency Time Series

Tucker S. McElroy* Brian C. Monsell†

Abstract

This paper provides analyses of multivariate daily immigration data, separating the long-term trend from annual and weekly patterns of seasonality via the use of so-called canonical models in an unobserved components framework. These canonical models are as stable as possible, having the maximal amount of white noise already removed, resulting in a less variable stochastic component. To further separate trend and annual seasonality, we employ a Hodrick-Prescott (HP) filter to the combined component, and using an implied models framework determine the uncertainty. To surmount the computational challenges, we implement forecast extension together with application of a bi-infinite filter composed of the HP and model-based aspects. Parameter estimates are obtained by a simple method-of-moments estimator, which is modified to ensure that no spurious co-integration effects are present in the model. These methods are demonstrated on the six high frequency immigration series, successfully decomposing the data into diverse components that each have correct (and disjoint) dynamic properties.

Key Words: Spectral analysis, signal extraction, moving holidays, outlier effects

Disclaimer

Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.

1. Introduction

A key challenge of high frequency time series data is separating low frequency content through signal extraction. For economic data that exhibits a notable trend, viewed heuristically as a frequency-zero phenomenon, the increase of the sampling frequency generates a confounding of trend with other salient patterns such as seasonality. This is because stochastic effects in a series that recur once a year have frequency proportional to the reciprocal of the period, which is increasing in the sampling frequency. For example, a seasonal effect for quarterly data has frequency $2\pi/4$; for monthly data the frequency is $2\pi/12$; for daily data the frequency is $2\pi/365$ (ignoring leap year). The frequency $2\pi/365$ is hard to distinguish numerically from zero, and hence trend and annual seasonality is easily confounded in daily time series. For an illustration, consider the six time series of New Zealand (NZ) immigration plotted in Figure 1.

This facet of high frequency data is counteracted by the abundance of sample: clearly, if 10 years of daily data are available, this should be as good or better – for the purposes of extracting trend and seasonality – as having 10 years of monthly data. However, note that the objective in the former case is more demanding than in the latter case: we now seek a daily trend, which is more difficult than producing a monthly trend. Although downsampling the high frequency data from daily to monthly would yield a time series where

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

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

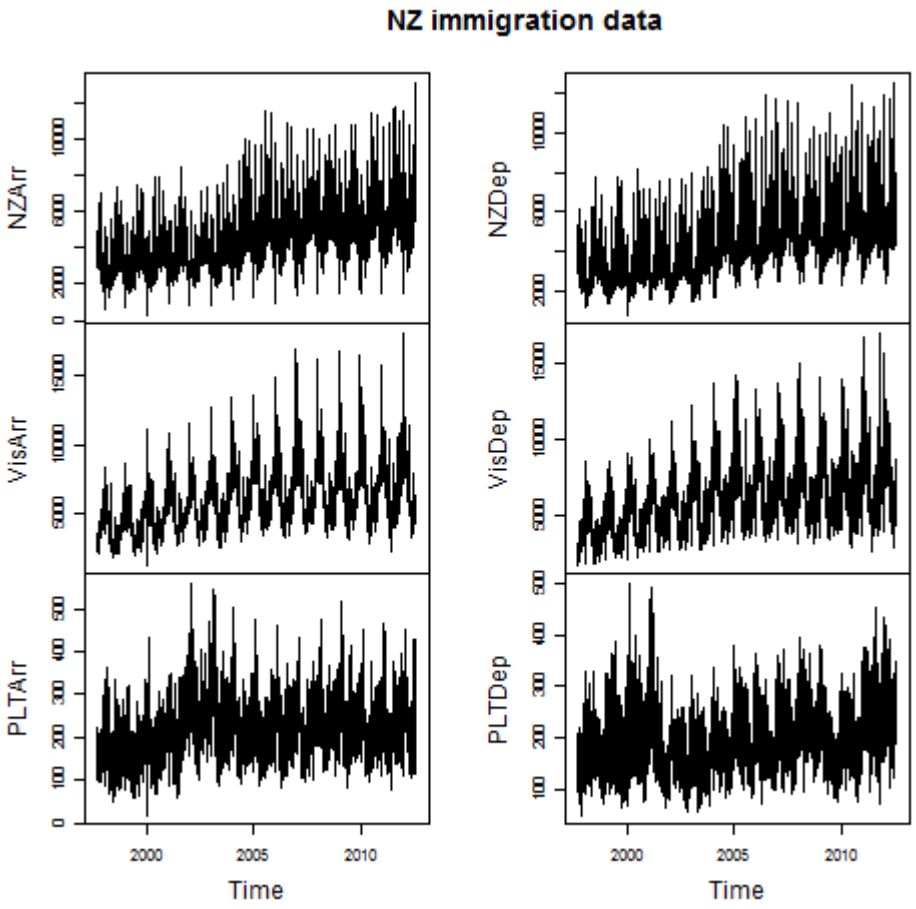

Figure 1: Six (6) New Zealand immigration series (September 1, 1997 through July 31, 2012).

there is a more obvious separation of the components in the frequency domain, filtering such – apart from doing ad hoc interpolation – is of no use in producing a daily trend. The lower frequencies of the data are closely entwined, in a way that is not present to such a degree in data that is sampled less frequently. We advocate a solution to this challenge through the concept of canonical models, together with use of the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997).

Canonical models were introduced by Hillmer and Tiao (1982), and essentially refer to stochastic time series models where the maximal amount of white noise has already been removed from the definition of the component. In other words, there is no extraneous noisiness inherent to the stochastic process (see Gómez (2001) for a canonical trend example). Having implemented such canonical models for atomic components in our custom software *sigex*, we apply these constructs to the modeling of New Zealand immigration data, wherein both a clear trend and a salient seasonal pattern are evident.

These models alone are unable to discriminate trend from annual seasonality (cycle, for short) based upon the available sample, and so we proceed by extracting a joint component, followed by application of the HP filter to separate the signal according to pass-band (trend) and stop-band (cycle). The combination of an ad hoc filter with a model-based paradigm has precedent in the univariate work of Kaiser and Maravall (2005). This is an effective means of separating content, although the signal-to-noise ratio must be set appropriately, and the filter must be combined with the model-based filter of the combined trend-cycle

component and applied to forecast-extended data. An additional challenge is determining the appropriate signal extraction uncertainty for the trend, or for more complex signals that incorporate the trend (such as the seasonal adjustment).

Another facet of our analysis is that the daily series are modeled multivariately, utilizing the general structures described in McElroy (2017), which allow for different types of seasonal co-integration among batches of series. These in turn yield informative econometric relationships with attractive interpretations, while simultaneously improving model fit and reducing signal extraction uncertainty. However, the advantages of a multivariate analysis are offset by the computational challenges associated with big data. With the NZ data, one faces the prospect of computing both the Gaussian likelihood and signal extraction estimates for six series and almost 6000 observations. In order to estimate the parameters, we utilize the method-of-moments (MOM) estimator proposed in McElroy (2017), with modifications to diminish the influence of spurious co-integration implied by reduced rank estimated covariance matrices. These modifications are very important in applications, because permitting false co-integration effects yields signal extraction filters with insufficiently strong noise suppression bands, resulting in signal leakage onto associated components.

These novel facets – canonical atomic models, separation of trend and cycle via HP filtering, and invertible MOM estimators – together yield a computational feasible decomposition of the immigration data into the three chief dynamics: trend, cycle (annual seasonality), and weekly seasonality. Along the way, we have a model-based interpretation for each component (including the trend and cycle, via the implied model paradigm of Kaiser and Maravall (2003)) that is capable of describing co-integration effects explicitly; should any component exhibit undesirable characteristics (e.g., seasonality in the trend component), the methodology can easily be tuned in an intuitive manner to yield extracted components possessing sufficient fidelity to their target process' dynamics.

Below (Section 2) we provide a description and exploratory analysis of the NZ data that motivated this work. Having had little success with basic atomic models, in our final analysis we utilized canonical atomic models, which are reviewed in Section 3. Estimation, and the invertibility modifications to the MOM estimators, is discussed in Section 4, while the implied models framework, and the choice of HP signal-to-noise ratio, is discussed in Section 5. Signal extraction is reviewed in Section 6, with details on how an ad hoc filter such as the HP can be combined with a model-based filter, and how the mean squared error of extracted components can be quantified. These techniques are applied to the data in Section 7, along with our final remarks in Section 7.3.

2. New Zealand Demographic Data

This paper studies six daily imigration series of New Zealand, covering the period September 1, 1997 through July 31, 2012. A description of each series is given below:

NZArr New Zealand residents arriving in New Zealand after an absence of less than 12 months.

NZDep New Zealand resident departures for an intended period of less than 12 months.

VisArr Overseas residents arriving in New Zealand for a stay of less than 12 months.

VisDep Overseas residents departing New Zealand after a stay of less than 12 months.

PLTArr Permanent and Long Term Arrivals includes overseas migrants who arrive in New Zealand intending to stay for a period of 12 months or more (or permanently), plus New Zealand residents returning after an absence of 12 months or more.

PLTDep Permanent and Long Term Departures includes New Zealand residents departing for an intended period of 12 months or more (or permanently), plus overseas visitors departing New Zealand after a stay of 12 months or more.

Figure 1 displays each of the series, and several patterns of seasonality are present, apart from the overall trend growth. There is an annual pattern, and secondarily (though not apparent) a weekly pattern. While a monthly pattern seems plausible, it is not present in this particular series.

In order to get a crude initial analysis of each series main features, we compute spectral density estimates via an autoregressive spetral estimator (Tiao and Tsay, 1983) and mark with vertical lines the chief frequencies of potential interest (Bell and Hillmer, 1984). In Figure 2 we display the spectral densities in units of a daily frequency; spectral peaks of a cusp-like convexity are known to correspond to periodic behavior in the process' autocorrelation function (Findley, 2005), where the period is given by 2π divided by the radians frequency of the peak. In the spectral plot, the x-axis has already been normalized, so that the ordinates correspond to the number of cycles per year (i.e., the number of times a phenomenon occurs per year). The peaks indicated by the vertical red lines correspond to $365/7$, $2 \cdot 365/7$, and $3 \cdot 365/7$, or roughly 52, 104, and 156 – hence once, twice, and thrice a week.

The green line denotes twelve cycles per year, i.e., the monthly frequency, and no such effect is apparent in the series. The annual or fundamental cycle, which corresponds to more conventional notions of seasonality, is the blue line. One of the key challenges with daily data is the entanglement of trend and fundamental seasonality. This is because the fundamental seasonal frequency is $2\pi/365$, or $.99$ degrees, which is very close to zero, the trend frequency – this poses a difficulty for likelihood evaluation and signal extraction computation.

In order to better measure the fundamental seasonality, we proceed to difference the data, which yields growth rates (versus the original levels). This will annihilate the trend peak in the spectrum, and may help to identify annual and monthly spectral peaks. Figure 3 displays the results: the weekly effects are still apparent; now the annual seasonality has been wiped out, and no monthly peak is apparent either.

Although down-sampling (or flow aggregation) allows us to separate the annual seasonality from the trend, the resulting data process would no longer be high frequency, defeating the objectives of analysis.

3. Modeling

Our basic model begins with the daily data, not paying particular attention to the day of week pertaining to each time index – this will be captured through the weekly seasonal effect. We propose a model for the N-variate series $\{y_t\}$, which has been appropriately transformed, providing an additive decomposition into independent latent components:

$$y_t = \mu_t + \xi_t + \iota_t + z_t. \tag{1}$$

See Harvey (1989) for background on this general approach to modeling time series. Here $\{\mu_t\}$ incorporates both the stochastic trend and the annual seasonality, and is denoted as the trend-cycle component. Also, $\{\iota_t\}$ is a stationary transient (or irregular), and the weekly

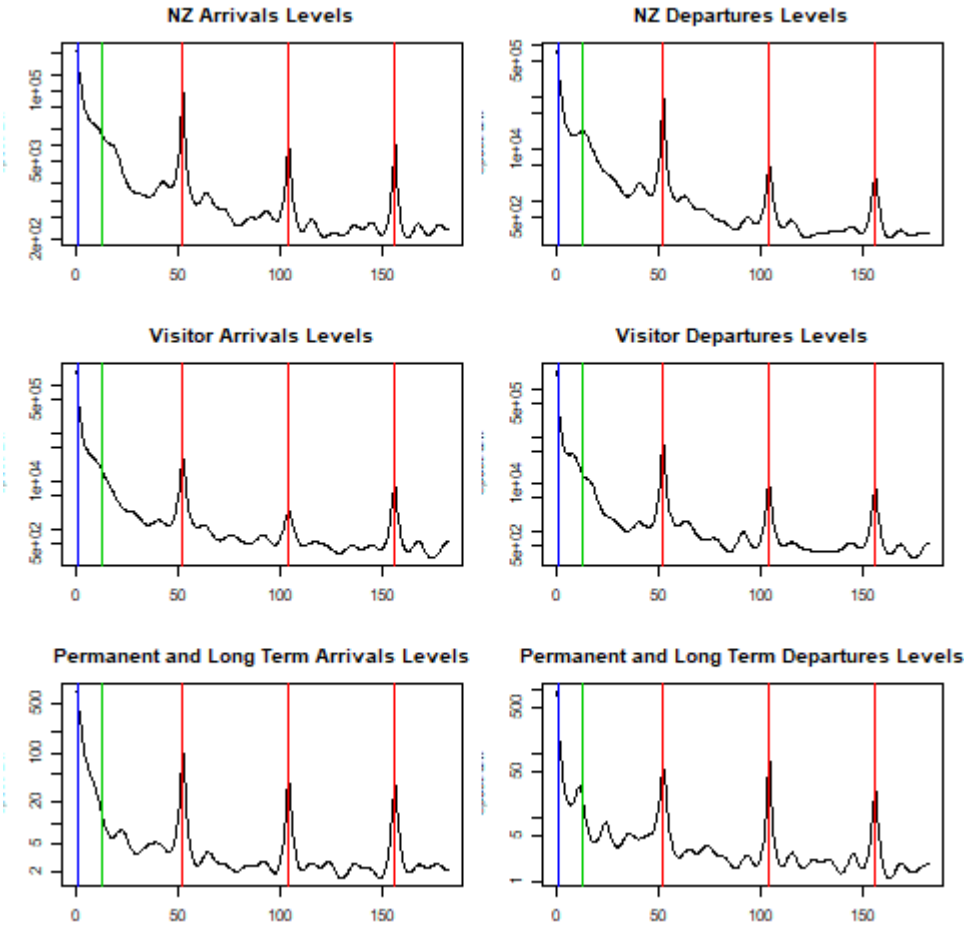

Figure 2: AR spectral density estimate of 6 New Zealand immigration series (September 1, 1997 through July 31, 2012). Vertical red lines correspond to once a week, twice a week, and thrice a week phenomena; blue line corresponds to annual phenomena, and green line corresponds to monthly phenomena.

seasonal component is $\{\xi_t\}$. We have omitted a monthly seasonal component for the retail data, as their inclusion was not really warranted by the spectral analysis. Fixed effects are incorporated through z_t, which consists of a set of regressors (e.g., moving holidays and outliers) – these may differ from series to series. Each of these latent processes involves model parameters, which will need to be estimated in a preliminary modeling stage. Conditional on the fitted model, one may proceed to forecasting and signal extraction, which will allow optimal estimation of trend, seasonal, fixed, or transient effects. The fixed effects z_t take the form

$$z_t = x_t' \beta.$$

Here x_t is a vector of r regressors, and β is the corresponding parameter. These regressors include the fixed effects, namely additive outliers (an indicator regressor) and moving holiday effects. The trend and seasonal latent processes will be supposed to be non-stationary, being defined such that differencing by a polynomial $\delta^\omega(B)$ yields a mean zero, stationary time series. Here, for any $\omega \in [0, \pi]$, we define

$$\delta^\omega(B) = 1 - 2\cos(\omega)B + B^2.$$

As a special case $\delta^0(B) = (1-B)^2$; however, we denote single differencing by $\nabla = 1 - B$. We henceforth suppose the total order d of trend differencing $(1-B)^d$ satisfies

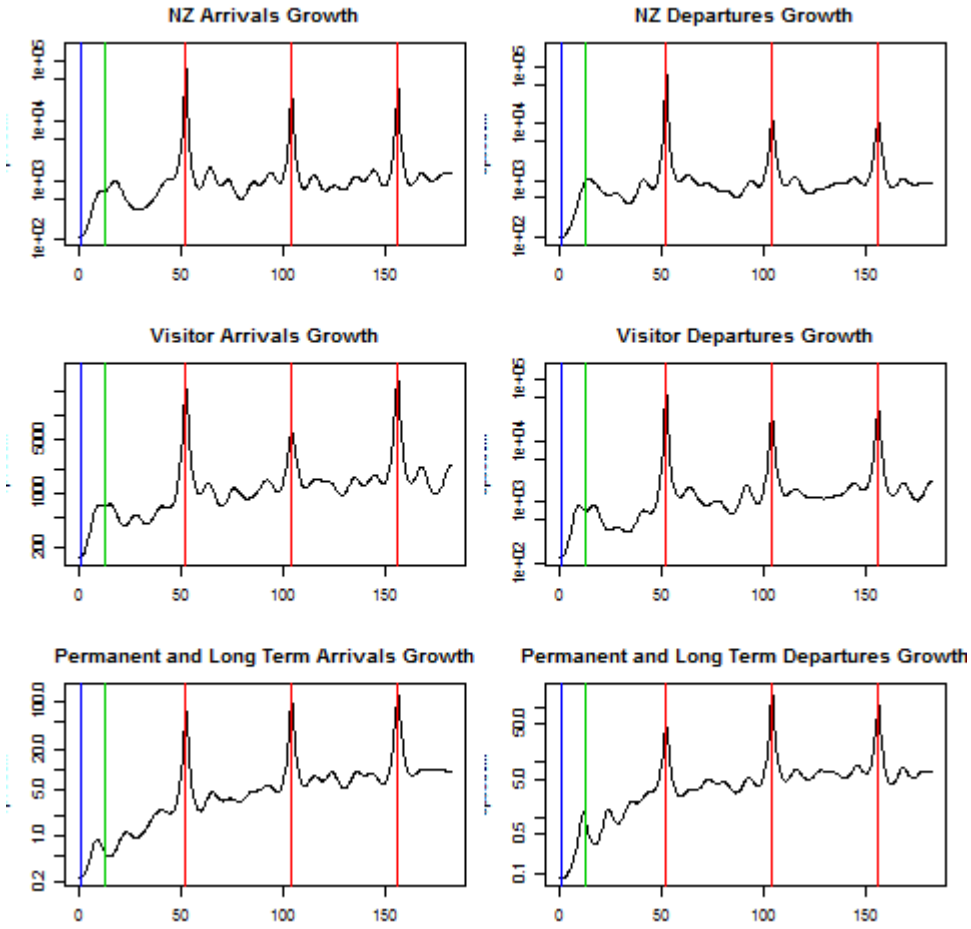

Figure 3: AR spectral density estimate of 6 New Zealand immigration series in growth rates (September 1, 1997 through July 31, 2012). Vertical red lines correspond to once a week, twice a week, and thrice a week phenomena; blue line corresponds to annual phenomena, and green line corresponds to monthly phenomena.

$d \leq 2$.

In general, the null space of $\delta^\omega(B)$ has the basis of time series $e^{\pm i\omega t}$, i.e., all time series x_t such that $\delta^\omega(B)x_t = 0$ can be written as a linear combination of the functions $e^{\pm i\omega t}$. Now the frequencies appropriate for annual and weekly effects in a daily time series are simply given by dividing the corresponding daily period into 2π, yielding $2\pi/365$ and $2\pi/7$. (If desired, leap year could be accounted for by taking the average year length over four years to be 365.25, but such subtleties make no difference to model fitting.) These are the chief harmonics, although higher multiples of such frequencies might also be considered. Combining the weekly frequency with its double and triple frequency aliases, we obtain

$$\delta^{2\pi/7}(B) \cdot \delta^{4\pi/7}(B) \cdot \delta^{6\pi/7}(B) = 1 + B + B^2 + B^3 + B^4 + B^5 + B^6 =: U(B),$$

which is verified by polynomial arithmetic. This $U(B)$ is the differencing operator for the whole weekly seasonal component, as discussed below. We use each of these differencing polynomials to define latent components, by imposing that the differenced component is a white noise process. In each case we denote the differenced component by an underscore.

Therefore we have

$$\begin{aligned}\underline{\mu}_t &= \nabla\mu_t \sim \text{WN}(0, \Sigma_\mu)\\ \underline{\xi}_t &= U(B)\xi_t \sim \text{WN}(0, \Sigma_\xi)\\ \iota_t &\sim \text{WN}(0, \Sigma_\iota).\end{aligned}$$

The innovation covariance matrices are unknown parameters – in this case there are three of them. The first component encapsulates the trend, along with any long-term movements. Because the annual frequency of $2\pi/365$ is so close to zero, such a trend component will also tend to capture annual seasonality. For this reason we refer to the component as the trend-cycle (where cycle is a short term for annual seasonality). As for weekly seasonality, the spectral plot in Figure 2 indicates that the three weekly peaks can have varying width. As argued in McElroy (2017), it is advantageous to entertain a more nuanced weekly specification, known as the atomic specification:

$$\begin{aligned}\xi_t &= \xi_t^{(1)} + \xi_t^{(2)} + \xi_t^{(3)}\\ \underline{\xi}_t^{(1)} &= \delta^{2\pi/7}(B)\xi_t^{(1)} \sim \text{WN}(0, \Sigma_1)\\ \underline{\xi}_t^{(2)} &= \delta^{4\pi/7}(B)\xi_t^{(2)} \sim \text{WN}(0, \Sigma_2)\\ \underline{\xi}_t^{(3)} &= \delta^{6\pi/7}(B)\xi_t^{(3)} \sim \text{WN}(0, \Sigma_3).\end{aligned}$$

This atomic specification yields a total of five covariance matrix parameters for the full model. In order to obtain a greater separation between components, we define them in such a way that the maximal amount of white noise is removed – this is referred to as stabilization. This specification takes the scalar portion of a component's pseudo-spectral density, namely $|\delta^\omega(e^{-i\lambda})|^{-2}$, and subtracts off its minimum value

$$c = \min_{\lambda\in[-\pi,\pi]} |\delta^\omega(e^{-i\lambda})|^{-2}.$$

Then the stabilized atomic model has pseudo-spectral density

$$\left(|\delta^\omega(e^{-i\lambda})|^{-2} - c\right)\,\Sigma.$$

This still defines a difference stationary process, but now the differenced component has a moving average structure instead of being white noise. In *sigex* the moving average polynomial is computed using spectral factorization, and the resulting stabilized atomic model is used to fit the data.

In order to fit these latent component models, it is necessary to obtained the reduced form representation for the observed process. Since all the differencing operators are distinct (i.e., the polynomials share no common roots), the minimum differencing polynomial that reduces the data to stationarity is given by their product, i.e.,

$$\delta(B) = \delta^{2\pi/365}(B)\cdot U^w(B) = \delta^{2\pi/365}(B)\cdot\delta^{2\pi/7}(B)\cdot\delta^{4\pi/7}(B)\cdot\delta^{6\pi/7}(B).$$

Applying $\delta(B)$ to (1) yields

$$\begin{aligned}\underline{y}_t = \delta(B)y_t = {}& \delta^{2\pi/365}(B)\cdot U^w(B)\mu_t\\ &+ U^w(B)\,\underline{\xi}_t^a\\ &+ \delta^{2\pi/365}(B)\,\underline{\xi}_t^w\\ &+ \delta^{2\pi/365}(B)\cdot U^w(B)\,\iota_t\\ &+ \delta^{2\pi/365}(B)\cdot U^w(B)\,z_t.\end{aligned}$$

In the case of atomic weekly seasonals, we also have the expression

$$\underline{\xi}_t^w = U^w(B)\xi_t^w = \delta^{4\pi/7}(B)\cdot\delta^{6\pi/7}(B)\,\underline{\xi}_t^{(1)} + \delta^{2\pi/7}(B)\cdot\delta^{6\pi/7}(B)\,\underline{\xi}_t^{(2)} + \delta^{2\pi/7}(B)\cdot\delta^{4\pi/7}(B)\,\underline{\xi}_t^{(3)}.$$

The autocovariance sequence of $\{\underline{y}_t\}$ is easily computed from these equations. In fact, the equation for $\underline{y}_t$ takes the form of three (five in the case of atomic weekly seasonals) independent vector moving average processes, of various orders, whose moving average polynomials are given by the various products of differencing polynomials, each being driven by independendent white noises of variances given above.

Regarding the first and last terms, the fixed effects: we simply apply the differencing operator to each component regressor of x_t – so long as this is not annihilated, the fixed effect is identifiable (otherwise, it is redundant with at least one of the components, and can safely be eliminated from the model). Let us denote this vector of differenced regressors, for the trend and non-trend effects, via

$$\underline{x}_t = \delta(B)\,x_t.$$

Then $\underline{y}_t - \underline{x}_t'\beta$ is mean zero, with autocovariance structure given by summing the autocovariances of the three (or five) latent moving average processes. Thus, the parameter vector is composed of β and the parametrizations of all of the covariance matrices, utilizing the Cholesky parametrization discussed in McElroy (2017).

4. Model Estimation

With a fitted model in hand, we can proceed to check the goodness-of-fit via examination of the time series residuals. If the model seems to be adequate, we may consider more parsimonious nested alternative models. For example, if a covariance matrix Σ has small eigenvalues, we might attempt to fit a nested model where Σ is enforced to have reduced rank. Estimation of the parameters can be accomplished in the *sigex* software via the MOM (method of moments) scheme; maximizing the Gaussian likelihood is not really feasible, due to the high dimension of the parameter space. The final parameter estimates will be denoted $\widehat{\theta}$.

The MOM estimator of McElroy (2017) involves a set of equations involving component and process spectral densities, which are solved for the various innovation covariance matrices Σ. The equations involve the autocovariance function of the differenced process $\{\underline{y}_t\}$ at various lags, and hence the parameter estimates are obtained by subsituting sample autocovariances in the same formulas. Whereas a single likelihood evaluation can take a minute or more for the data considered in this paper, the MOM calculation is complete in a few seconds. Consistency and asymptotic normality of these estimators has been established in a working paper; there is an efficiency loss as compared to MLE. Although the estimators converge in probability to the true covariance matrices, which are non-negative definite, the actual estimates can have negative eigenvalues in practice.

Because such estimates are not interpretable as covariance matrices, and are thereby useless for statistical applications such as signal extraction, it is necessary to modify the straight MOM to a non-negative definite version. This can be done by taking the maximum of each eigenvalue with zero. However, a singular covariance matrix leads to an interpretation of co-integration in the original series, as discussed in McElroy (2017), and furthermore indicates a very distinctive behavior in the signal extraction filter. McElroy and Trimbur (2015) showed that an implication of co-integration is that optimal signal extraction filters may have nonzero gain functions in the stop band. That is, in the simple

case where only a trend and irregular component is present, and supposing that the trend covariance Σ_μ is reduced rank, the noise extraction filter at frequency zero is

$$1_N - L\,(L'\,\Sigma_\iota^{-1}\,L)\,L'\,\Sigma_\iota^{-1},$$

where L is the lower triangular Cholesky factor of Σ_μ. Now when Σ_μ is full rank, the noise extraction filter at frequency zero is equal to the zero matrix, which corresponds to the full trend suppression needed to estimate the irregular; but in the case of a co-integrated trend – which corresponds to a reduced rank Σ_μ – the frequency response function can be nonzero. If the data really are not really co-integrated the noise suppression will be inadequate, and long-term movements will leach into the extracted irregular.

This is only a simple illustration of a more general phenomenon. In our case of five components, an unwarranted co-integration effect (at frequency zero, or at one of the weekly seasonal frequencies) will result in the inadequate suppression of that particular component when it is featured as noise – for instance, unwarranted co-integration at frequency $4\pi/7$ will generate a problem for extracting either the trend-cycle, the first seasonal, the third seasonal, or the irregular (or combinations thereof), with dynamics of the second weekly seasonal leaching out. For this reason, great care is needed to ensure that no co-integration effects are imputed to the model unless they are truly warranted. Simply thresholding the eigenvalues of a MOM estimate to zero can generate just such a difficulty, as the resulting estimate would have reduced rank.

To rectify this problem, we describe a procedure that modifies a given MOM estimate to a close positive definite covariance matrix, by modifying its correlations while leaving its variances intact. The Generalized Cholesky Decomposition (GCD) of a covariance matrix is discussed in Golub and Van Loan (1996). Also, condition numbers are defined in McElroy (2017) that describe near-singularity in a scale-free way: they are the logged diagonal entries in the GCD of the associated correlation matrix. Let a j-dimensional covariance matrix be denoted Σ_j, which can be decomposed as $\Sigma_j = L_j\,D_j\,L_j'$ for a unit lower-triangular L_j and a diagonal matrix D_j. Letting the lower row of L_j be denoted $[\ell_j', 1]$, it is known that

$$g_{j+1} = d_{j+1} + \ell_{j+1}'\,D_j\,\ell_{j+1},$$

where g_{j+1} denotes the lower right entry of Σ_{j+1}. Moreover, the $j+1$th condition number can be expressed as

$$\tau_{j+1} = -\log\left(g_{j+1}/d_{j+1}\right) = -\log\left(1 + \ell_{j+1}'\,D_j\,\ell_{j+1}/d_{j+1}\right).$$

These are recursive relations, so that for a given N-dimensional Σ we can examine the upper left submatrix Σ_j for $1 \leq j \leq N$ and recursively compute the condition numbers τ_j. Now $\tau_j = -\infty$ corresponds to singularity (it is shown in McElroy (2017) that certain partial correlations are equal to ± 1 in this case); hence, if we wish to modify a given Σ such that the condition numbers exceed some minimum threshold, then we must modify either ℓ_{j+1} or D_j. Whereas the former is related to linkages between variables, the each d_j is a partial variance (of the jth variable given the preceding variables 1 through $j-1$); choosing to modify these to some new $\widetilde{d}_j$, we see how the condition number changes accordingly. Thus, if we set τ_{j+1} equal to a desired threshold α, and given that we have already modified in a recursive fashion the preceding D_j to $\widetilde{D}_j$, we have the formula

$$\widetilde{d}_{j+1} = \left(\frac{\exp\{-\alpha\} - 1}{\ell_{j+1}'\,\widetilde{D}_j\,\ell_{j+1}}\right)^{-1}. \tag{2}$$

This suggests the following algorithm to produce a positive definite modification of a given Σ matrix: compute the GCD, and iteratively check whether $1/d_{j+1} < (\exp\{-\alpha\} - 1)/\ell'_{j+1}\,\widetilde{D}_j\,\ell_{j+1}$; if so, we replace d_{j+1} by $\widetilde{d}_{j+1}$ (2) and increment j. This yields a new

$$\widetilde{\Sigma} = L\,\widetilde{D}\,L',$$

where all the condition numbers are bounded below by α. In practice, the thresholds for α can be set according to the criteria discussed in McElroy (2017). For instance, $\alpha = -1.66$ corresponds to a partial correlation of .90; this is sufficiently distinct from unity, so as to ensure there is no signal leakage.

5. Implied Models from an Ad Hoc Filter

For model-based signal extraction an optimal extraction (i.e., minimal Mean Squared Error (MSE) among all linear estimators) of signal $\{s_t\}$ from $\{y_t\}$ is given by a filter $\Psi(B)$ with frequency response function

$$\Psi(e^{-i\lambda}) = f_s(\lambda)\,f_y(\lambda)^{-1},$$

where f_s and f_y are the pseudo-spectral densities of the signal and the data process. (See McElroy and Trimbur (2015) for details.) That is, $\widehat{s}_t = \Psi(B)y_t$ is the extraction and equals $\mathbb{E}[s_t|\{y_t\}_{t\in\mathbb{Z}}]$ when the process is Gaussian. This is known as the Wiener-Kolmogorov estimator. In contrast, we may consider *ad hoc* filters such as the Hodrick-Prescott (HP) filter. The HP filter has frequency response function (frf) given by

$$H(e^{-i\lambda}) = \frac{q}{q + (2 - 2\,\cos(\lambda))^2},$$

where $q > 0$ is the signal-to-noise ratio (snr). Taking q larger makes the filter resemble the identity, whereas low values of q indicate that more smoothing is necessary, and the frf has a low-pass shape. Clearly at frequency zero the frf equals one, and the value of λ yielding $H(e^{-i\lambda}) = 1/2$ is $\arccos(1 - \sqrt{q}/2)$. This suggests a relationship between the snr and the cutoff between low-pass and high-pass bands; see Harvey and Trimbur (2003). For example, to capture effects of a ten-year period in quarterly data, we can set $2\pi/\lambda = 40$ and obtain $q^{-1} = 1649$; Hodrick and Prescott (1997) advocated $q = 1/1600$ for quarterly data, which yields a cutoff period of 39.7 quarters.

However, this rule for separating phenomena still allows some leakage of signal effects into the noise; in order to screen the signal more completely from the noise, we should insist upon more damping at the cutoff frequency, i.e., determine λ according to the rule

$$\kappa = H(e^{-i\lambda}) \qquad \lambda = \arccos\left[1 + (1 - 1/\kappa)\sqrt{q}/2\right].$$

By setting $\kappa = .1$, we impose that only 10% of the variation at the cutoff frequency (and even less variation to the left of the cutoff frequency) seeps into the noise. In cases where the cutoff frequency λ is determined by a desired period $2\pi/\lambda$ that we take as given, the corresponding snr is

$$q = 4\,\frac{(1 - \cos(\lambda))^2}{(1 - 1/\kappa)^2}.$$

The trend-cycle component will now be decomposed into proper trend $\{\omega_t\}$ and cycle (annual seasonal) $\{\psi_t\}$:

$$\mu_t = \omega_t + \psi_t.$$

Letting f_μ denote the pseudo-spectral density of the trend-cycle, it follows from results in Bell (1984) that HP filtering yields a low-pass trend (from the pass-band) and a high-pass cycle (from the stop-band), with pseudo-spectral densities denoted

$$f_\omega(\lambda) = H(e^{-i\lambda})\, f_\mu(\lambda) \tag{3}$$

$$f_\psi(\lambda) = \left[1 - H(e^{-i\lambda})\right]\, f_\mu(\lambda). \tag{4}$$

(These expressions involve the product of a scalar function with a matrix-valued f_μ.) In other words, the outputs $H(B)\mu_t$ and $[1-H(B)]\mu_t$ of the HP low-pass and high-pass filters applied to the trend-cycle component are declared to be the trend and cycle components, by definition. From another viewpoint, if the trend and cycle components happened to have pseudo-spectral densities defined by (3) and (4), then $H(e^{-i\lambda})$ would be the Wiener-Kolmogorov (WK) filter for extracting trend from trend-cycle, and the HP filter would be optimal. However, $\{\mu_t\}$ is not observable, and so estimation of the trend would proceed by computing $H(B)\widehat{\mu}_t$, which is also optimal.

Proposition 1 *Suppose that the WK filter for extracting $\{\mu_t\}$ from $\{y_t\}$ is $\Psi(B)$, and the trend and cycle pseudo-spectra are defined via (3) and (4). Then $H(B)\widehat{\mu}_t$ is the minimum MSE linear estimator of ω_t from $\{y_t\}$, and $[1 - H(B)]\widehat{\mu}_t$ is the minimum MSE linear estimator of ψ_t from $\{y_t\}$.*

Proof of Proposition 1. The trend extraction error is

$$\omega_t - H(B)\,\Psi(B)\,y_t = (\omega_t - H(B)\,\mu_t) + H(B)\,\left[\mu_t - \Psi(B)\,y_t\right].$$

The second term is orthogonal to all linear functions of the data, by assumed optimality of the WK filter $\Psi(B)$. The first term is orthogonal to all linear functions of $\{\mu_t\}$ by the optimality of the HP (it is a WK filter for the trend from trend-cycle); the error $\omega_t - H(B)\,\mu_t$ only involves linear combinations of differenced trend and cycle processes $\{\underline{\omega}_t\}$ and $\{\psi_t\}$, which by assumption are orthogonal to the differences of each of the other non-trend-cycle components. Hence, using Assumption A of Bell (1984), the trend-cycle error is orthogonal to $\{y_t\}$ as well. □

For computation, one could first obtain $\widehat{\mu}_t$ for $1 \le t \le T$ as described in the next section, and then apply the HP filter. However, in order to obtain extractions at the sample boundary it is necessary to extend the calculation of $\widehat{\mu}_t$ to $1 - L \le t \le T + L$ for some horizon $L > 0$, where the HP filter is truncated to length $2L + 1$. Alternatively, the filter coefficients of the composite filter $H(B)\,\Psi(B)$ can be directly computed from the frf $H(e^{-i\lambda})\,\Psi(e^{-i\lambda})$, so that extractions are computed with a single pass of the filter on the extended data. Similar considerations apply to calculation of the cycle, although it is typically important that the cycle be a stationary process; this is handled by cancelling zeroes in the numerator of

$$1 - H(e^{-i\lambda}) = \frac{(2 - 2\,\cos(\lambda))^2}{q + (2 - 2\,\cos(\lambda))^2}$$

with any zeroes at frequency zero in the pseudo-spectral density f_μ. In the implementation discussed in Section 3, $f_\mu(\lambda) = f_{\underline{\mu}}(\lambda)\,(2 - 2\,\cos(\lambda))^{-d}$ for $d \le 2$ (we take $d = 1$ in our empirical analyses below), where $f_{\underline{\mu}}$ is the bounded spectral density of the differenced trend-cycle; hence

$$[1 - H(e^{-i\lambda})]\,\Psi(e^{-i\lambda}) = \frac{(2 - 2\,\cos(\lambda))^{2-d}}{q + (2 - 2\,\cos(\lambda))^2}\, f_{\underline{\mu}}(\lambda)\, f_{\underline{y}}(\lambda)^{-1}\, |U(e^{-i\lambda})|^2. \tag{5}$$

We can also compute the signal extraction error spectral densities, but these will actually follow from a more general result given below. The error spectral densities are useful for computing correlations of signal extraction errors, and the integral over $[-\pi, \pi]$ and normalized by 2π equals the signal extraction MSE.

Consider a component $\{\eta_t\}$ distinct from $\{\mu_t\}$, such that we wish to extract the mixed signal $\omega_t + \eta_t$. It follows that there must exist an $\{\epsilon_t\}$ (although it can be identically zero) such that

$$y_t = \mu_t + \eta_t + \epsilon_t$$

is a decomposition into orthogonal components. For example, $\{\eta_t\}$ might correspond to the irregular, so that $\{\omega_t + \eta_t\}$ is the seasonal-adjustment (i.e., the trend-irregular) and $\epsilon_t = \xi_t^w$. Whereas the extraction is given by $\widehat{\omega}_t + \widehat{\eta}_t$, the error spectral density can be computed in terms of the HP frf and the error spectra for the other components. In particular, letting g_η, g_μ, and g_ϵ denote the error spectral densities for components $\{\eta_t\}$, $\{\mu_t\}$, and $\{\epsilon_t\}$, it follows from McElroy and Trimbur (2015) that

$$g_\eta = f_\eta \, {f_y}^{-1} [f_\mu + f_\epsilon] \qquad g_\mu = f_\mu \, {f_y}^{-1} [f_\eta + f_\epsilon] \qquad g_\epsilon = f_\epsilon \, {f_y}^{-1} [f_\eta + f_\mu].$$

Because these are also the error spectral densities for the corresponding noise process in each case ($\mu + \epsilon$, $\eta + \epsilon$, and $\mu + \eta$ respectively), they are Hermitian and can be written with the first and third pseudo-spectral density swapped. In the next result, we let H and $1 - H$ denote as a short hand the frf of the HP low-pass and high-pass filters, as functions of λ.

Proposition 2 *Suppose that the trend and cycle pseudo-spectra are defined via (3) and (4), and the order of trend-cycle differencing is $d \leq 2$. Then the error spectral densities for the mixed signals are*

$$g_{\omega+\eta} = H\,(1-H)\,f_\mu - H\,(1-H)\,g_\mu + (1-H)\,g_\eta + H\,g_\epsilon$$
$$g_{\psi+\eta} = H\,(1-H)\,f_\mu - H\,(1-H)\,g_\mu + H\,g_\eta + (1-H)\,g_\epsilon.$$

Proof of Proposition 2. From McElroy and Trimbur (2015) and (3) and (4) we obtain

$$\begin{aligned}
g_{\omega+\eta} &= [f_\omega + f_\eta]\, {f_y}^{-1} [f_\psi + f_\epsilon] \\
&= [H\, f_\mu + f_\eta]\, {f_y}^{-1} [(1-H)\, f_\mu + f_\epsilon] \\
&= H\,(1-H)\, f_\mu\, {f_y}^{-1}\, f_\mu + (1-H)\, f_\eta\, {f_y}^{-1}\, f_\mu + H\, f_\mu\, {f_y}^{-1}\, f_\epsilon + f_\eta\, {f_y}^{-1}\, f_\epsilon \\
&= H\,(1-H)\, f_\mu - H\,(1-H)\, f_\mu\, {f_y}^{-1} [f_\eta + f_\epsilon] + (1-H)\, f_\eta\, {f_y}^{-1} [f_\mu + f_\epsilon] \\
&+ [1 - (1-H)]\, f_\eta\, {f_y}^{-1}\, f_\epsilon + H\, f_\mu\, {f_y}^{-1}\, f_\epsilon,
\end{aligned}$$

which simplifies to the stated formula (because $[f_\mu + f_\eta]\, {f_y}^{-1}\, f_\epsilon = g_\epsilon$). The expression for $g_{\psi+\eta}$ is similarly derived. □

Remark 1 Whereas the direct expression for the error spectrum $g_{\omega+\eta}$ is in theory available – as described in the proof – it is more convenient from the standpoint of software to have the formulas of Proposition 2, which allow us to compute in terms of g_μ, g_η, and g_ϵ; the product $(1-H)\, f_\mu$ can be computed using (5).

6. Signal Extraction

Once the modeling is complete, we can proceed towards signal extraction (Bell, 1984), which is concerned with extracting (estimating) the components of interest. Actually, in the case of fixed components there is nothing further to do, because we merely take

$$\widehat{z}_t = \underline{x}_t' \, \widehat{\underline{\beta}}.$$

But the stochastic components are more subtle to estimate. Note that this reflects a frequentist philosophy; a Bayesian analyst would weave model selection, parameter estimation, and signal extraction into a seamless garment. Although arguably less elegant, the frequentist approach is computationally simpler and still provides correct and relevant results.

The matrix formulas for signal extraction (McElroy and Trimbur, 2015) implemented in *sigex* are not feasible for large data sets. The full data vector for the six series has length 5448, which makes likelihood evaluation cumbersome and direct signal extraction infeasible, as the matrix formulas require inversion of a matrix of dimension 32688. In this paper we pursue an exact calculation of likelihood and signal extraction under Assumption A of Bell (1984), using MOM parameter estimates.

We here propose an alternative methodology to direct matrix formulas. It is known (McElroy and Monsell, 2015) that signal extraction in a finite sample can always be accomplished by applying the Wiener-Kolmogorov (WK) filter to the forecast-extended series, utilizing the same fitted model to do so. The WK filter is applicable to a full time series, and is described in McElroy and Trimbur (2015). This method has been implemented in *sigex*, proceeding with the calculation of the WK frequency response function – derived from the spectral densities of the component models – and computing the filter coefficients by integration. In this case, each filter coefficient is a 6×6 matrix. Forecasting and backcasting is straightforward, because the reduced form of the model has already been computed, and the recursive one-step ahead predictors lay at the heart of the Durbin-Levinson (D-L) algorithm.

In terms of speed, there is some cost associated with calculation of forecasts and backcasts, as it amounts to an evaluation of the Durbin-Levinson algorithm on an expanded sample size. (So if the sample size is $T = 5448$, and $H = 500$ forecasts are desired, then D-L must be run for sample size $T + H = 5948$.) There is also the cost associated with computing the WK filter, which in turn is contingent on the desired resolution of the frequency grid utilized for the frequency response function; for our applications, we chose 7000 grid points. Once the series has been extended, application of the WK filter via a convolution operation is extremely fast, requiring a negligible portion of time relative to the other calculations. In summary, the multivariate signal extraction can be feasibly calculated, with approximation error made as small as desired (by increasing L) while potentially increasing computation time.

To quantify the signal extraction uncertainty we compute the WK MSE, or the normalized integral of the error spectral density. (These calculations omit the extra MSE at the beginning and end of the sample due to backcast and forecast extension of the time series.) This is rapidly calculated, and in the figures each extracted signal is shaded with a band of two standard errors, defined as the square root of the WK MSE.

7. Results

We fit the mulitvariate model described in (1) to the New Zealand immigration series. In addition to that model, we are interested in estimating moving holiday effects for Easter and the first three school holidays for public schools in New Zealand.

The general methodology used to fit the multivariate model for the New Zealand immigration series is given below:

- Specify model, regressors for the six series.
- Perform MLE-estimation of fixed effects for the univariate series.
- Using t-statistics, determine which fixed effects will be estimated in the final model.
- Re-estimate final fixed effects for the univariate series using MLE-estimation.
- Remove fixed effects from the series.
- Perform MOM-estimation for the multivariate model on the series with fixed effects removed.
- Use WK filters to generate components, standard errors for the components.
- Use HP filter to separate trend-annual component into trend, fundamental seasonal components.

7.1 Moving Holidays and Y2K Regressors

For all of the moving holidays, the window of days around the specific holiday was kept small to capture a very local effect for the holiday. For example, the Easter regressor is set to one for the day before and the day of Easter for each year; every other observation is set to zero. Then, the Easter regressor is centered by removing the daily mean of the regressor over the entire series.

For the school holidays, an effect is estimated for the start and end of the holiday for the first three school vacations in each year. The school holiday regressor for the first day of the vacation period is set to one for that day and one for the day after, zero otherwise. These regressors will be referred to henceforth as `school1`, `school2`, and `school3` for the first, second, and third school holiday, respectively. The school holiday regressor for the last day of the vacation period is set to one for that day and one for the day before, zero otherwise. These regressors will be referred to henceforth as `school1e`, `school2e`, and `school3e` for the first, second, and third school holiday, respectively. Again, the school holiday regressors are centered by removing the daily mean of each of the regressors over the entire series.

Another regressor added for each series is a point outlier for January 1, 2000. Plots of the log of each of the series (Figure 4) show that for many of the series, there is a large decline on that day. A regressor was added to each of the series with a one for the day of January 1, 2000, zero otherwise. We'll refer to this regressor as the `AO(Y2K)` regressor.

In the first round of MLE estimation, all of the regressors described above were fit to each of the series univariately. Table 1 gives the t-statistics for the moving holidays and AO regressors. The statistics in bold represent those regressors that were found significant, and included in the final model for each series.

The significant fixed effects were removed from the series to prepare for the estimation of the multivariate model. Plotting the logged series adjusted for fixed effects showed additional outliers around January 1, 2000 for most of the series. For these series, AO regressors were included in the model and re-estimated; those regressors with significant t-statistics were included in the final model.

Figure 5 shows the AO adjustment factors related to observations around January 1, 2000. Note that the effects are quite large, and there are two AO regressors for all the series except for NZ Arrivals.

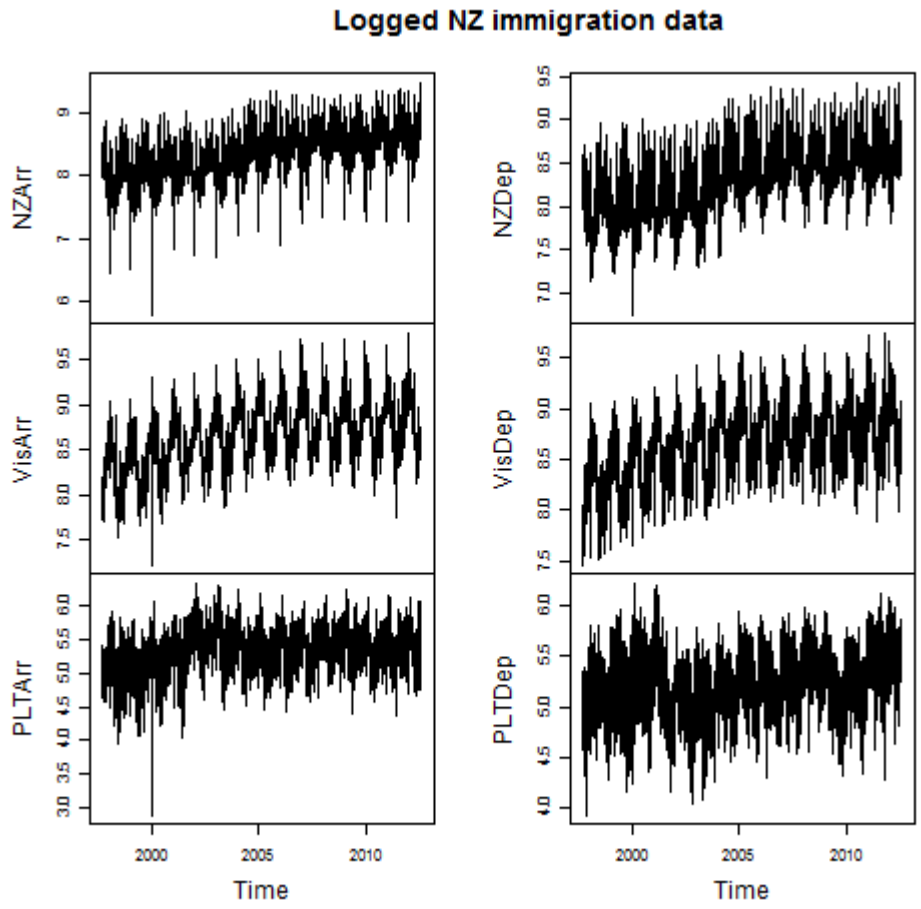

Figure 4: Log of six (6) New Zealand immigration series (September 1, 1997 through July 31, 2012). Point Outliers are evident in many series near January 1, 2000.

Figure 6 shows the Easter factors for the five series with significant Easter effects. Note that for most of the series, Easter by and large depresses activity around the Easter holiday, except for Permanent and Long Term Arrivals.

Figure 7 shows the combined effect of the moving holidays for the first school holiday regressors.

7.2 Signal Extraction Results

After the final fixed effects are removed from the series, the multivariate model is fit using MOM estimation. Then Wiener-Kolmogorov (WK) filters for the each of the components are derived, and the WK filter for each component is applied to the forecast and backcast extended series (transformed with fixed effects removed). The HP filter is then applied to the combined trend-seasonal component to generate a trend component and a fundamental seasonal component. From this a seasonally adjusted series can be generated; the seasonal adjustment generated by the multivariate model includes the fundamental seasonal bound with the trend. Standard errors are generated for each of the components.

A series of plots shows the components generated for one of the series, New Zealand Arrivals (NZArr). Figure 8 shows the seasonally adjusted series (in blue) against the original series (in grey). The uncertainty of the seasonally adjusted series is shown by the blue shading in the plot. The seasonal adjustment is quite smooth for this series, and the fun-

Table 1: **t-statistics for Holidays, AO for NZ Immigration Series**

Holiday	NZArr	NZDep	VisArr	VisDep	PLTArr	PLTDep
Easter	**-12.12**	**-4.63**	**-4.88**	**-5.03**	**2.45**	0.73
School1	**-6.26**	**4.22**	**6.88**	**-6.62**	**-2.97**	**-6.30**
School1e	**4.25**	**-4.54**	**-7.46**	2.32	**-4.80**	**-3.26**
School2	-0.12	**4.44**	**3.17**	**-3.68**	-0.89	**-5.97**
School2e	0.99	-1.82	**-6.64**	**4.24**	**-7.43**	-0.86
School3	-0.59	**3.24**	1.67	**-3.77**	-1.06	**-4.88**
School3e	1.68	**-3.21**	**-7.54**	**2.83**	**-8.03**	-1.98
AO(Y2K)	**-17.38**	**-3.50**	**-15.02**	**-3.91**	**-12.35**	**-5.12**

damental seasonality is not reflected in the seasonal adjustment. The AO(Y2K) outlier is apparent in the seasonally adjusted series.

Figure 9 shows the seasonally adjusted for the NZArr series generated from the transformed series with the fixed effects removed and another seasonally adjusted for the NZArr series generated from the transformed series without the fixed effects removed. Note that a more fully realized AO effect is contained in the adjustment done without the fixed effects; the AO effect for the adjustment without adjustment for fixed effects shows that some of the outlier effect is contained in either the fundamental or weekly seasonal effects.

Figure 10 shows the trend series (in orange) against the original series (in grey) for NZArr. The uncertainty in the trend is indicated with orange shading. The fundamental seasonality (in green) and the combined weekly effects (in purple) are also shown. Visually the fundamental seasonality is seen to be slightly larger than the combined weekly effects.

Figure 11 shows the original series (in black) and the individual weekly effects (in purple) for NZArr. Visually the first weekly component (for once a week) is larger than the other two weekly effects. Figure 12 shows the original series (in black) and the individual moving holiday effects (Easter in purple, `school1` and `school1e` in green), along with the AO(Y2K) effect (in blue).

One criteria for judging a seasonal adjustment is to look for residual effects - is there residual seasonal or weekly variation? One way of looking for this is to use spectral plots, as in Figure 2.

Figure 13 gives AR spectra for the seasonal adjustment of all six New Zealand Immigration series. Two adjustments are shown in the plot: one an adjustment from the original decomposition containing the fundamental seasonality (black line), and another the seasonal adjustment derived after the Hodrick-Prescott filter is used to estimate a trend component and the fundamental seasonality is isolated and removed (orange line). While there are troughs at the weekly frequencies for most of the series and the spectrum for the seasonal adjustment without the fundamental seasonality component is generally lower, one sign of trouble is the small peak at the thrice weekly adjustments for the Permanent and Long Term Departures series.

7.3 Future Work

We continue to look at different models and procedures for separating the long term trend from the fundamental seasonality. The size of the uncertainty bounds may lead to using a different estimation procedure for the multivariate model.

Another area of future work is how to adjust for outliers and extremes. Trying to test each observation, as is done in X-13ARIMA-SEATS, has the potential to be extremely time consuming. We will look at the method proposed in McElroy and Penny (2016), which uses

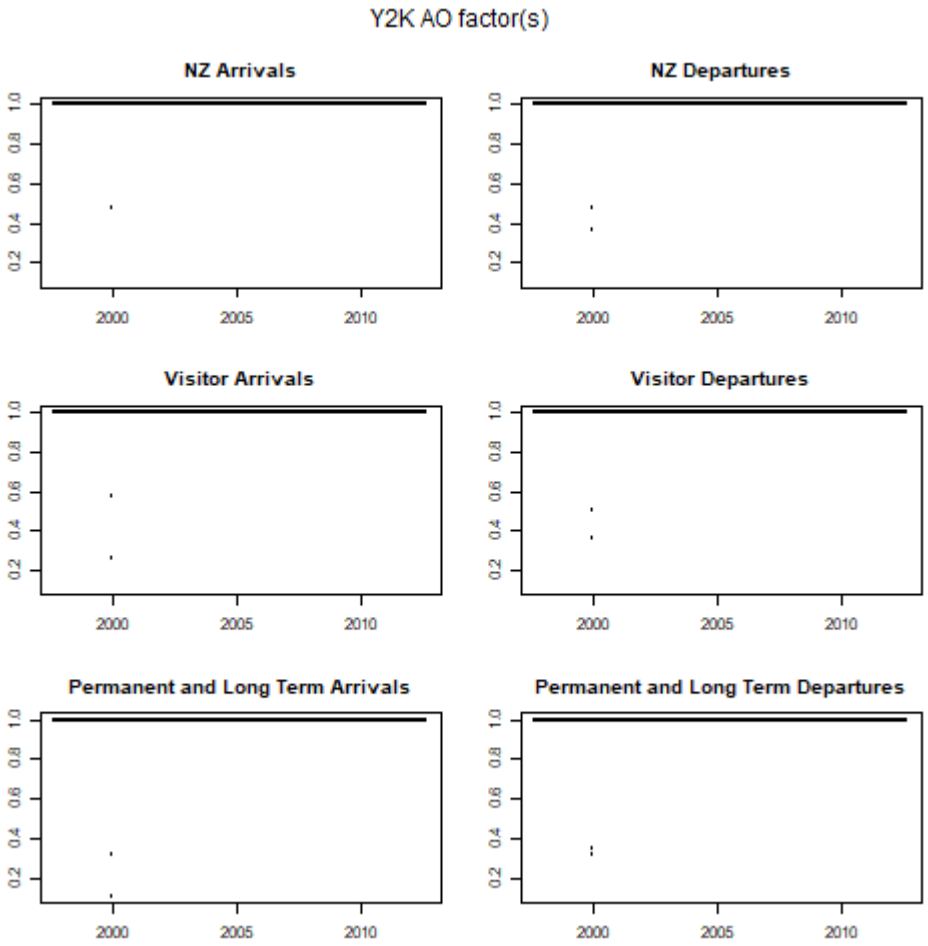

Figure 5: AO(Y2K) factors for six (6) New Zealand immigration series (September 1, 1997 through July 31, 2012).

entropy. A related issue is the issue of missing values.

Other moving holiday effects may also be considered, as well as changing the moving window length to try and find an optimal setting for existing holiday regressors. An additional outlier effect for the 2011 Rugby World Cup will also be considered.

REFERENCES

Bell, W.R. and Hillmer, S.C. (1984) Issues Involved with the Seasonal Adjustment of Economic Time Series. *Journal of Business & Economic Statistics* **2**, No. 4, 291–320.

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

Findley, D.F. (2005) Some recent developments and directions in seasonal adjustment. *Journal of Official Statistics* **21**, 343–365.

Golub, G.H. and Van Loan, C.F. (1996) *Matrix Computations*. Johns Hopkins University Press, Baltimore.

Gómez, V. (2001) The use of Butterworth filters for trend and cycle estimation in economic time series,. *Journal of Business and Economic Statistics* **19**, 365–73.

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

Harvey, A.C. and Trimbur, T.M. (2003) General model-based filters for extracting trends and cycles in economic time series. *Review of Economics and Statistics* **85**, 244–255.

Hillmer, S. C. and Tiao, G. C. (1982) An ARIMA-model Based Approach to Seasonal Adjustment. *Journal of the American Statistical Association* **77**, 63-70

Hodrick, R. and Prescott, E. (1997) Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking* **29**, 1–16.

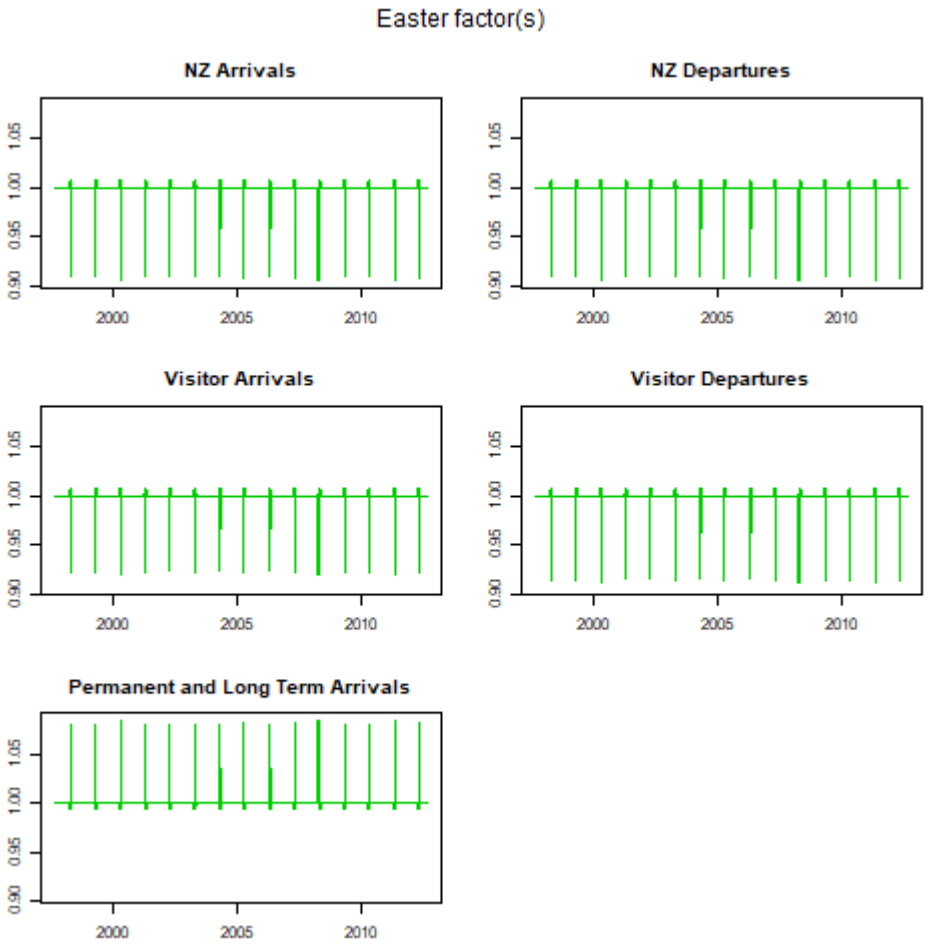

Figure 6: Easter factors for six (6) New Zealand immigration series (September 1, 1997 through July 31, 2012).

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

McElroy, T. (2017) Multivariate Seasonal Adjustment, Economic Identities, and Seasonal Taxonomy. *Journal of Business and Economics Statistics* **35**(4), 511–525.

McElroy, T. and Monsell, B. C. (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 http://dx.doi.org/10.1080/01621459.2014.978452

McElroy, T. and Penny, R. (2017) Maximum entropy extreme-value seasonal adjustment. Research report series, Statistics 2017-02. U.S. Census Bureau.

McElroy, T. and Trimbur, T. (2015) Signal Extraction for Nonstationary Multivariate Time Series with Illustrations for Trend Inflation. *Journal of Time Series Analysis* **36**, 209–227. Also in "Finance and Economics Discussion Series," Federal Reserve Board. 2012-45 http://www.federalreserve.gov/pubs/feds/2012/201245/201245abs.html

Tiao, G. and Tsay, R. (1983) Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Annals of Statistics* **11**, 856–871.

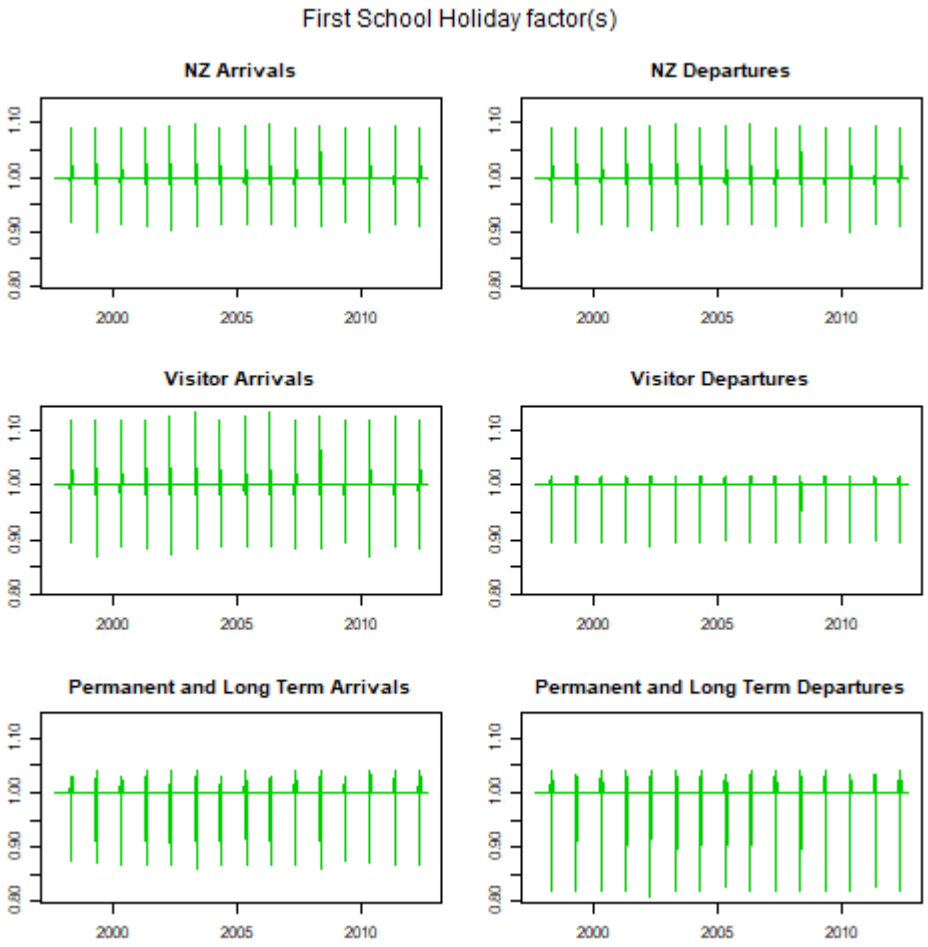

Figure 7: First semester NZ school holiday factors for six (6) New Zealand immigration series (September 1, 1997 through July 31, 2012).

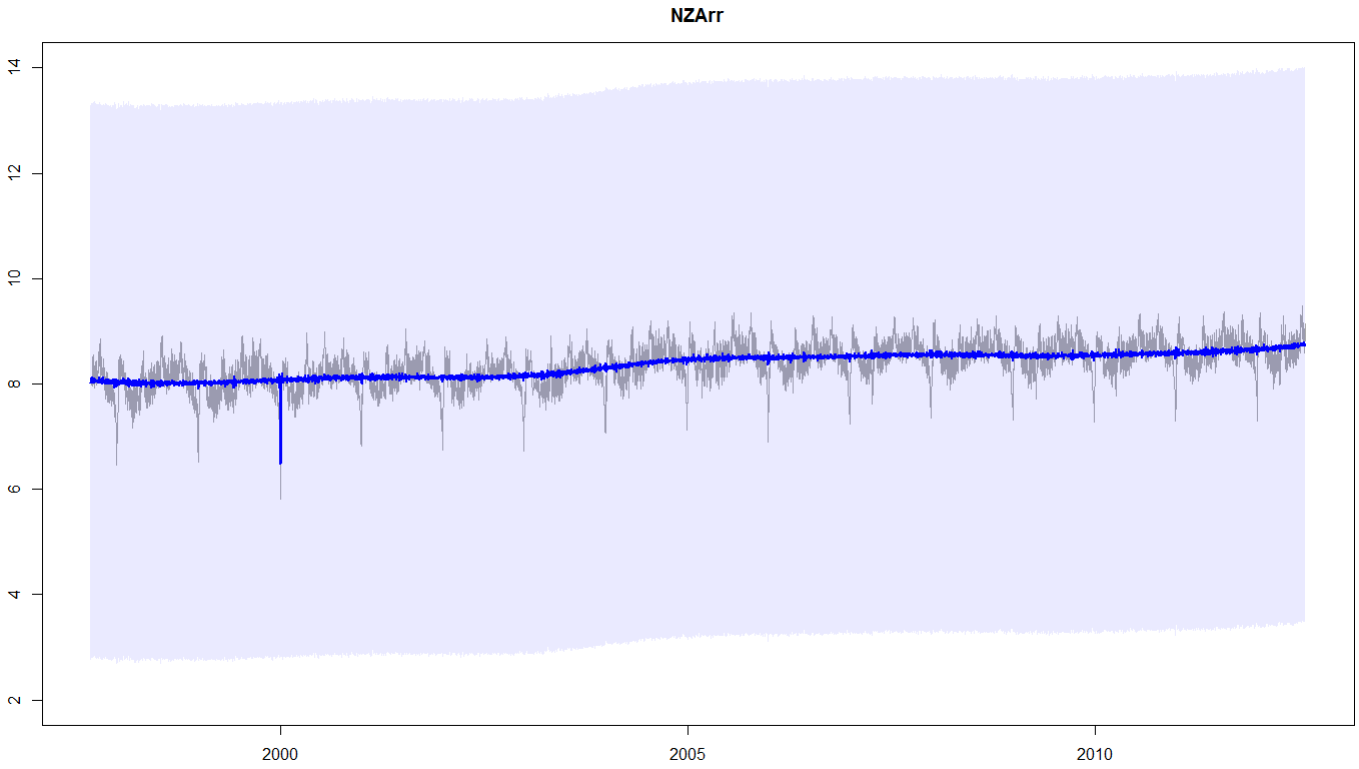

Figure 8: Seasonal adjustment for the logged NZ Arrivals series (September 1, 1997 through July 31, 2012). The original series is shown in grey, and the seasonally adjusted series is given in blue. The uncertainty in the seasonal adjustment is indicated with blue shading.

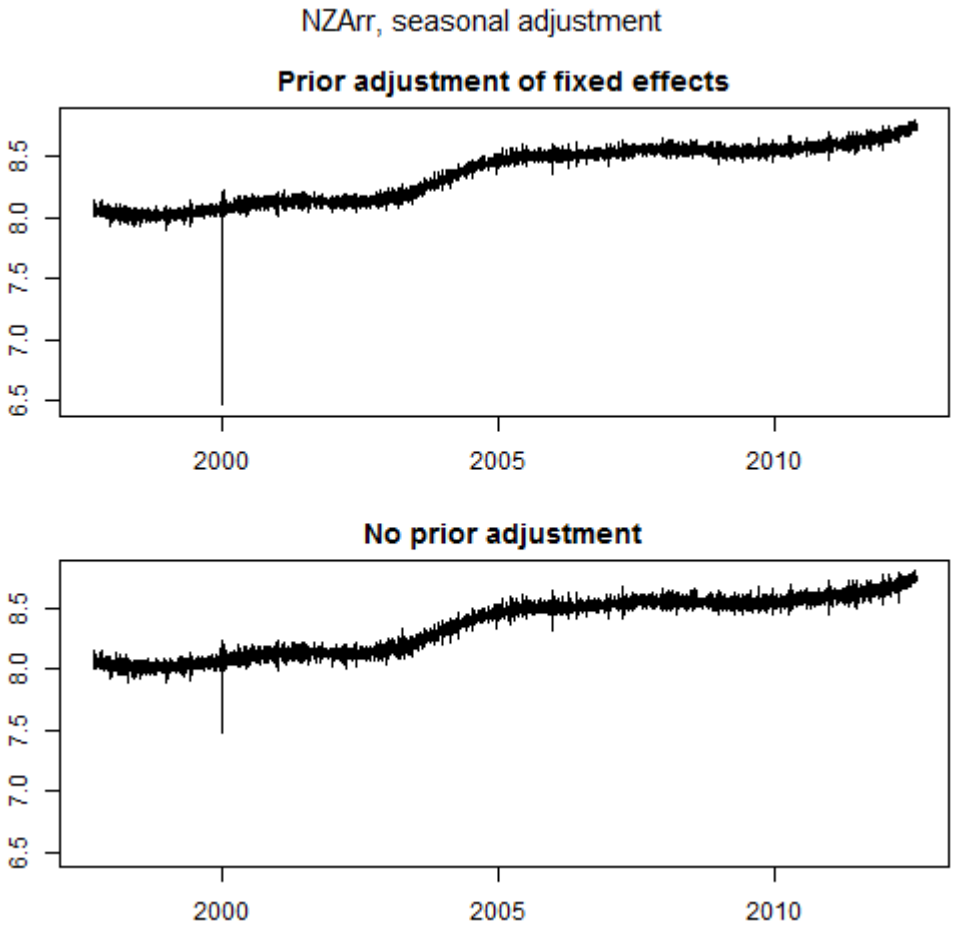

Figure 9: Seasonally adjusted series for the logged NZ Arrivals series (September 1, 1997 through July 31, 2012). The top panel is the seasonally adjusted series where the fixed effects were removed before signal extraction, the bottom panel is the seasonally adjusted series where no fixed effect were removed prior to signal extraction.

Figure 10: Trend component for the logged NZ Arrivals series (September 1, 1997 through July 31, 2012). The original series is shown in grey, and the trend series is given in orange. The uncertainty in the trend is indicated with orange shading. The fundamental seasonality (in green) and the combined weekly effects (in purple) are also shown.

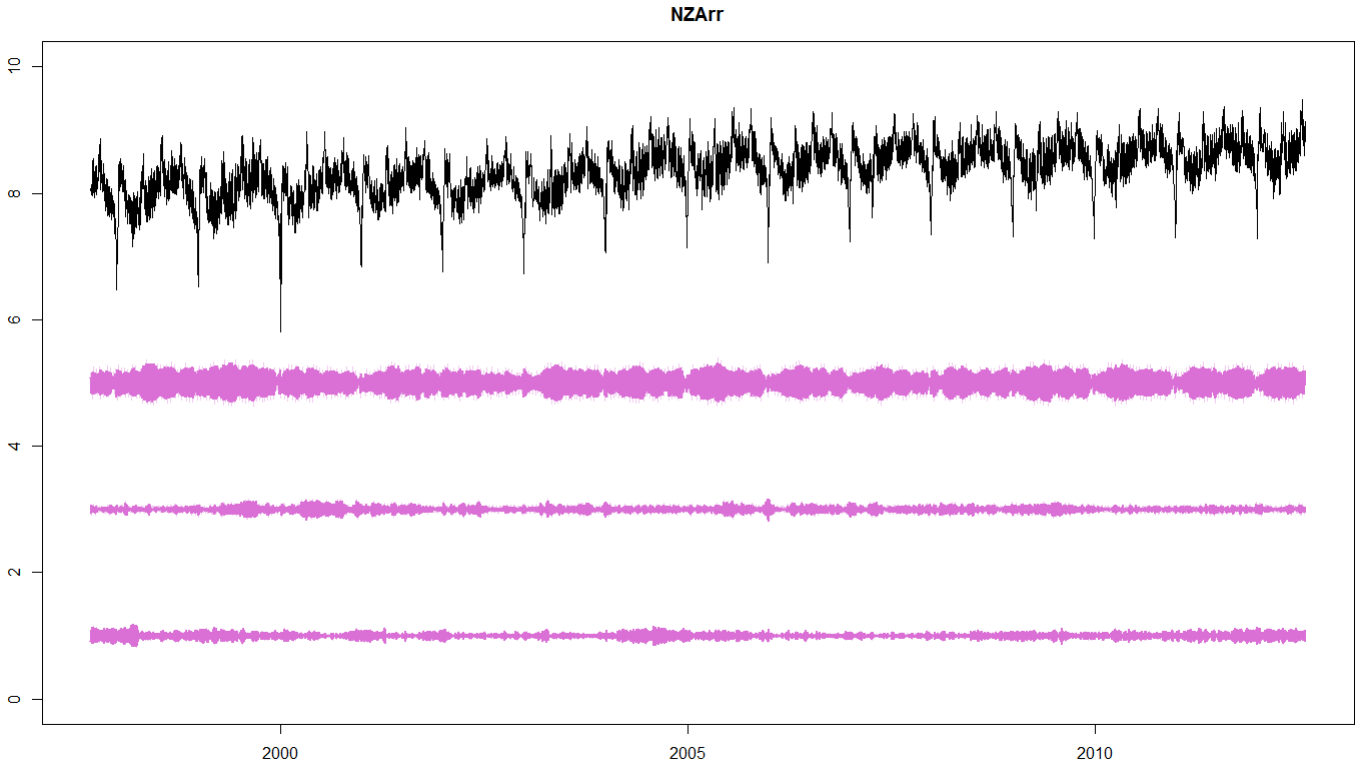

Figure 11: Weekly effects for the logged NZ Arrivals series (September 1, 1997 through July 31, 2012). The original series is shown in black, and the atomic weekly effects are also given in purple.

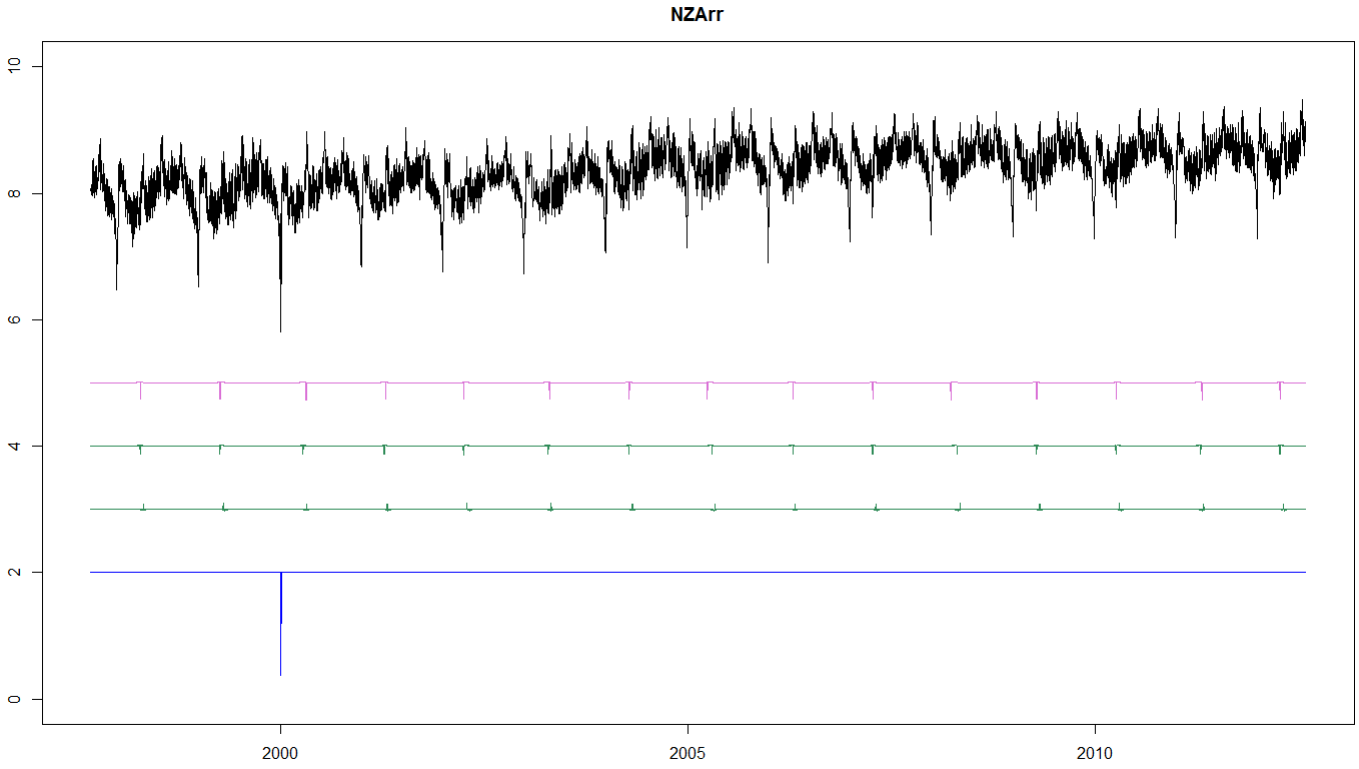

Figure 12: Moving holiday effects for the logged NZ Arrivals series (September 1, 1997 through July 31, 2012). The original series is shown in black. The moving holidays shown are Easter (in purple), NZ school holidays (in green), and point outliers for Y2K (in blue).

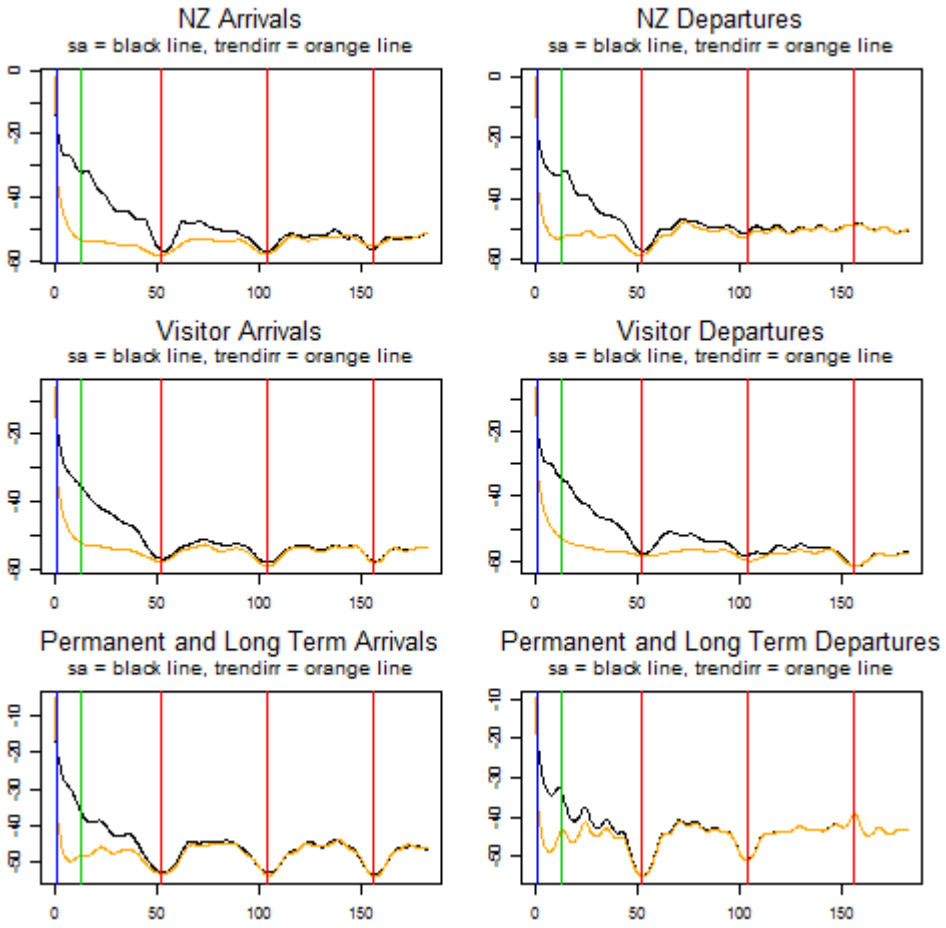

Figure 13: AR spectral density estimate of seasonal adjustments for 6 New Zealand immigration series (September 1, 1997 through July 31, 2012). Vertical red lines correspond to once a week, twice a week, and thrice a week phenomena; blue line corresponds to annual phenomena, and green line corresponds to monthly phenomena. Two adjustments are shown in the plot: one an adjustment from the original decomposition containing the fundamental seasonality (black line), and another the seasonal adjustment derived after the Hodrick-Prescott filter is used to estimate a trend component and the fundamental seasonality is isolated and removed (orange line).