ZAR Models: Improved Long Range Forecasting[1]

Granville Tunnicliffe Wilson* and John Haywood†

14 March 2011

1 Autoregressions and long range prediction

The aim of this report is to illustrate the application of a class of linear time series models, called ZAR models, which have been introduced to the literature in recent years. They are a generalization of linear autoregressive (AR) models that have the potential to improve the long range prediction of time series by extending the dependence on past values to high lags. We will review earlier work on these models at the end of this section, we describe them in section 3, and present their properties and estimation procedures in the following sections. A ZAR model is specified by its order, p, and a smoothing coefficient, θ, and reduces to an AR(p) model when $\theta = 0$. Standard AR (and ARIMA) models are widely and successfully used, but there are examples which suggest that we should be able to improve upon the predictions that they furnish. The specific example we use to illustrate the topics in this report is the series of the monthly USA unemployment rate from January 1968 to August 2009. In this section we consider long range forecasting issues typified by this example series, and in the next section we present forecasts of the series which illustrate the ability of the ZAR model to address these issues.

The upper two plots in Figure 1 show the monthly and annual raw unemployment rates. The lower two plots show the seasonally adjusted (S-A) rate and its logarithms, which we use for most of our statistical analyses for reasons given at the start of the next section. We comment that the main features are the same in all these series: the slow rise and fall of the overall level similar to a random walk (we will call this the *walk* feature) and the approximate five year economic cycle (the *cycle* feature). Seasonal adjustment is carried out because the seasonality evident in the first series obscures the underlying level that is

*Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK. email: g.tunnicliffe-wilson@lancaster.ac.uk

†School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, PO Box 600, Wellington 6140, NZ. email: john.haywood@vuw.ac.nz

[1]A version of this report will appear in Holan et al. (to appear), a book to be published in honour of David Findley's retirement from the US Census Bureau.
For further details of some of the derivations and proofs, see Reale and Tunnicliffe Wilson (to appear).

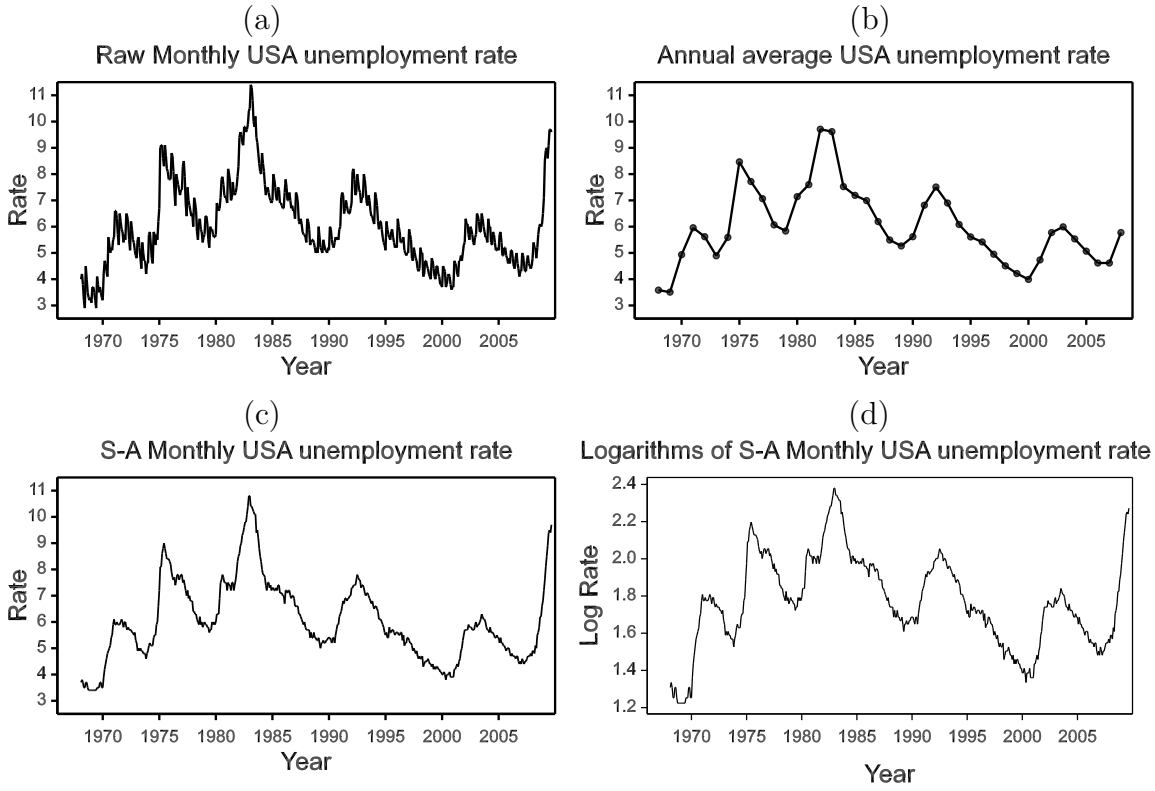

Figure 1: The USA unemployment rate: (a) raw values from January 1968 to August 2009, (b) the annual average rate from 1968 to 2008, (c) seasonally adjusted values from January 1968 to August 2009 and (d) logarithms of seasonally adjusted values from January 1968 to August 2009.

important for monitoring movements in the series from one month to the next. Forecasters are interested in predicting the series for periods of possibly several years ahead and it is usual to use S-A monthly data for this purpose. Two considerations that arise from the specific characteristics of this example are that:

1. the most notable features (walk and cycle) of the data are associated with low frequencies and long periods relative to the length of the series and the monthly sampling frequency, and

2. forecasts are desired to high lead times with respect to the monthly sampling frequency.

The first of these points raises the issue that standard AR models, selected using information criteria, do not have the flexibility to fit features within the small fraction of the frequency range that is associated with these long period movements. The upper plots in Figure 2 show the (unsmoothed) log spectra of the logarithms of the raw and adjusted series. Evidence for the walk and cycle features in the series is only barely visible in a collection of narrow peaks below frequency 0.02 in both these plots, *i.e.* within 5% of the range shown. The second point is an issue because standard AR models are fitted to predict just one step ahead. A further consideration in this respect is the effect of seasonal adjustment on the series. Figure 2(c) shows those parts of the log spectrum of the adjusted series for which the spectrum

is less than one half of that of the raw series. These illustrate that, in general, seasonal adjustment can introduce dips in the spectrum of the adjusted series around the frequencies of the seasonal harmonics.

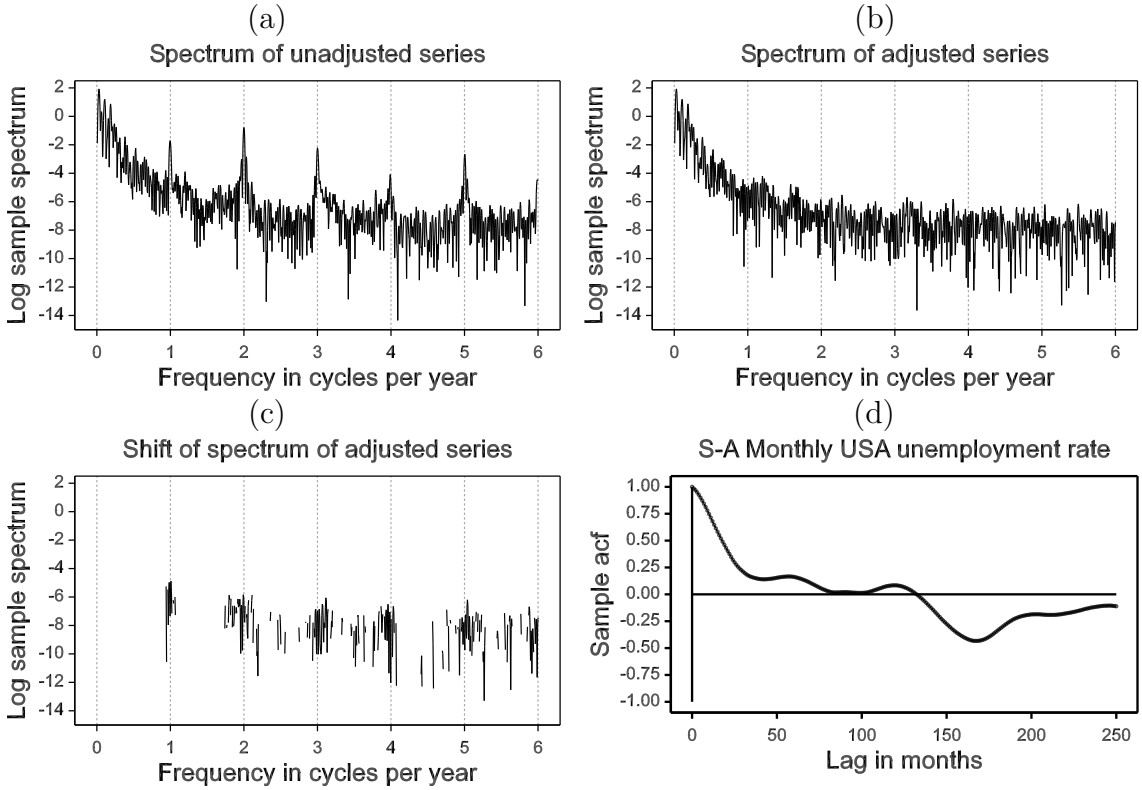

Figure 2: Log scale spectra of logarithms of the monthly unemployment rate (a) for the unadjusted series, (b) for the S-A series, (c) those parts of the log spectrum of the adjusted series for which the spectrum is less than one half of that of the raw series and (d) the sample acf of the adjusted series.

The low frequency spectral peaks are strongly evident in the un-logged spectra (not shown). However, a model fitted to minimize the one-step ahead error sum of squares is generally influenced by fluctuations of the series spectrum on the log scale. The structure associated with dips in the spectrum of the adjusted series at seasonal harmonic frequencies may then have an influence on the fitting of a parametric model that is comparable with that of the low frequency peaks. Given a sufficiently long series a standard AR model selected by the AIC, Akaike (1973), will consistently estimate the whole spectrum, Shibata (1980). However, for the series we are considering, the selected model may be inadequate to provide a good fit both to the cyclical features in the lower 5% of the range and to the pattern over the remaining 95% that is affected by seasonal adjustment. Note, from Figure 2(c), that the lower frequencies associated with the walk and cycle features are not affected by seasonal adjustment.

The time domain viewpoint is that the AR model essentially captures the lagged sample correlations (acf) of a series up to the order of the model, and the order of model required might be expected to be at least comparable with the time period of the cycles, of about 60 months. Figure 2(d) shows the sample autocorrelations of the logarithms of the seasonally

adjusted series. Evidence for the approximate five year economic cycle in the acf is obscured by the effect of the walk, and lags up to several multiples of the cycle period are needed to reveal it.

In section 3, and those following, we describe the ZAR model, and methods for fitting this model, which address the issues raised in the previous paragraphs as follows:

1. the flexibility of the model lies in fitting features at low frequencies,
2. the method of fitting allows more weight to be placed on lower frequencies.

The corresponding time domain properties are that the model places weight on past data up to higher lags, and the fitting method also seeks to minimize the prediction error over high lead times.

For many years there has been an ongoing interest in methods for improved forecasting at higher lead times. This is based on the recognition that minimizing single-step forecast error variance, which is equivalent to maximum likelihood model estimation, can be far from optimal for higher lead time forecasting if the model is not correct. This may be true even if the mis-specification may appear relatively minor in the sense that residual diagnostics do not readily reveal model inadequacy. Much ground-breaking work in this area was presented in Findley (1983, 1985, 1990, 1991) and discussion with Dr Findley when he visited Lancaster lead to the development of Haywood and Tunnicliffe Wilson (1997), on fitting models by minimizing squared multistep-ahead errors. Our extension of this idea to the construction of a test for improved multi-step forecasting, Haywood and Tunnicliffe Wilson (2009), includes a review of more recent work by other authors in this general area. We showed that for commonly occurring models our approach implicitly gave greater weight to the lower frequencies in the data, but this weight was model dependent.

The ZAR modeling approach is based on a very different idea, but is similar in that it has the capacity to give greater weight to lower frequencies in the data. However, this weighting is prescribed and not model dependent. The ideas of ZAR modeling may be traced back to the generalized continuous time shift operator and Laguerre filters of Wiener (1949). Discrete time versions have, more recently, been used extensively in systems modeling, Wahlberg (1991). These ideas have been introduced by Wahlberg and Hannan (1993) in a class of time series models very closely related to the ZAR models of this report. Our formulation of the discrete time ZAR model developed from a continuous time model presented in Belcher et al. (1994). It is first found in Morton and Tunnicliffe Wilson (2004) where the model is equivalent to that in Wahlberg and Hannan (1993), and slightly different from the one presented in this report. Underlying the models is the idea of basis functions for the space of past observations and an excellent mathematical background to these concepts can be found in Partington (1997). Multivariate time series applications are presented in Morton and Tunnicliffe Wilson (2001), Tunnicliffe Wilson et al. (2001) and Tunnicliffe Wilson and Morton (2004), as also are continuous time versions of the model. The Ph.D. theses of Morton (2000), Ibañez (2005) and Lo (2008) study, respectively, the multivariate continuous time model, non-linear ZAR models and multi-step prediction properties of ZAR models. The findings of the second thesis are summarized in Ibañez and Tunnicliffe Wilson (2007).

2 Some forecast comparisons

In this section we compare predictions of the unemployment series using both the standard AR (and ARIMA) model and the ZAR model. All the models are fitted to the logarithms of the data then transformed back to display forecasts. The results are not, however, particularly sensitive to this transformation. There is not a great difference in appearance between Figures 1(c) and (d), but transformation does give a more similar appearance to the five or so peaks seen in the series. The asymmetry in the rise and fall of the peaks is evidence of some non-linear behavior which is not treated by the logarithmic transformation. The models presented in this report are all linear, but nevertheless have the capacity to reflect, in their predictions, the pattern of cycles evident in the past. All predictions are out of sample, *i.e.* the models are fitted to data up to the chosen origin then forecasts made over the remaining period. The error limits shown on all forecasts are designed for 90% coverage at any given point.

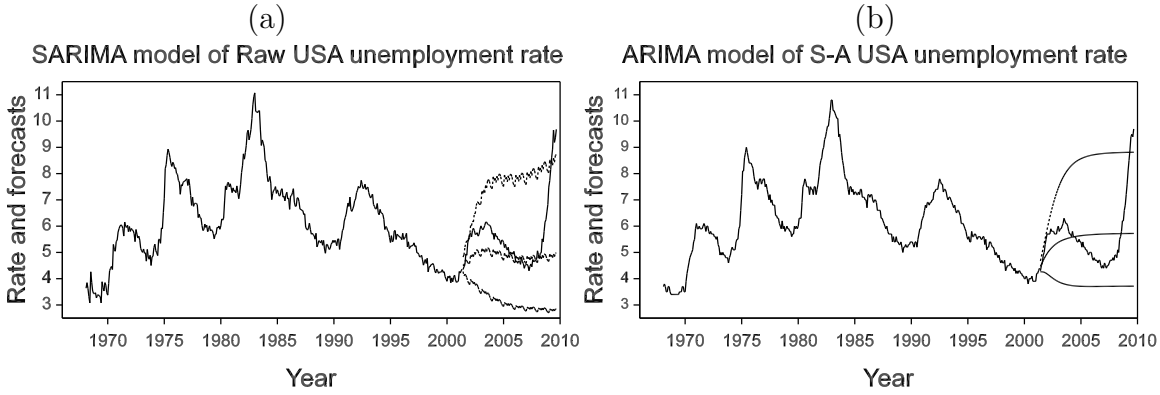

Figure 3: Forecasts of the USA unemployment rate from May 2001: (a) for the raw series with results corrected for a fixed annual cycle, and (b) for the seasonally adjusted series.

Consider first the raw monthly series. A Box-Jenkins seasonal ARIMA model, Box and Jenkins (1970), fits this well with non-seasonal autoregressive, differencing and moving average orders set to $p = 2$, $d = 0$ and $q = 1$, corresponding seasonal orders $P = 0$, $D = 1$ and $Q = 1$, and seasonality set to $s = 12$. This model is encompassed by the default set of models used by the PICKMDL procedure of X-12-ARIMA. Residual diagnostics were good except for a hint of some calender effects. The estimated non-seasonal autoregressive operator characterizes irregular cyclical behavior, and this is to some extent reflected in the forecasts from May 2001, shown in Figure 3(a), which hint at the oscillation observed in future values. In order to show the pattern of forecasts more clearly we have subtracted from the series and forecasts displayed in this graph a fixed annual cycle, obtained by regression on the set of seasonal harmonics. Because the seasonality is not constant some evidence of this remains in the figures.

Forecasts of the seasonally adjusted series are shown, from the same origin, in Figure 3(b). The non-seasonal ARIMA model used for this had orders $p = 3$, $d = 0$ and $q = 2$. All parameters were significant and residual diagnostics were very good. The forecasts from this model, shown in Figure 3(b), are not very markedly different from those for the raw series,

but fail to show any hint of the future cycle.

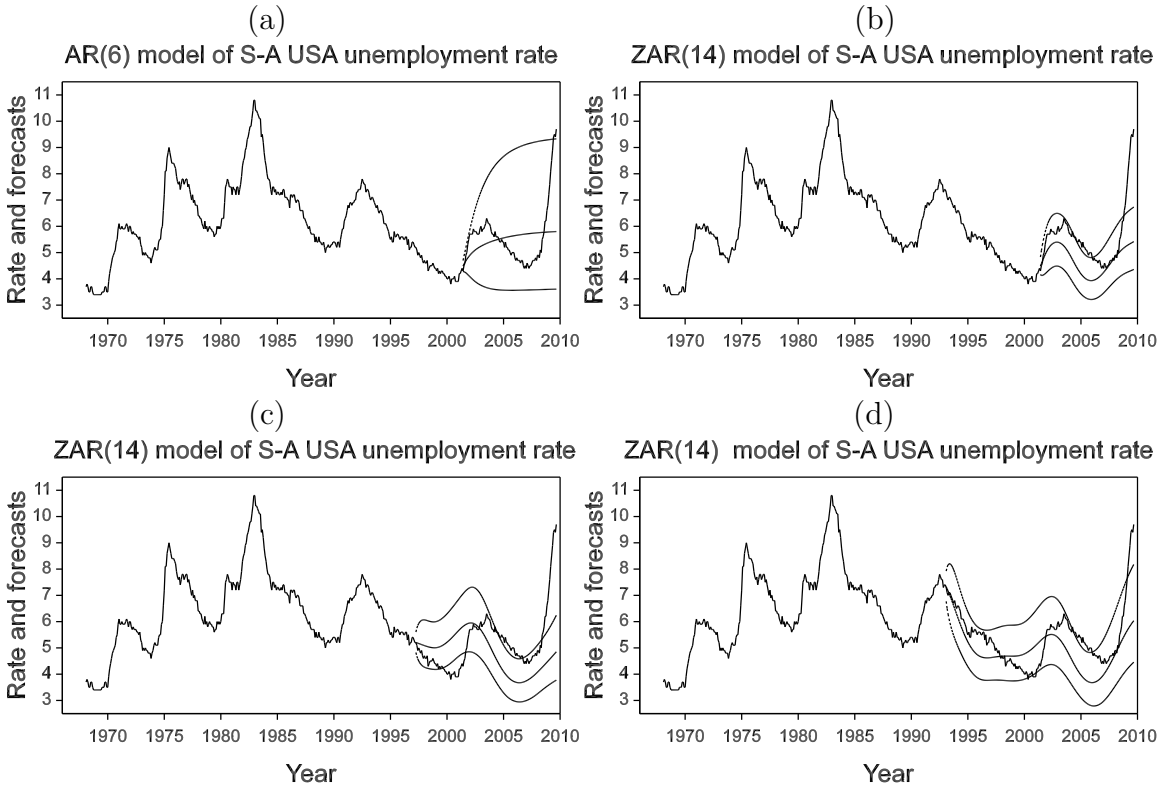

Figure 4: Forecasts of the S-A monthly USA unemployment: (a) from May 2001, using an AR(6) model, (b) from May 2001, using a ZAR(14) model, (c) from March 1997, using a ZAR(14) model, and (d) from January 1993, using a ZAR(14) model.

Illustrations of forecasts obtained by application of the ZAR model to the seasonally adjusted series are shown in Figure 4, except that in Figure 4(a) the results are from the model restricted to the special case of the standard AR model for which AIC selects the order 6 and the forecasts are similar to those in Figure 3(b) from the ARMA model. The remaining frames of this figure show forecasts made using a ZAR model of order 14 from origins of May 2001, March 1997 and January 1993, which are respectively 400, 350 and 300 months from the start of the series. The forecast functions in all cases reflect quite well the cyclical pattern in the future values with realistic error limits, which is encouraging. However, on the logarithmic scale of the series, the ratio of the forecast error sum of squares up to 6 years ahead for the ZAR(14) model, is greater than that of the AR(6) model by a factor of 1.38 for the forecasts from May 2001 shown in Figure 4(a) and (b). Simple models with forecast functions showing persistence, or mean reversion as in Figure 4(a) can be hard to beat, though, overall, the more accurate forecast might be judged to be that of the ZAR(14) model in Figure 4(b) when the much more precise error limits are taken into account. A similar comparison of forecasts from March 1997 and April 1993 results in ratios of respectively 0.79 and 0.33, strongly favoring the ZAR(14) model. This model, *i.e.* its order $p = 14$ and ZAR coefficient $\theta = 0.94$, was selected in preference to the mean-reverting AR(6) model, using an information criterion described in section 6. This is an extension of the AIC which provides protection against the over-fitting of these more complex models. We recall the cautious advice of Tukey (1986) regarding attempts to extract too much structure

from data, but he also makes a positive point that we should "look for appearances" in the data, and the cycles are quite apparent here, so a cautious attempt to include them in the results is recommended. Another point made in that article, specific to time series, is that one is fortunate if a sufficient length of data is available to determine its structure, without some change of structure occurring over that time span. For the unemployment series it would not be surprising if structural breaks and lack of stationarity occurred over the 40 year span we are modeling: the last few points reflect the most recent dramatic break and lie well outside the forecast limits of the ZAR model. Nevertheless, substantial economic shocks have previously had their impact on the recorded unemployment, and the dynamic behavior of the series seems to persist. The application to the unemployment series is very useful for illustrating the motivation and methodology, but it is a testing example, which we offer for critical re-assessment by anyone who may wish to investigate these models further. We have used ZAR models and methods successfully in other contexts, including extensions to vector time series, and are confident of their value and potential in appropriate situations.

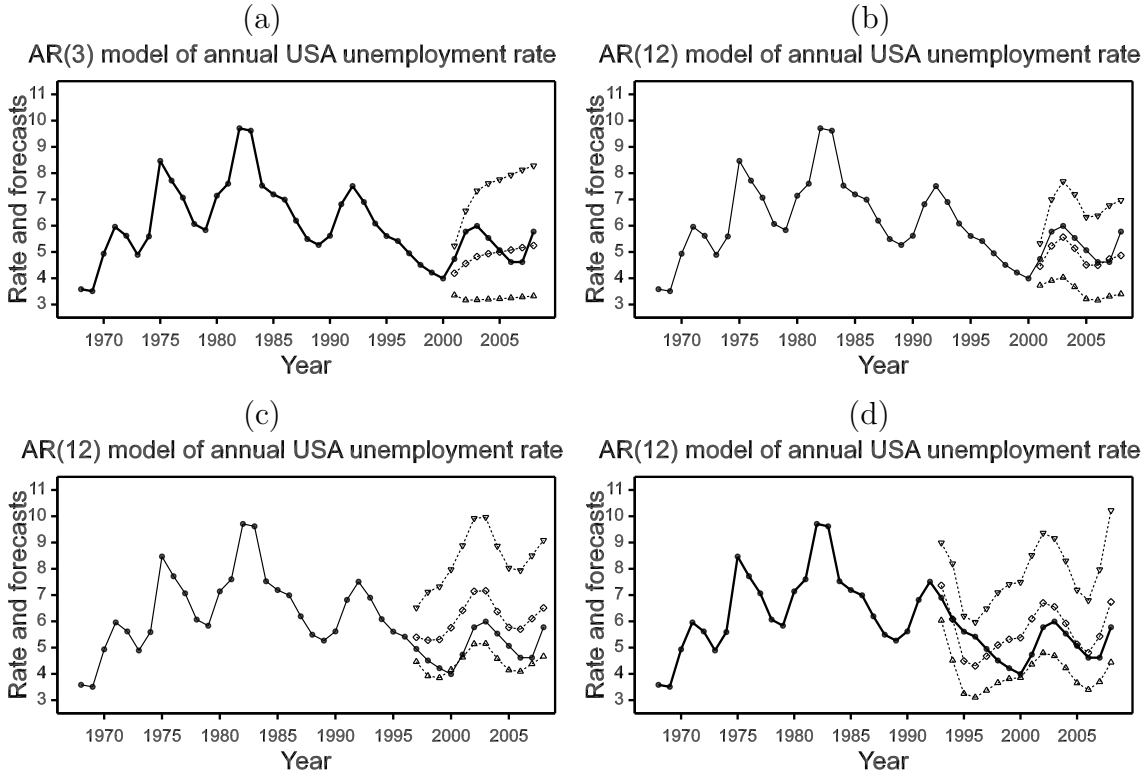

Figure 5: Forecasts of the average annual USA unemployment rate: (a) from the year 2000, using an AR(3) model,(b) from the year 2000, using an AR(12) model, (c) from the year 1996, using an AR(12) model and (d) from the year 1992, using an AR(12) model.

To complete this section we consider the annual average series. Our point is to emphasize that our advocacy of the ZAR model is related to the higher, monthly, sampling frequency that makes it challenging for the standard AR to model the series parsimoniously. In contrast, the standard AR model can be applied to give annual forecasts that parallel those shown for the monthly ZAR model in Figure 4. Inspection of information criteria displayed in Figure 6 suggested two possible orders, 3 or 12. The modest forecasting performance of the AR(3) is shown in Figure 5(a), and is comparable with the forecast of the S-A monthly series in

Figure 4(a). Forecasts from the same origin for the AR(12) model are shown in Figure 5(b). These are quite close to the actual and comparable to those from the ZAR(14) model for the S-A monthly series in Figure 4(b). Similarly, forecasts from the earlier origins of 1996 and 1992, shown in Figures 5(c) and (d), though not so close, are comparable to those in Figure 4(c) and (d): they reflect well the cyclical pattern of the future values.

The order 12 was determined by application of the AIC to the whole series. This is shown in Figure 6(a) together with three other criteria. The first is the modified AIC of Hurvich and Tsai (1989), which it is appropriate to consider when the order of model is a substantial fraction of the series length, as in this application: an order 12 model is fitted to just 29 points in the example shown in Figure 5(c). The second is the Hannan and Quinn (1979) criterion, and the third is the Schwarz (1978) criterion. All these select order 3 except the AIC which selects order 12. Knowing the tendency of the AIC to over-estimate the true order of a model, one might hesitate to select the order 12 as over-fitting the data. To avoid the loss of degrees of freedom suffered by simple lagged regression, which is serious for short series and high order models, we have used the exact likelihood for estimation, and in the information criteria. However, the exact log-likelihood can be far from quadratic in its parameters, and care is needed to locate the MLE.

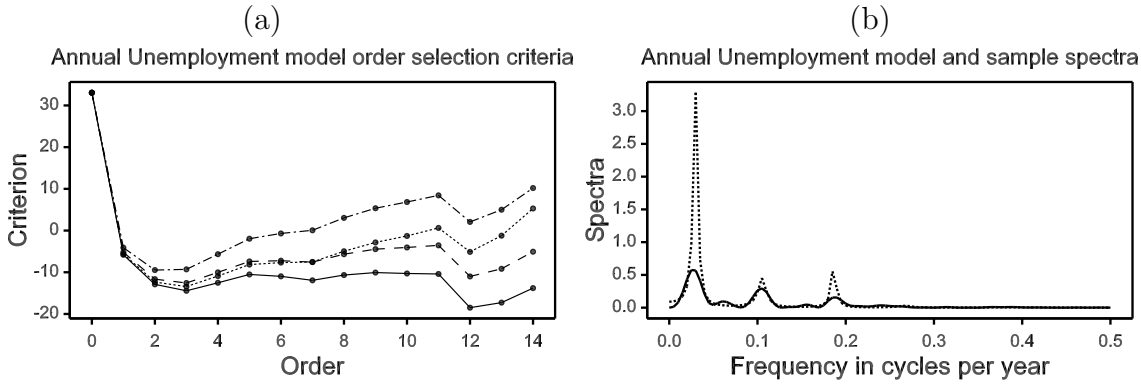

Figure 6: (a) Autoregressive modeling of the average annual unemployment rate: (a) order selection criteria, respectively the AIC, the Hannan-Quinn, the Hurvich-Tsai and the Schwarz criteria in increasing value at order 14, and (b) the sample (solid line) and model (dotted line) spectra of the series fitted by an AR(12).

The comparison between the AR model for the annual and the ZAR model for the monthly series is further illustrated by Figure 6(b) which shows how the spectrum of the AR(12) model matches the sample spectrum of the annual series. This is the raw (unsmoothed) spectrum and the tallest peak is at the lowest harmonic frequency of the series. Such a peak is typical of a mean corrected (near) random walk process which gives the illusion of a cycle with period close to the series length. Such behavior appears to be a component feature of the series. The fact that the spectral peaks of the fitted model are higher than those of the sample spectrum is due to the use of exact maximum likelihood estimation of the model. The peak just below frequency 0.2 one would associate with the economic cycle which is the other main component feature with approximate period of 5 years. The peak just above frequency 0.1 (period 10 years) corresponds to no immediately visually evident feature in the series. The testing of spectral peaks to avoid spurious detection of cycles has

a long history, see Priestley (1981, p. 406) and we must interpret this apparent cycle with caution, even though supported by an AR model. The successful extrapolations from this model suggest, however, that the implied spectral peaks are important.

To return to the comparison with modeling the monthly S-A series, this is much longer, 500 instead of 41 values, but it does not contain much more information about the cycles. It just contains the extra 95% or so of higher frequency spectral components shown in Figure 2. One strategy for obtaining well-fitting AR models from high frequency data is subsampling as in this example, using annual averages of monthly values. But to forgo monthly records and monthly forecasts for such a reason is not desirable. ZAR models enable us to avoid the sub-sampling strategy whilst retaining well-fitting models.

3 ZAR models

In the remainder of this report we present the ZAR model, state and explain its properties including how it is fitted to data, and illustrate its application to the unemployment series. The reader is referred for derivations, proofs and further explanations to the papers referenced earlier, and to the forthcoming book, Reale and Tunnicliffe Wilson (to appear). The models are based on the generalized shift operator Z, acting on present and past values of a series and defined in terms of the backward shift (or lag) operator B as

$$Z = \frac{B-\theta}{1-\theta B} = -\theta + (1-\theta^2)(B + \theta B^2 + \theta^2 B^3 + \ldots)$$

where θ is a specified smoothing coefficient, or discount factor, which lies in $[0, 1)$. We shall also write Z_θ to indicate the dependence of the operator upon θ, except that for convenience of notation we shall only use this form when some other symbol than θ is used, and will omit the subscript when the symbol is θ. The acronym ZAR is motivated by the use of this notation. In the case $\theta = 0$ we have $Z = B$. In practice the calculation of $s_t = Z\,x_t$ is by the recursion:

$$s_t = x_{t-1} - \theta x_t + \theta s_{t-1} \tag{1}$$

and Z may be applied repeatedly to construct a set of series that we will call ZAR states $s_t^{(k)} = Z^k\,x_t$, including $s_t^{(0)} = x_t$. The effect of Z on a slowly varying series is similar to applying a lag of $\ell = (1+\theta)/(1-\theta)$. This is illustrated using $\theta = 0.9$ in Figure 7(a) which shows the S-A unemployment series x_t with $Z\,x_t$ and $Z^3\,x_t$ for which the approximate lags are 19 and 57 months. Initial values of the states have to be assigned to start the recursions in (1). We describe in section 5 how these can be set to reduce the errors generated in subsequent states. However, these errors are transient, decaying in general like θ^t.

Algebraic manipulation shows that $Z^{-1} = (B^{-1}-\theta)/(1-\theta B^{-1})$ and is an operator on present and future values. The operator Z is unimodular, *i.e.* if we take $B = \exp 2\pi i f$, a value on the unit circle $|B| = 1$, then $Z = \exp 2\pi i g$, is also a value on the unit circle $|Z| = 1$. A consequence is that for any k, $Z^k x_t$ has exactly the same lagged covariances and spectrum

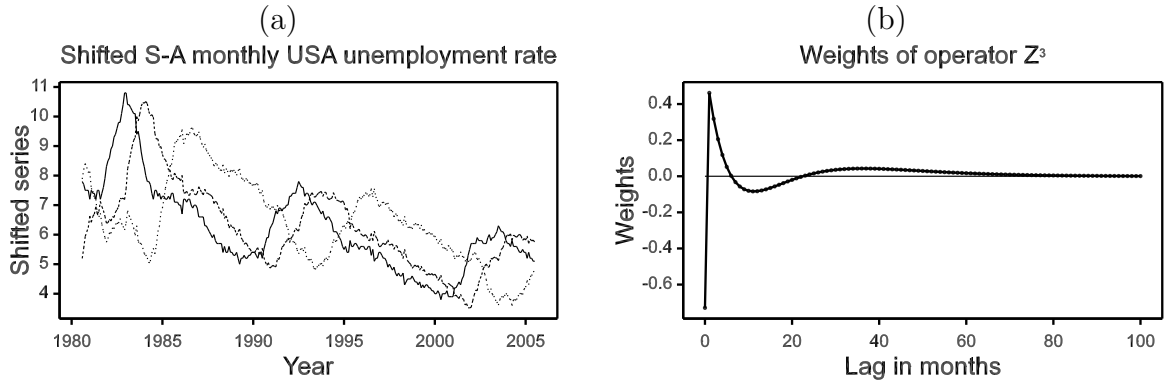

Figure 7: (a) The S-A monthly unemployment series x_t (solid line) with $Z\,x_t$ (dotted line) and $Z^3\,x_t$ (fine dotted line) formed using the generalized shift operator with $\theta = 0.9$, (b) the weights implicitly applied by Z^3 to lagged values of x_t in the construction of $Z^3\,x_t$.

as x_t (just as is true for $B^k x_t$). The dependence of g on f is given by

$$\cos 2\pi g = \frac{(1+\theta^2)\cos 2\pi f - 2\theta}{1+\theta^2 - 2\theta\cos 2\pi f} \tag{2}$$

and is illustrated later in Figure 8(a) for $\theta = 0.6$, where it is used to give insight into ZAR models. It is known as a *frequency warp*, a term which has a long history; see *e.g.* Braccini and Oppenheim (1974). We shall also find useful the expression for the derivatives

$$\frac{dg}{df} = \frac{1-\theta^2}{1+\theta^2 - 2\theta\cos 2\pi f} \quad \text{and} \quad \frac{df}{dg} = \frac{1-\theta^2}{1+\theta^2 + 2\theta\cos 2\pi g}. \tag{3}$$

Our ZAR models are motivated by expressing the predictor of the *future* value $Z_\rho^{-1}x_t$ as a linear combination of the finite set of present and *past* values x_t, $Z\,x_t$, …, $Z^{p-1}\,x_t$, where we note that the operator Z_ρ^{-1} acting on the future may in general be chosen to have a different discount coefficient from that used to construct the predictors. In fact $Z_\rho^{-1}x_t$ attaches weight of $-\rho$ to the present value of x_t, but it is useful to consider it essentially as a function of the future. In constructing this predictor the choice of the explicitly specified coefficient ρ provides a robust alternative to the choice of the simple one step ahead predicted value x_{t+1}, although this possibility can be included by setting $\rho = 0$. For a slowly varying series, $Z_\rho^{-1}x_t$ is an approximation to $x_{t+\ell}$ where $\ell = (1+\rho)/(1-\rho)$, and we shall see from our example that choosing a quite modest value of $\rho = 0.5$, for which $\ell = 3$, can substantially improve the accuracy of multi-step forecasts generated by the ZAR model. The predictors $Z^k x_t$ in the ZAR model depend on the coefficient θ which is not explicitly displayed. The choice of θ and the model order p allows the prediction to depend on values of x_t at lags up to and somewhat beyond $p(1+\theta)/(1-\theta)$, as illustrated in Figure 7(b). We shall again see from our example that this is essential if the predictor is to use information from the past in an efficient manner. In section 6 we will show how an information criterion that generalizes the AIC can be used to select both the coefficient θ and order p, but the choice of ρ remains to a large extent subjective. This is similar to the choice of lead time to use when fitting models for multi-step prediction. As for classical ARMA modeling, if the specified

model is correct, *i.e.* it does exactly represent the series, estimation is optimally achieved by minimizing the one-step ahead forecast error, *i.e.* by setting $\rho = 0$. However, in a realistic situation, approximating a multi-step predictor by using a value of $\rho > 0$ can give improved prediction over a range of lead times. Moreover, the loss of efficiency from using $\rho > 0$ even when the model is correct, is not necessarily substantial.

We now specify the ZAR(p, θ) model. We shall in fact set down three forms of the model: first the *general* form, with ρ unrestricted. The second is the *predictive* form which is motivated by restricting $\rho = 0$ and the third is the *natural* form, in which ρ is set equal to θ. The important point is that these three forms are *exactly equivalent* – there is only one ZAR(p, θ) model, which is most sensibly expressed as the predictive form. The prediction coefficients in any one of these models can be algebraically transformed to those of any other form in a quite simple manner. The coefficient ρ is important for influencing the *fitting* of the model but is *not* included as a parameter of the model with p and θ. The coefficients of the general form of model obtained by regression of $Z_\rho^{-1} x_t$ on x_t, $Z\, x_t$, $\ldots$, $Z^{p-1}\, x_t$, will depend on the choice of ρ, but can be transformed to those of the predictive form of the model which would realize *exactly* the same prediction of $Z_\rho^{-1} x_t$.

For a mean-corrected stationary process we express the models in the conventional manner, with the *present value* x_t given in terms of a set of predictors and an error term. To derive this from the motivating prediction of $Z_\rho^{-1} x_t$ by x_t, $Z\, x_t$, $\ldots$, $Z^{p-1}\, x_t$, we simply multiply all terms by Z_ρ. The general form of model is then:

$$x_t = Z_\rho \left(\zeta_1\, x_t + \zeta_2\, Z\, x_t + \cdots + \zeta_p\, Z^{p-1}\, x_t \right) + n_t. \tag{4}$$

We remark immediately that, except in the case $\rho = 0$, the (so-called) predictors on the RHS of this model are linear functions of the present value x_t, as well as lagged values. Before considering further the implications of this fact, particularly regarding the error term n_t, we consider the predictive form of model obtained by setting $\rho = 0$, so that $Z_\rho = B$:

$$x_t = \xi_1\, x_{t-1} + \xi_2\, Z\, x_{t-1} + \cdots + \xi_p\, Z^{p-1}\, x_{t-1} + e_t \tag{5}$$

The predictors are now proper linear combinations of past values alone, and we assume e_t to be white noise uncorrelated with all past values x_{t-k} for $k > 0$, *i.e.* the linear innovation process of x_t. We also write the predictive form of model in operator notation as

$$\{1 - B\, \xi(Z)\} x_t = e_t,$$

where

$$\xi(Z) = \xi_1 + \xi_2\, Z + \cdots + \xi_p\, Z^{p-1}.$$

Returning to the general form of model (4), the assumption that the prediction error in $Z_\rho^{-1} x_t$ is orthogonal to x_{t-k} for $k \geq 0$ implies that n_t follows the AR(1) model

$$n_t = \rho\, n_{t-1} + \varepsilon_t. \tag{6}$$

The process ε_t is white noise, proportional to the innovation series e_t of the predictive form:

$$\varepsilon_t = M e_t.$$

We can then express the general form of model in operator notation as

$$(1 - \rho\, B)\,\{1 - Z_\rho \zeta(Z)\}\, x_t = \varepsilon_t,$$

from which the algebraic equivalence with the predictive form of model is found as

$$(1 - \rho\, B)\,\{1 - Z_\rho \zeta(Z)\} = M\{1 - B\,\xi(Z)\}.$$

Depending on which of $\xi(Z)$ and $\zeta(Z)$ is to be derived from the other, we can determine $M = \{1 + \rho\,\zeta(-\theta)\}$ or $M = (1 - \rho^2)/\{1 - \rho\xi(-\tau)\}$, where $\tau = (\theta - \rho)/(1 - \theta\rho)$.

Finally, the natural form of model is obtained by setting $\rho = \theta$ and hence $Z_\rho = Z$ in the general form, to give:

$$x_t = \varphi_1 Z\, x_t + \varphi_2 Z^2\, x_t + \cdots + \varphi_p Z^p\, x_t + n_t,$$

where $(1 - \theta\, B) n_t = \varepsilon_t$. We will also write this model as $\varphi(Z) x_t = n_t$ where

$$\varphi(Z) = 1 - \varphi_1 Z - \varphi_2 Z^2 - \cdots - \varphi_p Z^p.$$

There are several important properties of these models that we now state.

1. Each model represents a stationary process if, when transformed to the natural form, the operator $\varphi(Z)$ satisfies the stationarity condition of the standard AR model, *i.e.* considering Z as a complex variable, $\varphi(Z)$ has no zeros inside or on the unit circle.

2. For any non-deterministic stationary process x_t, let coefficients of the model in the general form be determined for fixed ρ, θ and p by projecting x_t on to the predicting variables $Z_\rho Z^k\, x_t$, $k = 0, 1, \ldots, p-1$, on the RHS of (4). Then the ZAR model so derived will satisfy the stationarity condition. Moreover, in the limit as $p \to \infty$, the projection error n_t will converge in mean square to a process that follows the AR(1) model (6) with ε_t proportional to the innovation series of x_t. In this sense the models can approximate x_t to any arbitrary level of precision.

3. The model can be expressed as a restricted ARMA$(p, p-1)$ of the form

   $$\phi(B) x_t = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) x_t = (1 - \theta B)^{p-1} e_t.$$

   This expression is useful for deriving some properties of the model but is of little practical value because for models of reasonably high order, the parameterization typically leads to collinearity problems if used in model fitting. Note however that the reciprocal zeros r_k of $\phi(B)$, which characterize the dynamics of the process x_t, *i.e.* the decay rate and period of any cyclical component of the correlations, are related to the recipro-

cal zeros q_k of $\varphi(B)$ by $r_k = (\theta + q_k)/(1 + \theta\, q_k)$, and will generally be much closer to the unit circle. The ZAR model can therefore capture long-term dynamics with a parameterization that does not suffer the problems of zeros of $\varphi(B)$ close to the unit circle, which can lead to highly collinear estimates of model coefficients in standard AR models, due to the highly correlated predictors. The predictors of the ZAR model are in general less highly correlated.

4. There is a simple state space representation of the predictive form of the model, in terms of the ZAR states. The state transition equation is:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \theta & 1 & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & \theta & 1 \end{pmatrix} \begin{pmatrix} s_t^{(0)} \\ s_t^{(1)} \\ \vdots \\ s_t^{(p-1)} \end{pmatrix} = \begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_p \\ 1 & \theta & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & 1 & \theta \end{pmatrix} \begin{pmatrix} s_{t-1}^{(0)} \\ s_{t-1}^{(1)} \\ \vdots \\ s_{t-1}^{(p-1)} \end{pmatrix} + \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

or $L\, S_t = R\, S_{t-1} + E_t$, where S_t is the state vector at time t. The first row of the equation is simply the predictive model and the remaining rows represent the recursive calculation of the states. The conventional form of a state space representation is obtained on premultiplying through by L^{-1}, leading to the state transition matrix $T = L^{-1}R$. The observation equation is simply $x_t = s_t^{(0)}$. This representation is useful for calculating model properties, constructing predictions and model estimation. In particular the covariance matrix V_S of the state vector S_t can be calculated using standard methods, and then $V_{S,k} = \mathrm{Cov}\,(S_t, S_{t-k}) = T^k V_S$. The first element in this matrix is $\mathrm{Cov}\,(x_t, x_{t-k})$.

4 Understanding the ZAR model

Insight into the nature of the ZAR model for any stationary process x_t is given by defining, for any fixed t, a related process:

$$X_k = s_t^{(-k)} = Z^{-k} x_t.$$

Then X_k, $k = \ldots, -1, 0, 1, 2, \ldots$ is also a stationary process, with the terms for $k \leq 0$ providing a basis for $x_s, s \leq t$ and the terms for $k \geq 0$ providing a basis for $x_s, s \geq t$. That X_k is stationary is verified by deriving its lagged covariances from the spectrum $S_x(f)$ of x_t:

$$\Gamma_{X,v} = \mathrm{Cov}\,(X_k, X_{k+v}) = \int_{-0.5}^{0.5} Z^{-k} \bar{Z}^{-(k+v)} S_x(f) df = \int_{-0.5}^{0.5} Z^v S_x(f) df, \tag{7}$$

which depends only upon v. To clarify this derivation, we have set $B = \exp 2\pi i f$ in $Z = (B - \theta)/(1 - \theta B)$ within the integral and used the property that $\bar{Z} = Z^{-1}$. Further insight is now obtained by transforming the integral (7) by substituting f in terms of g as defined

in (2). Then, on including the Jacobian of the transformation via (3),

$$\Gamma_{X,v} = \int_{-0.5}^{0.5} \exp(2\pi ivg) S_x\{f(g)\} \frac{1-\theta^2}{1+\theta^2+2\theta\cos 2\pi g} dg.$$

We deduce that the spectrum of X_k is given in terms of that of x_t by:

$$S_X(g) = S_x\{f(g)\} \frac{1-\theta^2}{1+\theta^2+2\theta\cos 2\pi g}.$$

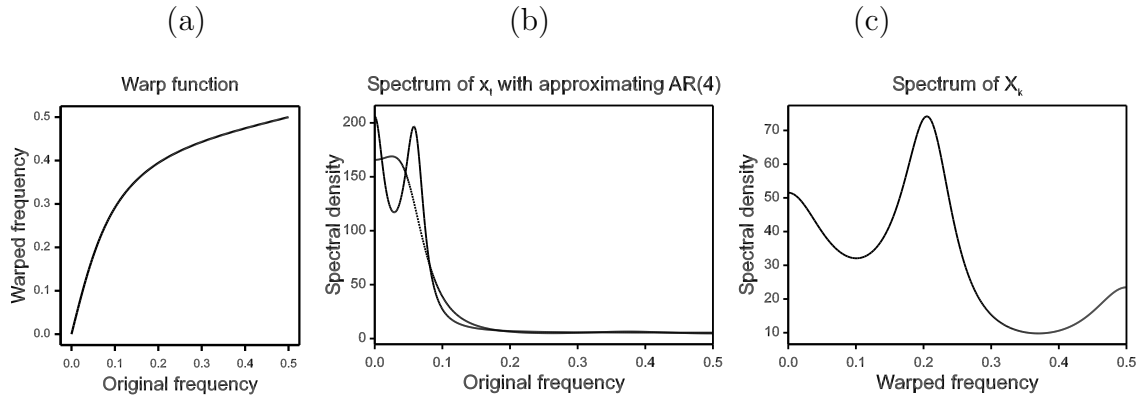

Figure 8: The frequency warp map shown in (a) transforms the bimodal spectrum in (b) of the process x_t into the spectrum in (c) of the process X_k. The dotted line spectrum also shown in (b) is that of a Yule-Walker approximating AR(4) model; a ZAR(4,0.6) approximating spectrum is indistinguishable from that of x_t and is not plotted. The warped spectrum of this ZAR approximation is the Yule-Walker AR(4) approximation to the spectrum in (c) and is again not shown because it is almost indistinguishable from the spectrum of X_k.

Figure 8 illustrates the effect of the frequency warp, shown in Figure 8(a), on transforming the spectrum of a univariate process x_t, shown by the bimodal solid line plot in Figure 8(b), into the (warped) spectrum of the corresponding process X_k, shown in Figure 8(c). The process x_t used in this illustration is constructed as the sum of three independent components, an AR(1) which contributes the spectrum peak at frequency zero, an AR(2) which contributes the spectral peak close to frequency 0.07, and a uniform white noise. The value of $\theta = 0.6$ was used in this illustration. At frequency zero the gradient (3) of the transformation (2) is $(1+\theta)/(1-\theta) = 4.0$, so the frequency range of the main features of the spectrum in Figure 8(b) is expanded by a factor of approximately 4 in Figure 8(c).

If x_t follows the natural ZAR(p, θ) model $\varphi(Z)x_t = n_t$, by applying Z^{-k} we obtain that X_k can be represented by $\varphi(Z)X_k = N_k$ where Z is the shift operator on k and $N_k = Z^{-k}n_t$. Moreover the spectrum of n_t is warped into the spectrum of N_k, which is that of white noise. This representation of X_k is therefore a standard AR(p) model with the same coefficients as the natural ZAR model. The autocovariances of X_k, and hence the covariances between the ZAR states $s_t^{(k)}$ for fixed t, can then be derived as those of a standard AR(p). Furthermore, the approximation of a general process x_t by a natural ZAR(p, θ) model then corresponds exactly to the Yule-Walker approximation of the process X_k with the warped spectrum. This is also illustrated in Figure 8(b) and (c). A Yule-Walker approximating AR(4) model

spectrum for x_t fails to resolve the two peaks in Figure 8(b). An approximating ZAR(4,0.6) model spectrum is, however, so close as to be indistinguishable and is not plotted. The warped spectrum of this ZAR approximation is the Yule-Walker AR(4) approximation to the warped spectrum in Figure 8(c) and is again not shown because it is almost indistinguishable from the spectrum of X_k.

The concept of the frequency warp gives real insight into the capacity of the ZAR model to approximate certain processes much better than the standard AR model, as illustrated in Figure 8. There are two aspects to this, but both apply when the spectrum of x_t is generally confined to low frequencies. The first aspect is that the warp spreads out the low frequency features of the spectrum so that it is more readily approximated by a standard AR spectrum of relatively low order. Looked at in reverse, on the scale of the original frequencies, the ZAR model has much more flexibility for approximation at low frequencies. This is determined by the choice of θ. Secondly, the warp affects the weight applied at different frequencies in approximating the spectrum. It can be shown that fitting the general ZAR model corresponds to minimizing a weighted Whittle criterion on the original scale of frequencies:

$$\int_{-0.5}^{0.5} w(f) \left\{ \log S_m(f) + \frac{S_x(f)}{S_m(f)} \right\} df. \tag{8}$$

Here $S_m(f)$ is the spectrum of the model to be fitted and $S_x(f)$ is the spectrum of the process to be approximated, or the sample spectrum of this process if it is required to fit to a sample series. The weight function $w(f)$ depends only on the choice of the coefficient ρ which defines the future value $Z_\rho^{-1} x_t$ that is to be predicted in fitting the model. It is the spectrum of an AR(1) model with unit variance and coefficient ρ given by

$$w(f) = \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(2\pi f)}.$$

Even for the modest value of $\rho = 0.5$ which approximates a lead time of 3, almost 50% of the weight is attached to frequencies below 0.1, and for $\rho = 0.7$ the relative weight falls to nearly 25% at frequency 0.1. In particular, for a seasonally adjusted monthly time series the parts of the spectrum close to the seasonal harmonic frequencies, that are generally depleted by seasonal adjustment, will carry much less weight in the fitting criterion. An inevitable consequence of using a value of $\rho > 0$ is some loss of efficiency in estimation of a correct ZAR model, but this need not be large. A simple illustration is in the fitting of a standard AR(1) model $x_t = \phi\, x_{t-1} + e_t$, but viewed as a ZAR(1,0.0) model. The coefficient ϕ can be recovered from the coefficent ζ of the regression of x_t on $Z_\rho x_t$ in the general model form, by $\phi = (\zeta + \rho)/(1 + \zeta\rho)$. It is left as an exercise for the reader to show that the relative efficiency of an estimate of ϕ found in this manner is $(1 - \rho^2)/(1 - \rho^2\phi^2)$, which has the value 94% for the quite realistic coefficients $\phi = 0.9$ and $\rho = 0.5$.

5 ZAR model estimation

We present, in outline, three approaches to estimation of the general form of the ZAR(p, θ) model for a process x_t, given observations from $t = 1$ to n. For simplicity we will assume that the series has been mean corrected. A major challenge is to handle the end effects, which are quite evident for ZAR models because the ZAR states $s_t^{(k)} = Z^k\, x_t$ at any given time depend to some extent upon unobserved values of x_t for $t \leq 0$. We will refer to the three approaches as the Yule-Walker, regression and likelihood methods. The estimation methods are a prerequisite for constructing criteria for selecting both the order p and discount factor θ of the model. These criteria will also depend on the value of the discount factor ρ of the general form of model used for estimation. We do not propose any objective criterion for selecting ρ, but suggest that selection of p and θ be carried out for a small number of values of ρ, chosen to investigate the sensitivity of the selected model to values other than zero.

The Yule-Walker method for determining the parameters in the general form of model is to project the variable $Y = Z_\rho^{-1}\, x_t$ on the variables $X_{-k} = Z^k\, x_t$ for $k = 0, 1, \ldots, p-1$. The equations for the coefficients in this projection involve the estimated covariances between these variables. We suggest that these are constructed numerically using frequency domain methods as

$$C_{Y,X,k} = \widehat{\mathrm{Cov}}(Z_\rho^{-1}\, x_t, Z^k\, x_t) = \int_{-0.5}^{0.5} Z_\rho Z^k S_x^*(f) df,$$

and

$$C_{X,v} = \widehat{\mathrm{Cov}}(Z^k\, x_t, Z^{k+v}\, x_t) = \int_{-0.5}^{0.5} Z^v S_x^*(f) df,$$

where

$$S_x^*(f) = (1/n) \left| \sum_{t=1}^{n} x_t \exp(2\pi i f t) \right|^2$$

is the sample spectrum of the observed series. The estimated coefficients are the solution of:

$$C_{Y,X,k} = \zeta_1\, C_{X,k} + \zeta_2\, C_{X,k-1} + \cdots + \zeta_p\, C_{X,k-p+1} \quad \text{for} \quad k = 0, 1, \ldots, p-1,$$

and the variance of the error is

$$S_n^2 = C_{X,0} - \zeta_1 C_{Y,X,0} - \zeta_2 C_{Y,X,1} - \cdots - \zeta_p C_{Y,X,p-1}.$$

Note that in the case of the natural model for which $\rho = \theta$, we obtain $C_{Y,X,k} = C_{X,k+1}$ and the equations are identical in form to the classical Yule-Walker equations. If $\theta = 0$ they reduce to the standard Yule-Walker equations in the usual sample covariances. This spectral approach seemingly overcomes the end effect problem, but in fact it implicitly substitutes zero for the unknown values of x_t for $t < 0$ and $t > n$. This results in estimation bias as in the case of classical Yule-Walker equations, but the bias can be substantially reduced if tapering is used in the construction of the sample spectrum, as in Zhang (1992). We could go further into the properties of the estimates obtained in this way, but our main point is that they are of value as providing rapidly estimated and consistent starting values for estimation by the likelihood method.

For the regression method we first construct a response vector with elements $y_t = Z_\rho^{-1} x_t$ and regression vectors with elements $s_t^{(k)} = Z^k x_t$ by applying the respective operators to x_t taking all unknown values of x_t for $t \leq 0$ and $t > n$ as zero. Thus we use $y_t = x_{t+1} - \rho x_t + \rho y_{t+1}$ for $t = n, n-1, \ldots, 1$, setting x_{n+1} and y_{n+1} to zero to start the recursions. Similarly we use $s_t^{(1)} = x_{t-1} - \theta x_t + \theta s_{t-1}^{(1)}$ for $t = 1, 2, \ldots, n$ setting x_0 and $s_0^{(1)}$ to zero, and similarly for the higher order states $s_t^{(k)}$. We then introduce further regressors to compensate for the transient errors introduced by this treatment of the end effects. The transient error in y_t is $\rho^{n-t}(x_{n+1} + \rho y_{n+1})$, so we introduce the regressor ρ^{n-t} to allow for this. The transient errors in the regression vector $s_t^{(k)}$ depend on $s_0^{(j)}$ for $j \leq k$ and span a space of dimension k. The space for $k = p-1$ contains the space for all lower values of k so that the effect of these transient errors on the prediction also belongs to this space. To allow for these errors we therefore include in the regression a further set of $p-1$ basis vectors of this space, which are easily generated as impulse responses of the unknown series value x_0 on $s_t^{(1)}, \ldots, s_t^{(p-1)}$. If $\rho = 0$ the regressor for the transient error in y_t is simply an indicator for the last time point and if $\theta = 0$ the regressors for the transient errors in the states are simply indicator variables for the time points $t = 1, \ldots, p-1$. These time points are then effectively removed and the effect is exactly the same as in lagged regression where the starting time point is taken as $t = p+1$ so as to include only known values in lagged variables. The regression approach for the ZAR model is therefore a direct generalization of that for lagged regression.

The large sample properties of the estimates are that, under wide conditions, $n^{\frac{1}{2}}(\hat{\zeta} - \zeta) \sim N(0, V_\zeta)$ where V_ζ is defined in terms of the error variance σ_n^2, the variance V_S and covariances $V_{S,k} = \mathrm{Cov}(S_t, S_{t-k})$ of the p dimensional state vector S_t and the transition matrix T:

$$V_\zeta = \sigma_n^2 V_S^{-1} \left(\sum_{k=-\infty}^{\infty} \rho^k V_{S,k} \right) V_S^{-1} = \sigma_n^2 \left\{ V_S^{-1}(I - \rho T)^{-1} + (I - \rho T')^{-1} V_S^{-1} - V_S^{-1} \right\}. \quad (9)$$

This may be consistently estimated from the fitted model parameters. Note that V_ζ reduces to $\sigma_e^2 V_S^{-1}$ for the predictive model with $\rho = 0$.

Lagged regression for the standard AR(p) model does however lose information if the order p is large, and the so-called exact likelihood estimation method has been developed to overcome this. We now generalize this method for the ZAR(p, θ) model and first describe its form in the case $\rho = 0$ for which the likelihood of the observations is what we require. Important to the derivation is the idea that *given* the vector s of initial (unknown) states $s_t^{(0)}, \ldots, s_t^{(p-1)}$ at $t = 0$, the vector $e = e_1, \ldots, e_n$ of innovations may be directly constructed from the predictive form of model after recursively generating the subsequent states from the observations up to time $n-1$. From this the initial state vector s may be estimated, *i.e.* its expected value $\tilde{s}$ found as a linear function of the known observations, and the *series innovations* vector $\tilde{e}$ regenerated using this initial vector. Furthermore, the initial state vector may be transformed to a set of *state innovations* $\tilde{f} = R\tilde{s}$, where R depends only on the model parameters through the natural form. The value of minus twice the log-likelihood, which we will call the deviance, $D(\xi)$, may then be expressed (up to a fixed constant) as

$$D(\xi) = SS/\sigma_e^2 + n \log(\sigma_e^2) - 2 \log |R| + \log |W|$$

where $SS = a'a$ is the sum of squares of a concatenated innovations vector $a = (\tilde{e}, \tilde{f})$, $\sigma_e^2 = \mathrm{Var}(e_t)$ and W is derived from the regression matrix used to estimate the initial state vector. Both R and W are functions of the model parameters only, excluding σ_e^2. The first term in the deviance is the dominant one and maximum likelihood estimation by numerical minimization of the deviance is generally not difficult. One of the advantages of maximum likelihood, shared by the Yule-Walker method, though not by the regression method, is that the estimated parameters are constrained to satisfy the stationarity conditions, because $-2\log|R|$ diverges to infinity at the boundary of the stationarity region.

The extension of this approach to the general form of model is to construct a modified, or quasi-deviance, in which the vector $\tilde{e}$ is replaced by the vector with terms which approximate the prediction error in $Z_\rho^{-1}x_t$ of that model. These are generated as $\tilde{n}_t = \rho\tilde{n}_{t+1} + M\tilde{e}_t$ for $t = n, n-1, \ldots, 1$, where we recall that M is the factor relating the general model errors to the innovations. We take $\tilde{n}_t = 0$ for $t = n+1$ because there is no information about this term in the observed series. However, these modified errors must also be inversely weighted in the sum of squares by their relative standard deviations $d_t = M\{(1-\rho^{2(n-t+1)})/(1-\rho^2)\}^{\frac{1}{2}}$, and the corresponding term $\sum_{t=1}^{n} 2\log d_t$ added to the quasi-deviance. Because ρ is fixed it is actually sufficient just to add the term $2n\log M$ to the deviance.

We call this a quasi-deviance because the error terms in the sum of squares are now correlated. The sum of squares term, as a function of the model parameters, will be asymptotically equivalent to the sum of squares in the regression method. However, besides avoiding the loss of information suffered by the regression approach, the likelihood approach retains the contribution from the initial states which adds the stability constraint on the parameters. The general model parameters ζ_k are not however directly estimated by this approach. Construction of the quasi-deviance is more conveniently implemented using the parameters of the predictive form.

6 ZAR model selection

We will use the quasi-deviance function of the likelihood approach in the selection of the order p and coefficient θ, given a fixed value of ρ in the general form of model. We use the concept of the final prediction error (FPE) in constructing a penalty term to add to the minimized quasi-deviance to form a criterion for selecting p and θ. The FPE requires the calculation of two terms; the *bias correction* or under-estimation of the error variance due to fitting the model and the *excess variance* arising in the error from the use of estimated parameters for out of sample prediction. For the general ZAR(p,θ) model the expected values of these two terms, as a proportion of the variance of the model error n_t, are both equal in large samples of size n, to $n^{-1}b(p,\theta)$ where from (9)

$$b(p,\theta,\rho) = \sigma_n^2 \,\mathrm{tr} V_\zeta V_S = \mathrm{tr}\sum_{-\infty}^{\infty} \rho^{|k|}T^{|k|} = p\frac{(1+\rho\theta)}{(1-\rho\theta)} - 2\rho\frac{(1-\theta^2)}{(1-\rho\theta)^2}\frac{\varphi'(-\tau)}{\varphi(-\tau)}.$$

Here, $\varphi'(Z)$ is the derivative of the natural model form operator $\varphi(Z)$ and, as before, $\tau = (\theta - \rho)/(1 - \rho\theta)$. The relationship between the eigenvalues of T and the zeros of $\varphi(Z)$ are used to derive the final formula, which reduces to the standard value of p when $\rho = 0$. The model selection criterion is then

$$ZIC(p, \theta) = D(\hat{\zeta}) + 2\, b(p, \theta, \rho). \tag{10}$$

In practice we have also applied a modification of Hurvich and Tsai (1989) to $b(p, \theta, \rho)$, which for smaller n gives a slight improvement in accuracy. Figure 9(b) shows a plot of the mean deviance bias of a ZAR(4,θ) model for a range of values of θ. The mean bias is the average from 1000 simulations of a series of length 500, of the difference between the deviance evaluated for the simulated series at the true and estimated model parameters. A value of $\rho = 0.7$ was used for fitting all the models. Also plotted are the mean bias correction evaluated for the fitted model parameters of the simulations and the bias correction evaluated for the true model parameters used in the simulation. These two lines are so close that they are barely distinguishable. Figure 9(a) shows a typical sample series generated by the simulation model with a value of $\theta = 0.9$ with appearance broadly similar to that of the monthly S-A unemployment rate. These simulations reassure us of the accuracy of the bias correction formula for use in model selection.

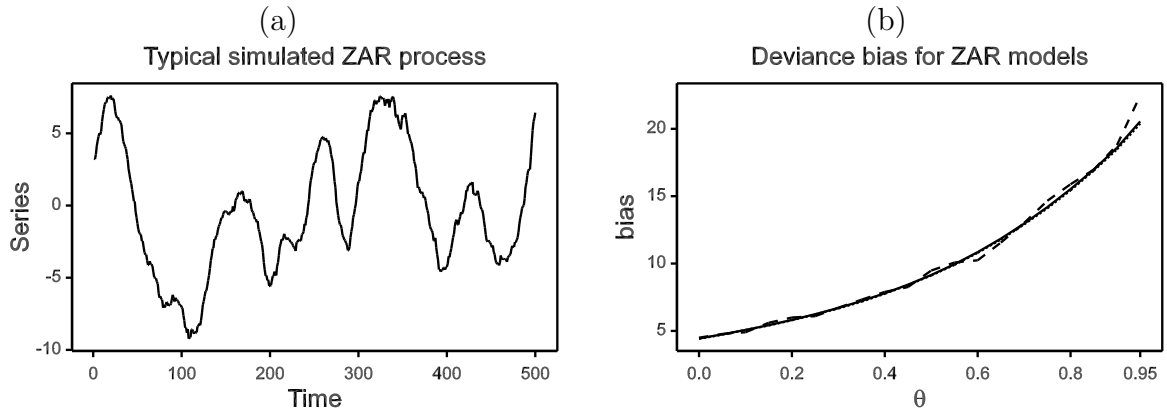

Figure 9: (a) A typical ZAR(4,0.9) series used in simulations with appearance broadly similar to that of the monthly S-A unemployment rate and (b) mean deviance bias from 1000 simulated samples (dashed line) with large sample bias correction calculated from true model parameters (solid line) and the mean of the large sample bias correction calculated from estimated model parameters (dotted line, almost co-incident with solid line). A value of $\rho = 0.7$ and the range of θ shown was used in model simulation and estimation.

The bias correction formula is strictly only applicable to a valid model, *i.e.* one that is fitted with an order at least equal to that of the true model. The second term of the formula is therefore unreliable if evaluated for a model that is fitted with an order less than this. To avoid this difficulty we advocate an initial strategy of selecting an order $p(\theta)$ for each of a range of values of θ using only the first term $p(1 + \rho\theta)/(1 - \rho\theta)$. This is the only term that changes as the order of the fitted model is *increased* from that of the *true* model. For orders less than that of the true model and for sufficiently large sample sizes ZIC will be decreasing with high probability. Figure 10 shows the results of a simulation exercise to illustrate the order selection for a fixed value of θ that was used in the ZAR(4,0.9) model from which the

samples were generated. We investigated a further modification that is illustrated in the figure. This is the use of the extra factor $\log(\log(n))$ as a multiplier of the penalty term in the ZIC, motivated by the criterion of Hannan and Quinn (1979).

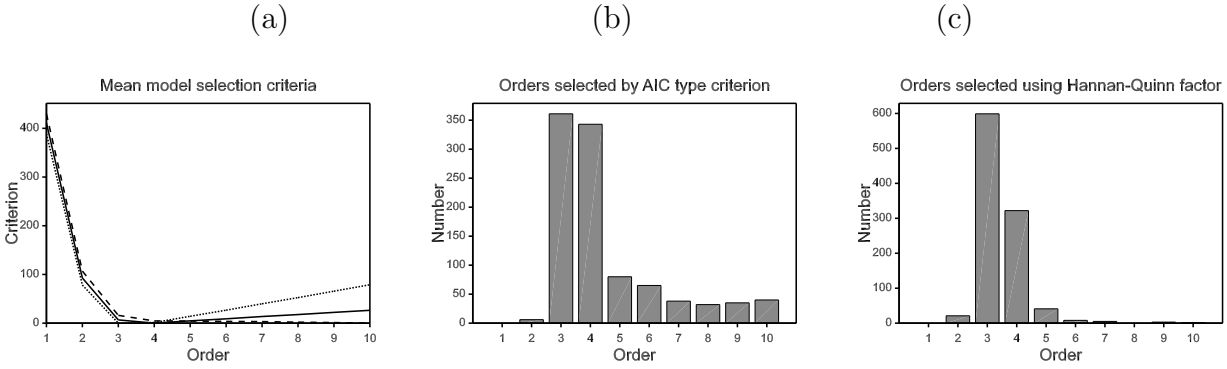

Figure 10: (a) Mean model selection criteria derived from 1000 simulated samples of a ZAR(4,0.9) process of length 500 estimated from the general model form with $\rho = 0.7$: the bias corrected deviance (dashed line), AIC type penalized deviance (solid line) and Hannan-Quinn type penalized deviance (dotted line). (b) The distribution of the model orders selected using the AIC type criterion and (c) The distribution of the model orders selected using the Hannan-Quinn type criterion.

Figure 10(a) shows plots of the mean values from 1000 simulations of the deviance corrected only for the bias, *i.e.* $D(\zeta) + b(p, \theta, \rho)$, the ZIC as in (10) and the ZIC with the Hannan and Quinn (1979) modification, *i.e.* $D(\zeta) + 2\log(\log(n))\, b(p, \theta, \rho)$. The first of these levels off from the true model order $p = 4$, the second has a minimum at the true model order but the third has its minimum at $p = 3$. Figure 10(b) shows the distribution of model orders selected using the ZIC and Figure 10(c) shows the distribution when the Hannan and Quinn (1979) modification is applied. Order 3 is actually most frequently selected by both, which reflects the fact that the fourth order coefficient is quite small and would only be more certainly identified using a much larger sample. The use of the Hannan and Quinn (1979) modification would appear to be advantageous because it greatly reduces the instances of over-estimation of the model order.

The final step of the model selection strategy is to plot the criterion ZIC over the chosen range of θ, using the order of model selected for each of those values. The full bias formula (10) is used for this, applied to the parameters estimated for the selected order at each value of θ. The minimum of this plot is used to select θ. We have not yet carried out any simulation study to support this strategy, but show its outcome for the S-A unemployment series in Figure 11.

As advocated earlier, we applied this procedure for two values of ρ: $\rho = 0$ and $\rho = 0.5$. Figure 11(a) shows the model orders selected over the range of θ for each value of ρ and Figure 11(b) shows the selection criteria over the range of θ. The range of θ is transformed to its equivalent lag $(1+\theta)/(1-\theta)$ so as to understand better the implications of the model. It is striking that for $\rho = 0.5$ the procedure selects models with equivalent lags for θ in the region of 30 to 50 months, with much higher orders. The model with $\theta = 0.94$ (equivalent lag 32) and order $p = 14$ is the one selected for the forecasts in Figure 4(b),(c) and (d). For

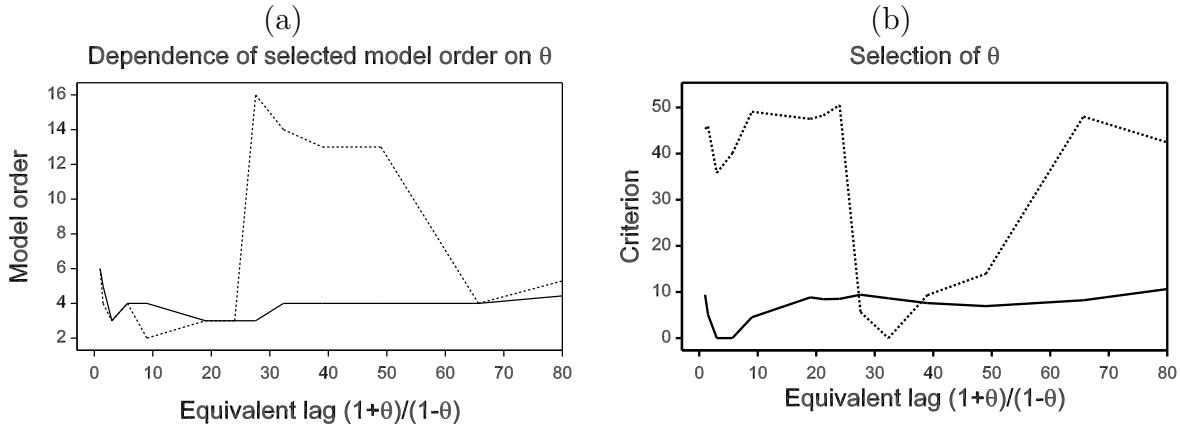

Figure 11: (a) The order p of the ZAR(p, θ) model selected by the Hannan-Quinn type criterion for the monthly unemployment rate series, for a range of values of θ; (b) the value of ZIC evaluated using the selected order and estimated parameters for each of a range of values of θ. The solid line shows results from setting $\rho = 0$ and the dotted line the results for $\rho = 0.5$ in the general model form used for estimation. Each line in (b) is corrected to have its minimum value equal to zero.

model selection we did use the whole series, but all the forecasts were out of sample with the parameters re-estimated from past values alone. For $\rho = 0$ the selected model order remains relatively low over the range of θ with the selected parameters being $p = 3$ at $\theta = 0.5$. These resulted in rapidly mean-reverting forecast functions very close to that for the model with $p = 6$ at $\theta = 0$, shown in Figure 4(a). The ZAR modeling methodology we have set out therefore appears to achieve its objectives in this example, and we hope that others will be encouraged to apply it to their own modeling tasks.

References

Akaike, H. (1973). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AC–19(2):716–723.

Belcher, J., Hampton, J. S., and Tunnicliffe Wilson, G. (1994). Parameterisation of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society Series B*, 56:141–155.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Braccini, C. and Oppenheim, A. (1974). Unequal bandwidth spectral analysis using digital frequency warping. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 22:236–244.

Findley, D. (1983). On the use of multiple models for multi-period forecasting. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association Conference*, pages 528–531. American Statistical Association.

Findley, D. (1985). Model selection for multi-step ahead forecasting. In Barker, H. A. and Young, P. C., editors, *Proceedings of the 7th Symposium on Identification and System Parameter Estimation*, pages 1039–1044, Oxford. Pergamon.

Findley, D. (1990). Making difficult model comparisons graphically. In *Proceedings of the Survey Research Methods Section of the American Statistical Association Conference*, pages 622–627. American Statistical Association.

Findley, D. (1991). Model selection for multistep forecasting. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association Conference*, pages 243–247. American Statistical Association.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society Series B*, 41:190–195.

Haywood, J. and Tunnicliffe Wilson, G. (1997). Fitting time series models by minimizing multistep-ahead errors: a frequency domain approach. *Journal of the Royal Statistical Society Series B*, 59:237–254.

Haywood, J. and Tunnicliffe Wilson, G. (2009). A test for improved multi-step forecasting. *Journal of Time Series Analysis*, 30:682–707.

Holan, S. H., Bell, W. R., and McElroy, T. S. (eds., to appear). *Economic Time Series: Modeling and Seasonality*. Chapman & Hall/CRC, Boca Raton.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Ibañez, J.-C. (2005). *New Tools for Multi-Step Forecasting of Non-linear Time Series*. PhD thesis, Department of Mathematics and Statistics, Lancaster University.

Ibañez, J.-C. and Tunnicliffe Wilson, G. (2007). Multi-stage time series forecasting using Gaussian regression and generalised lags. In Lendasse, A., editor, *European Symposium on Time Series Prediction, ESTSP'07*, pages 49–58. Multiprint Oy / Otamedia. ISBN: 978-951-22-8601-0.

Lo, M. T. (2008). *Improvement in Multi-Step Prediction by a Class of Extended Autoregressive Models*. PhD thesis, Department of Mathematics and Statistics, Lancaster University.

Morton, A. and Tunnicliffe Wilson, G. (2001). Extracting economic cycles using modified autoregressions. *The Manchester School*, 69:574–585.

Morton, A. and Tunnicliffe Wilson, G. (2004). A class of modified high order autoregressive models with improved resolution of low frequency cycles. *Journal of Time Series Analysis*, 25:235–250.

Morton, A. S. (2000). *Spectral Analysis of Irregularly Sampled Time Series Data using Continuous-Time Autoregressions*. PhD thesis, Department of Mathematics and Statistics, Lancaster University.

Partington, J. R. (1997). *Interpolation, Identification and Sampling*. Clarendon Press, Oxford.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

Reale, M. and Tunnicliffe Wilson, G. (to appear). *Models for Dependent Time Series*. Chapman & Hall/CRC, Boca Raton.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:1461–464.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, 8:147–164.

Tukey, J. W. (1986). Sunset salvo. *The American Statistician*, 40(1):72–76.

Tunnicliffe Wilson, G. and Morton, A. (2004). Modelling multiple time series: achieving the aims. In Antoch, J., editor, *Proceedings in Computational Statistics, 2004*, pages 527–538, Heidelberg. Physica Verlag.

Tunnicliffe Wilson, G., Reale, M., and Morton, A. (2001). Developments in multivariate time series modelling. *Estadistica*, 53:353–395.

Wahlberg, B. (1991). System identification using Laguerre filters. *IEEE Transactions on Automatic Control*, 36:551–562.

Wahlberg, B. and Hannan, E. (1993). Parametric signal modelling using Laguerre filters. *The Annals of Applied Probability*, 3:467–496.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Cambridge, New York.

Zhang, H.-C. (1992). Reduction of asymptotic bias of AR and spectral estimators by tapering. *Journal of Time Series Analysis*, 13:451–469.