DYNAMIC REGRESSION IN ARIMA MODELING

Ross Bettinger, Analytical Consultant, Seattle, WA

ABSTRACT

Box-Jenkins time series models that contain exogenous predictor variables are called *dynamic regression* models. These models require analysis of correlations between current and lagged values of response, exogenous inputs, and errors to accurately reflect the relationship between response and predictors.

We discuss the analysis of exogenous variables as they are used in dynamic regression and how PROC ARIMA may be used to incorporate them into transfer functions for time series model building. Time series in this paper are assumed to be stationary and nonseasonal to simplify the discussion, but since the analysis of nonstationary and seasonal time series proceeds according to the standard Box-Jenkins methodology there is no restriction in practice on the use of exogenous predictors.

KEYWORDS

Dynamic regression, ARMA, ARMAX, transfer function, time series, model building, exogenous predictor, autoregressive series, moving average series, Box-Jenkins time series methodology

INTRODUCTION

Time series models differ from ordinary regression models in that the sequence of measurements of a process over time introduces autocorrelations between measured values. The serial nature of the measurements must be addressed by careful examination of the lag structure of the model.

TIME SERIES MODELS

A time series is a set of measurements $\{Y_t\}$ collected sequentially at equally-spaced points in time. Time series models may be characterized in terms of lag structure.

- For an autoregressive model (AR) of order *p*, the response variable, Y_t, is a linear combination of *p* past values and an error term:

  $$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

  The error term, ϵ_t, is assumed to be independent and identically distributed (IID) as $N(0, \sigma^2)$.[1]

- For a moving average model (MA) of order *q*, the response variable is a linear combination of *q* past errors and an AWGN error term:

  $$Y_t = \theta_0 - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \epsilon_t$$

- For an autoregressive moving average model (ARMA) of order (*p*,*q*), the response variable is a linear combination of *p* past responses and *q* past errors with an AWGN error term:

  $$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \epsilon_t$$

REGRESSION MODELS

Regression models may be characterized according to their inclusion of time as a component in the model [1].

- An ordinary regression model has no explicit time component. The response variable is a linear combination of independent predictors and an AWGN error term[2]:

  $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

- The response variable for a time series regression model with independent predictors is a linear combination of independent predictors measured in the same time frame as the response variable:

---

[1] A random variable that is IID as $N(0, \sigma^2)$ is also described as "additive white Gaussian noise", or AWGN. The error term is also called a disturbance or shock term because it perturbs the response in a random, unpredictable manner at the level of an individual observation.

[2] The error term is also called a disturbance or shock term because it perturbs the response in a random, unpredictable manner.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \epsilon_t$$

- A dynamic regression model is a time series model with lagged independent predictors.
  - In the univariate case,

$$Y_t = \beta_0 + \omega_{10} X_{1,t} + \omega_{11} X_{1,t-1} + \omega_{12} X_{1,t-1} + \cdots + \omega_{1m} X_{1,t-m} + \epsilon_t$$

  - In the multivariate case,

$$\begin{aligned} Y_t = \beta_0 &+ \omega_{10} X_{1,t} + \omega_{11} X_{1,t-1} + \omega_{12} X_{1,t-1} + \cdots + \omega_{1m} X_{1,t-m} \\ &+ \omega_{20} X_{2,t} + \omega_{21} X_{2,t-1} + \omega_{22} X_{2,t-1} + \cdots + \omega_{2m} X_{2,t-m} \\ &+ \cdots \\ &+ \omega_{k0} X_{k,t} + \omega_{k1} X_{k,t-1} + \omega_{k2} X_{k,t-1} + \cdots + \omega_{km} X_{k,t-m} \\ &+ \epsilon_t \end{aligned}$$

THE BACKSHIFT OPERATOR

The backshift operator, $B(Y_t)$, has the property $B(Y_t) = BY_t = Y_{t-1}$. It is useful in expressing time series models in polynomial form. Applying it to the AR(*p*) model, we have:

$$Y_t = \phi_0 + \phi_1 B Y_t + \phi_2 B^2 Y_t + \cdots + \phi_p B^p Y_t + \epsilon_t$$

The AR(*p*) model in polynomial form becomes

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = \epsilon_t$$

and the MA(*q*) model in polynomial form becomes

$$Y_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \epsilon_t$$

THE LINEAR FILTER

A linear filter describes a process by which a univariate time series $\{X_t\}$ is transformed into another time series Y_t:

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i X_{t-i}$$

for $t = \cdots, -1, 0, 1, \cdots$. A filter is linear if it has the following properties [2]:

- It is time-invariant, e.g., the coefficients $\{\psi_i\}$ do not depend on time.
- It is causal if ψ_i=0 for $i < 0$, e.g., it is physically realizable because Y_t depends only on current and past values of X_t.
- It is stable if $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$, e.g., the sum of the coefficients is bounded and hence finite.

Expressing the linear filter in terms of the backshift operator, we have

$$Y_t = \sum_{i=-\infty}^{+\infty} \psi_i X_{t-i} = \Psi(B)\, X_t$$

where $\Psi(B) = \sum_{i=-\infty}^{\infty} \psi_i B^i$ is called a *transfer function*.

MODEL IDENTIFICATION

A time series model is described in terms of its lag structure and the estimates of the lagged coefficients of the model, ϕ_i for AR models, and θ_j for MA models. A model is identified by examining the relationship between the response, Y_t, and a realization (observed value) of Y_t, denoted as y_t.We define the covariance at lag *k* between two time series separated by a delay of *k* time units as

$$\gamma_k = \mathrm{Cov}(y_t, x_{t-k}) = E[(y_t - \mu_y)(x_{t-k} - \mu_x)]$$

When $x_t = y_t$, the set of the values of γ_k, *k* = 0, 1, 2, ⋯ is called the autocovariance function.

THE AUTOCORRELATION FUNCTION

The autocovariance of γ_k at $k = 0$ is the variance of the time series: $\gamma_0 = \sigma^2$. Then the autocorrelation coefficient at lag *k* is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{E[(y_t - \mu)(y_{t-k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]\, E[(y_{t-k} - \mu)^2]}} = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$

Analogous to the autocovariance function, the set of the values $\{\rho_k\}$ is called the autocorrelation function (ACF). When the ACF is computed from samples of a time series of finite length, it is an estimate of the true ACF and is called the sample ACF.

The ACF may be used to identify the order of a MA(*q*) process because the ρ_k diminish to 0 after a finite number of lags. Statistically speaking, the ρ_k become insignificantly different from 0 after *q* lags.

THE CROSSCORRELATION FUNCTION

The crosscovariance function is

$$\gamma_{xy}(j,k) = Cov(x_j, y_k) = E[(x_j - \mu_x)(y_k - \mu_y)]$$

and

$$Cov[x_t, y_{t-j}] = \gamma_{xy}(j)$$

which depends only on *j* = 0, 1, 2, ⋯.

The crosscorrelation function (CCF) is the set of values of $\gamma_{xy}(j)$, *j* = 0, 1, 2, ⋯

$$\rho_{xy}(j) = Corr(x_t, y_{t-j}) = \frac{\gamma_{xy}(j)}{\sqrt{\gamma_x(0)\gamma_y(0)}}$$

When the CCF is computed from a finite sample of a time series, it is called the sample CCF. The lag structure of the CCF may be used to identify the exogenous time series.

THE PARTIAL AUTOCORRELATION FUNCTION

The partial autocorrelation function (PACF) is based on the concept of partial correlation. It is derived from the Yule-Walker equations for an AR model and is described at length in [2]. Briefly stated, the PACF represents the correlation between Y_t and Y_{t-k} after removing the influence of prior lagged values of $Y_{t-1}, Y_{t-2,} \cdots, Y_{t-k+1}$.

The PACF may be used to identify the order of an AR(*p*) model using a similar interpretation as the ACF function for a MA(*q*) process because the $\phi_k$become (statistically) insignificantly different from 0 after *p* lags.

THE INVERSE AUTOCORRELATION FUNCTION

The inverse autocorrelation function (IACF) is the dual of the ACF.[3] It is used to identify the order of an AR model and is similar to the PACF in that regard. For an MA(*q*) model, we attempt to identify the ACF and expect the ρ_k to produce values that are not significantly different from 0 after lag *q*. For an AR model, the PACF may show a pattern of decaying exponential factors and damped sinusoidal ϕ_k with statistically insignificant values of ϕ_k after lag *p*.

The IACF is the dual of the ACF in that for an AR(*k*) model,

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_k Y_{t-k} + \epsilon_t$$

and the dual MA(*k*) model is

$$Y_t = \theta_0 + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_k \epsilon_{t-k} + \epsilon_t$$

---

[3] "Let W_t be generated by the ARMA(p,q) process $\Phi(B)Z_t = \Theta(B)a_t$ where a_t is a white noise sequence. If $\Theta(B)$ is invertible (that is, if Θ considered as a polynomial in B has no roots less than or equal to 1 in magnitude), then the model $\Theta(B)Z_t = \Phi(B)a_t$is also a valid ARMA(q,p) model. This model is sometimes referred to as the dual model." See reference [3]. That the PACF has oppositely-signed spikes from the IACF is due to the definitions of the two functions. Consider the AR(p) process $X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \epsilon_t$. Its IACF is of the same form as the corresponding inverse MA(p) process, $X_t = \epsilon_t - \sum_{i=1}^{p} \alpha_i\, \epsilon_{t-i}$ (Personal communication from Donna Woodward, SAS Technical Support, to the author).

We may use the IACF for an AR model as if it were an ACF for an MA model and then interpret the results as they apply to the AR model.

The following table from [2] demonstrates the duality between ACF and PACF.

Model	**ACF**	**PACF**
MA(*q*)	Cuts off after lag *q*	exponential decay and/or damped sinusoid
AR(*p*)	exponential decay and/or damped sinusoid	Cuts off after lag *p*
ARMA(*p*,*q*)	exponential decay and/or damped sinusoid	exponential decay and/or damped sinusoid

Table 1: Behavior of Theoretical ACF and PACF for Stationary Process

THE TRANSFER FUNCTION-NOISE MODEL

If we add AWGN to a causal linear filter with a finite time delay *k*, we have

$$Y_t = \sum_{i=0}^{k} \psi_i X_{t-i} + N_t = \psi(B)\, X_t + \theta(B)\epsilon_t$$

which can be reformulated as

$$Y_t = \frac{\omega(B)}{\delta(B)} X_t + \frac{\theta(B)}{\phi(B)} \epsilon_t$$

for some appropriate lag structures $\{\delta_i\}$ and $\{\phi_i\}$. We may obtain initial estimates $\widehat{\psi_t}$ by using OLS regression but there is a significant risk of producing inaccurate estimates if the exogenous predictors $X_{t,i}$ are strongly autocorrelated or crosscorrelated. Since the $X_{t,i}$ are time series, and since the processes measured in the context of a given problem are seldom independent of each other, this risk must be mitigated through a process called *prewhitening.* Prewhitening will be discussed below.

THE ARMA TRANSFER FUNCTION

The ARMA(*p*,*q*) model in polynomial form is

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\epsilon_t$$

where ϵ_t is an AWGN process. Then we can write the ARMA(*p*,*q*) model as

$$\Phi(B) Y_t = \Theta(B)\epsilon_t$$

so that

$$Y_t = \frac{\Theta(B)}{\Phi(B)} \epsilon_t$$

and we see that the response variable, Y_t, has been expressed in terms of the ratios of the numerator MA polynomial and the denominator AR polynomial. The rational polynomial $\Theta(B)/\Phi(B)$ is a transfer function. It describes the relationship of the response variable, Y_t, to lagged values of Y_t and lagged values of disturbance terms or shocks, ϵ_t. The transfer function of a time series is analogous to the coefficient $\beta_1$in the time series regression model $Y_t = \beta_0 + \beta_1 X_{1t} + \epsilon_t$. Advantages of expressing an ARMA(*p,q*) model as a transfer function are, from [1]:

- The ratio of the numerator and denominator polynomials of orders *q* and *p*, respectively, is a finite order approximation to the ratio of two infinite-order polynomials.
- The ratio of polynomials requires the estimation of fewer parameters than a pure numerator polynomial (MA model).
- The length of time series in the dynamic regression formulation avoids estimating a large number of parameters.
- The rational transfer function formulation of an ARIMA model reduces the limitation of finite data.

These approximations introduce errors in models for the sake of ease of solution, which is why G. E. P. Box said that “[E]ssentially, all models are wrong, but some are useful.”

THE ARMAX[4] TRANSFER FUNCTION

The ARMAX time series model is a dynamic regression model such that the response variable is a function of exogenous inputs as well as lagged values of Y_t and lagged values of disturbance terms. Its form for a single exogenous predictor is

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)(Y_t - \sum_{i=0}^{m} \omega_i X_{t-i}) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\epsilon_t$$

or

$$\Phi(B)(Y_t - \sum_{i=0}^{m} \omega_i X_{t-i}) = \Theta(B)\epsilon_t$$

so that the underlying regression structure is revealed [5]:

$$Y_t = \boldsymbol{x}_t'\boldsymbol{\omega} + \frac{\Theta(B)}{\Phi(B)}\epsilon_t$$

where $\boldsymbol{x}_t' = [\, x_t, x_{t-1}, x_{t-2}, \cdots, x_{t-m}]$.

PREWHITENING

If we assume that X_t can be modelled as an ARMA process, then

$$\phi_x(B)X_t = \theta_x(B)\eta_t$$

where η_t is an AWGN process. Then

$$\eta_t = \theta_x(B)^{-1}\phi_x(B)X_t$$

and we have expressed the noise as a function of a *linear filter* applied to the exogenous input series X_t. A properly-designed prewhitening filter $\theta_x(B)^{-1}\phi_x(B)$ reduces the X_t to a white noise time series. In the case that there are several exogenous input series, each series must be prewhitened independently of all others.

ARMAX MODEL BUILDING

The ARMA model can be used to represent stationary processes for the response variable Y_t. If a process depends on exogenous inputs, the ARMA model can be extended to include them by using transfer functions. These models are called ARMAX models and can be built using the Box-Jenkins methodology.

Creating an ARMAX model follows the same steps as for an ARMA model. The Box-Jenkins methodology of *identification* of a suitable model, *estimation* of parameters, and *forecasting* future values applies to ARMAX models as well as ARMA models. There is an additional step added to remove autocorrelations among exogenous variables. This step is called *prewhitening* and is necessary when an exogenous predictor is autocorrelated. Since exogenous predictors are time series in themselves, the CCF of the response variable and the exogenous predictor is confounded by the lag structure of the exogenous predictor. The exogenous series are transformed into white noise to remove autocorrelations, then the CCF is computed to reveal the lag structure of Y_t as a function of prewhitened X_t.

EXAMPLE: A SINGLE EXOGENOUS PREDICTOR

This example of using PROC ARIMA to identify a dynamic regression model is taken from [4]. A stochastic process Y is described by the model

$$Y_t - 10 - .8(Y_{t-1} - 10) = 3[(X_{t-2} - 5) - .4(X_{t-3} - 5)] + N_t$$

where $N_t \sim AWGN$. The model can be written as

$$Y_t - \mu_y - .8(Y_{t-1} - \mu_y) = 3[(X_{t-2} - \mu_x) - .4(X_{t-3} - \mu_x)] + N_t$$

where

$$X_t - \mu_x = .5(X_{t-1} - \mu_x) + \epsilon_t$$

generates values of the series $\{X_t\}$, which is an AR(1) model.

---

[4] ARMAX models are ARMA models that contain e**X**ogenous inputs.

We can express the transfer function as

$$Y_t = 10 + \frac{3[(1-.4B)(X_{t-2}-5)]}{(1-.8B)} + \frac{1}{(1-.8B)}N_t$$

Figure 1 shows the first 501 observations of Y_t.

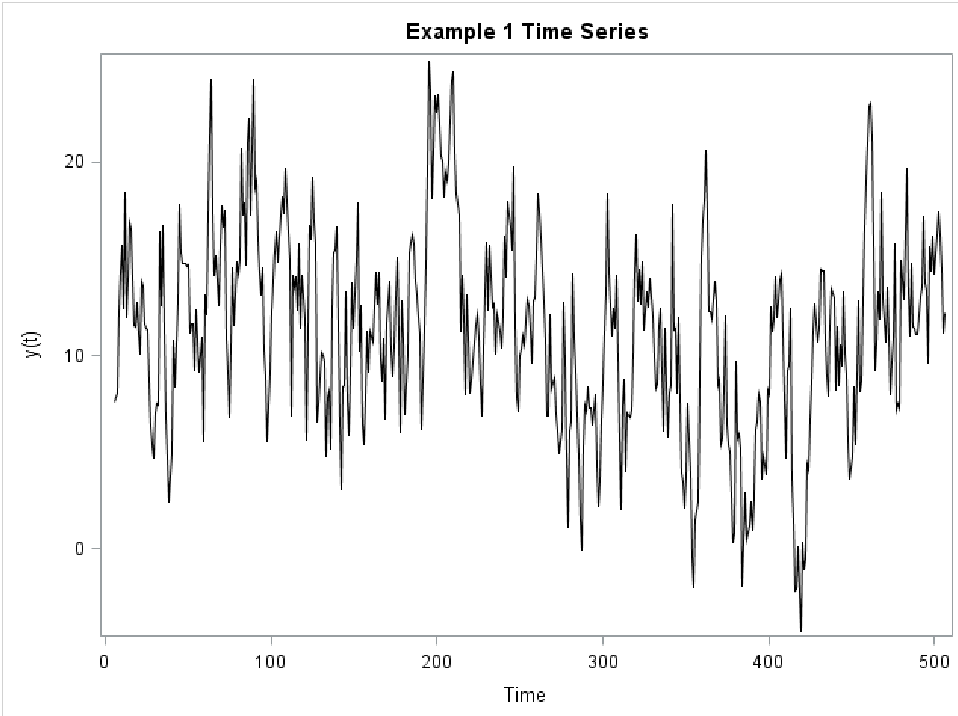

Figure 1: Example 1 Time Series

First, we must identify the X_t series so that we can *prewhiten* the series prior to using X_t to identify Y_t. The code

```
proc arima data=example1 ;
    identify var=x_t ; run ;
```

produces results for autocorrelation and identification of X_t.

Autocorrelation Check for White Noise									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	98.59	6	<.0001	0.394	0.178	0.067	-0.003	0.014	0.060
12	113.81	12	<.0001	0.108	0.102	0.072	0.004	-0.022	0.044
18	124.22	18	<.0001	0.068	0.113	0.031	-0.039	-0.002	0.013
24	144.87	24	<.0001	0.060	0.079	0.096	0.124	0.067	0.022

Table 2: Autocorrelation Check for White Noise, X_t

There is significant autocorrelation at lag 1, which is to be expected since X_t is an AR(1) model. The autocorrelations at higher lags are statistically significant but negligible in magnitude.

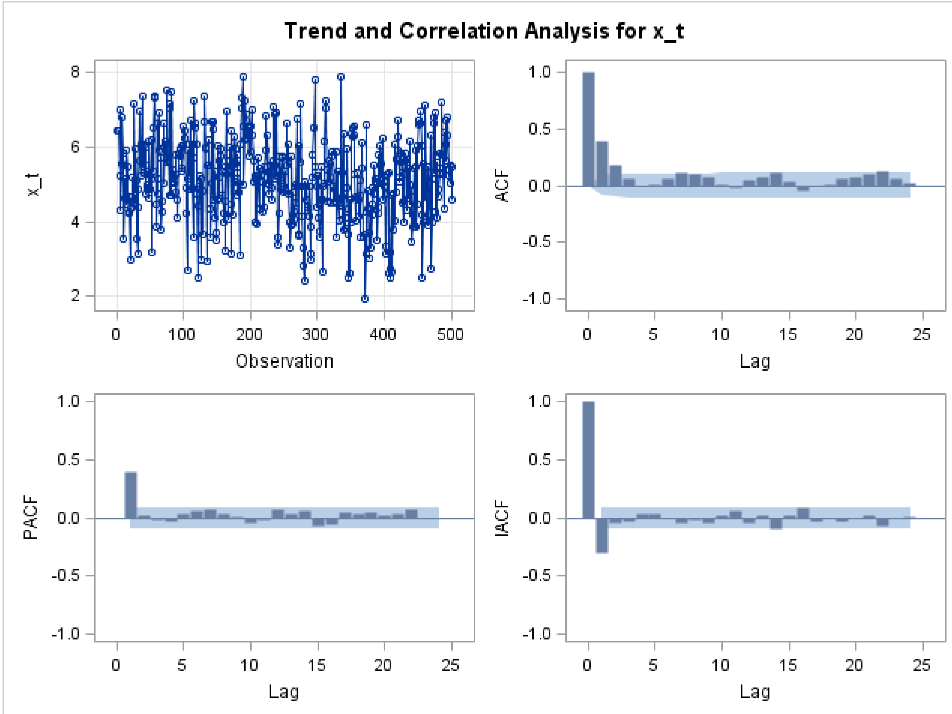

Figure 2: Trend and Correlation Analysis for x_t

The ACF shows a mixture of pure exponential decay and sinusoidal damping. It suggests that errors could be modelled by an MA(2) process. The PACF and IACF jointly suggest that an AR(1) process will suffice. The opposite sign of the PACF and IACF is expected.

Since we value parsimony of description, we will fit an AR(1) model to Y_t and see if it suffices. We submit the following estimation statement

```
estimate p=1 ml plot ; run ;
```

to fit an AR(1) model to the X_t series using the method of maximum likelihood to estimate the parameters.

For reasons of brevity, the results are summarized. The parameter estimates, μ_x and ϕ_1, are 5.13575 and .39481, respectively. They are both significant at the p<.0001 level. The autocorrelation check of residuals indicates no significant autocorrelations at any lag. Figure 3 below indicates, by the absence of any significant lags, that the AR(1) model suffices to convert the series X_t to white noise, as was to be desired. Hence, there is no need to consider a MA model.

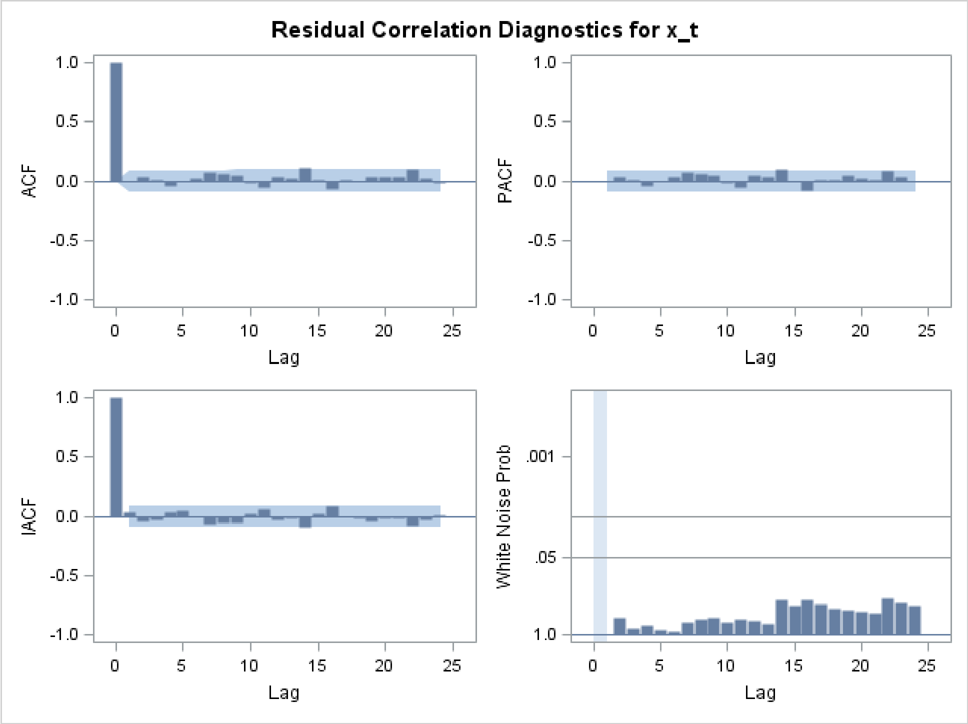

Figure 3: Residual Correlation Diagnostics for x_t

The next step in the process is to identify Y_t. We submit the following identification statement:

```
identify var=y_t crosscorr=( x_t ) nlag=24 ; run ;
```

The `crosscorr` option indicates that X_t is an exogenous predictor that is to be prewhitened and whose residuals are to be crosscorrelated with the residuals of the prewhitened series Y_t. The crosscorrelations of the prewhitened series are shown in Table 3.

Crosscorrelation Check Between Series

To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	422.43	6	<.0001	0.012	-0.015	0.792	0.294	0.298	0.201
11	467.96	12	<.0001	0.120	0.109	0.113	0.137	0.132	0.125
17	484.74	18	<.0001	0.061	0.021	0.056	0.050	0.127	0.087
23	489.11	24	<.0001	-0.001	0.032	0.011	0.045	0.054	0.052

Table 3: Crosscorrelation Check Between Series

We see significant crosscorrelation $\rho_{xy}(2)$ =0.792. The ACF of the prewhitened Y_t indicates exponential decay until lag 10, the PACF indicates an AR(1) model with a shift of 2 time periods, and the IACF indicates an AR(1) model.

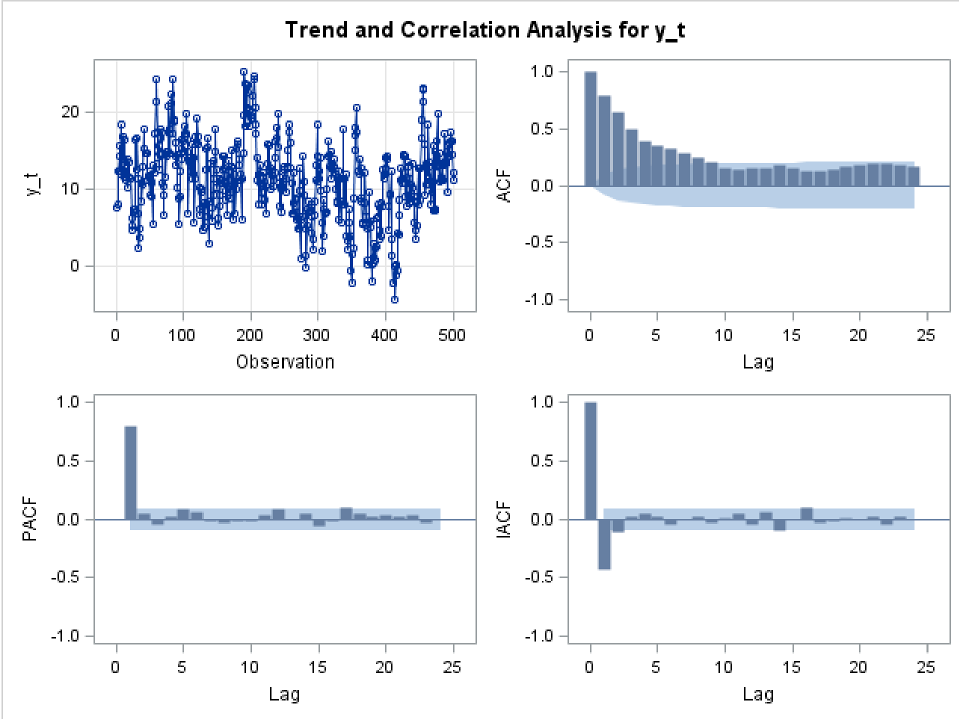

Figure 4: Trend and Correlation Analysis for y_t

The crosscorrelation plot of the prewhitened series is shown in Figure 5,

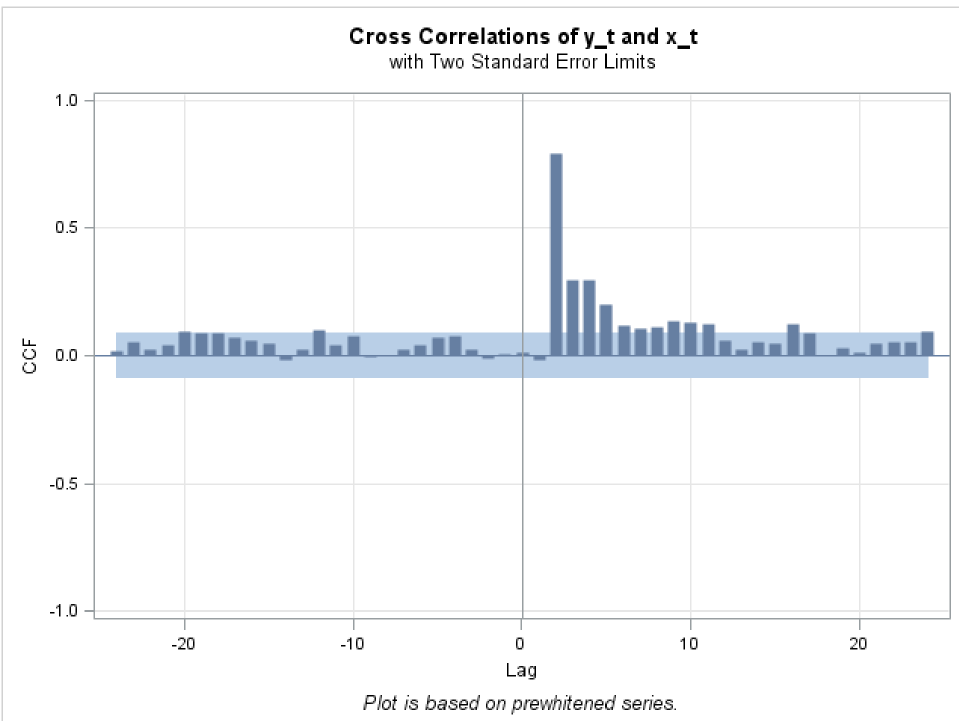

Figure 5: Cross Correlations of y_t and x-t

and it suggests that Y_t is influenced by X_{t-2}, which indicates a delay of two time periods. The model appears to have both an AR component and an MA component, so let us initially estimate it as an ARMA(1, 1) model with a delay of two time periods and test the adequacy of fit. We submit the statement

```
estimate p=1 q=1 input=( 2 $ ( 1 )/( 1 ) x_t ) ml plot ; run ;
```

We summarize the results: all of the parameters estimated were significant at the *p*<.0001 level, the residuals showed no significant autocorrelation and were not significantly different from white noise. They were approximately normally distributed. The crosscorrelation of the residuals with the exogenous predictor X_t indicated no significant lag structure. We conclude that the ARMA(1,1) model is adequate to explain the behavior of Y_t dynamically regressed on X_t.

The model, in transfer function form, is

$$Y_t = -30.2975 + 2.910951 \frac{(1 - .38739B)}{(1 - .77905B)} X_{t-2} + \frac{1}{(1 - .822B)} \epsilon_t$$

where

$$(1 - .39481B)X_t = \epsilon_t$$

A plot of the values of Y_t for $t \geq 450$ is shown in Figure 6. We note the range of variability in the forecast values where $t \geq 502$. Even though we identified the series Y_t, we still must accept the imprecision of the forecasts.

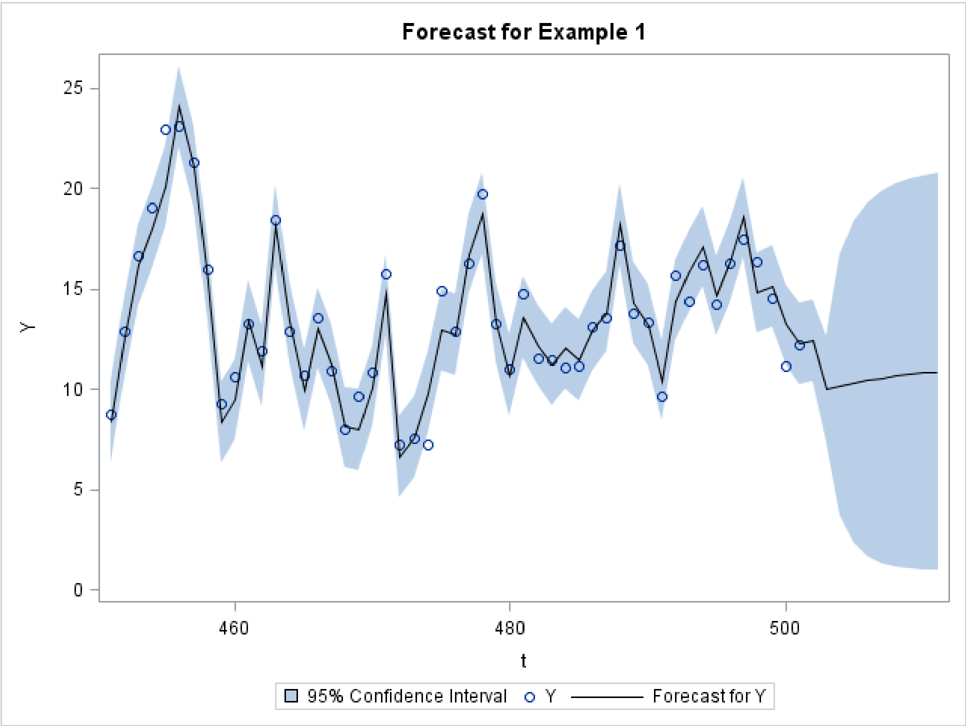

Figure 6: Forecast for Example 1

SUMMARY

We have briefly reviewed the Box-Jenkins time series modeling methodology, with emphasis on prewhitening exogenous predictors used in dynamic regression. While the methodological steps of model identification, parameter estimation, and forecasting are the same for ARMAX as well as ARMA modeling, the potential complications of confounding due to autocorrelations amongst exogenous predictors are removed by prewhitening them before the identification of the response series is attempted.

We applied the ARMAX modeling strategy to a known model and discussed each phase of the modeling process. We observed equivalent results to previously-published work and discussed each step in detail.

REFERENCES

1. Dickey, David and Terry Woodfield (2011). *Forecasting using SAS® Software: A Programming Approach*, SAS Institute, Inc. Cary, NC, USA.
2. Montgomery, Douglas C., Cheryl L. Jennings, and Murat Kulahci (2008). *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons, Inc. Hoboken, NJ.
3. http://support.sas.com/documentation/cdl/en/etsug/65545/HTML/default/viewer.htm#etsug_arima_details01.htm
4. Brocklebank, John C. and David A.Dickey (1987). *Transfer Functions Using PROC ARIMA*, SUGI 12, SAS Institute, Inc, Cary, NC.
5. Hyndman, Rob J. (2010). *The ARIMAX Model Muddle,* http://www.r_bloggers.com/the-arimax-model-muddle.

BIBLIOGRAPHY

Brocklebank, John C., and David A. Dickey (2003). *SAS(R) for Forecasting Time Series, 2nd Ed,* SAS Institute, Inc., Cary, NC, USA.

ACKNOWLEDGMENTS

We thank Dr. David Dickey, Samantha Dalton, Sam Iosevich, Joy Joseph, Dr. Joseph Naraguma, and Dr. Kenneth Sanford, who reviewed preliminary versions of this paper and contributed their helpful comments.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ross Bettinger

rsbettinger@gmail.com