

# Forecasting with ARMA Models

**Sanjiv Jaggia**

California Polytechnic State University, San Luis Obispo, CA

*Increasing awareness and data availability has made statistical analysis an integral part of managerial decision making. Forecasting, in particular, has become an essential skill for all financial analysts, economists, marketing and accounting professionals. Arguably all business students should be educated in forecasting techniques with an emphasis on ARMA modeling. While statistical packages such as Eviews have increasingly become user friendly, the discussion and the implementation of the important diagnostic analysis of ARMA models in popular undergraduate textbooks has not been totally satisfactory. This paper succinctly outlines this important diagnostic step for ARMA modeling with Eviews, using monthly data on 3-month government Treasury bills rate for illustration. The presentation is accessible to readers with an intermediate level of statistics.*

Keywords: xxx

## Introduction

Forecasting is an important aspect of statistical analysis that provides guidance for decisions in all areas of business. It is important to be able to make sound forecasts for variables such as sales, production, inventory, interest rates, exchange rates, real and financial asset prices for both short and long term business planning. In essence, the success of any business or government organization depends on the ability to accurately forecast its revenue and expenditure. Auto-regressive moving average (ARMA) models provide a unifying framework for forecasting. These models are aided by the abundance of high quality data and easy estimation and evaluation by statistical packages.

There are numerous applications where researchers have employed basic as well as advanced ARMA models for forecasting. While a seasoned researcher is well versed in these models, the details and easy to follow steps of AR-

MA modeling, which may be invaluable to a business student or a practitioner, are not available. For instance, popular textbooks in econometrics, including *Using Econometrics* (2011) by Studenmund and *Introductory Econometrics: A Modern Approach* (2009) by Jeffrey Wooldridge provide no formal discussion of ARMA models. While specialized books such as *Elements of Forecasting* (2008) by Francis Diebold and *Business Forecasting* (2009) by Hanke and Wichern discuss ARMA models, their suggested diagnostic analysis of the models has not been comprehensive.<sup>1</sup> For instance, these books do not provide a unified treatment of in-sample and out of sample analysis of competing models. This paper succinctly outlines the important step of diagnostic analysis of ARMA models with Eviews, using monthly data on 3-month government Treasury bills rate for illustration.

---

<sup>1</sup> Advanced textbooks such as Enders (2005) do the job properly.

There are a number of useful issues that students and practitioners must understand regarding the analysis of ARMA models. First, an intuitive understanding of stationarity is essential since these models can only be identified for stationary series. Informal graphs can be used to look for departures from the constancy of the mean and the variance. We generally difference the series to achieve constancy of the mean. However, if both the mean and the variance are time-varying, we also need a dampening or a pre-differencing transformation. This simple explanation is left out in the textbooks by Diebold as well as Hanke and Wichern. Since economic theory provides no guidance on transformations, the natural log or square root pre-transformations may produce better forecasts. Students and practitioners must be advised to identify several ARMA models with various transformations and use model selection criteria to pick the most appropriate model.

Another issue that has not been explicitly stated is a consistent definition of the residuals that works for all models, with or without pre-transformations. Statistical packages usually store the residuals of the estimated models on the basis of pre-transformations. Therefore, the model selection measures that emerge from such residuals are not comparable.<sup>2</sup> Finally, since forecasting is essentially an out-of-sample problem, competing models must also be compared on the basis of their performance outside the sample period. This important analysis of the out of sample forecast errors based on static and dynamic methods has been left out in popular textbooks.

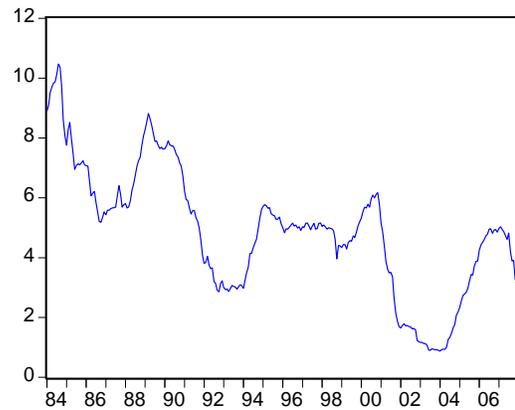
**Model Identification**

Let  $y_1, y_2, \dots, y_T$  represent a sample of T observations of a variable of interest  $y$ . Consider a time series  $\{y_t\}$  representing the monthly interest rate of 3-month Treasury bills for a sample period of January 1984 to December 2007 (see Table 1). This application is used throughout the paper to motivate and implement ARMA forecasting models.

Since stationarity is essential for the identification of an ARMA model, the first step is always to test for stationarity of the underlying series. While many macroeconomic, business, and financial series are not stationary, the series can easily be made stationary by differencing with or without pre-transformations. Formally,  $\{y_t\}$  is said to be

stationary if the mean,  $E(y_t) = \mu$ , the variance,  $Var(y_t) = E(y_t - \mu)^2 = \sigma^2$ , and the covariance,  $Cov(y_t, y_{t-s}) = E(y_t - \mu)(y_{t-s} - \mu) = \gamma_s$ , are all stable over time. For instance, for the series to be stationary, it must not exhibit any stochastic trend (changing mean) or varying volatility (changing variance). Whereas the stochastic trend is generally removed by first differencing of the series, the varying volatility can be reduced with a natural log or square root pre-differencing transformation. As mentioned earlier, the important discussion of pre-transformations is left out in the above textbooks although they include the applications of growth rates based on log transformations.

A simple plot of the series is often used as an informal test for stationarity. The series is considered stationary if the level and the spread do not exhibit any visual changes over time. The plot of the autocorrelation function also serves as another visual test for stationarity.<sup>3</sup> The series is considered stationary if the ACF cuts off, or dies down, fairly quickly. A plot of the monthly interest rate series in Figure 1 indicates non-stationarity as it exhibits an obvious trend.



**Figure 1.** 3-month Treasury Bill Rate

The differenced series specified in terms of a lag operator  $L$ ,  $y_t - y_{t-1} = (1 - L)y_t$ , is plotted in Figure 2. As is often the case with graphical tests, it is not clear whether or not the variance of the differenced series is time invariant. In such situations, it is recommended to also consider pre-transformations and use model selection criteria to determine the most appropriate model for forecasting.

<sup>2</sup> In the regression context, it is analogous to spuriously comparing a linear with a log-linear model on the basis of the computer generated  $R^2$ .

<sup>3</sup> In Eviews, choose 'Correlogram' from 'Series Statistics' to get the autocorrelation (ACF) and the partial autocorrelation (PACF) functions of the series.

**Table 1. 3-Month Treasury Bill Rate**

Date	Tbill										
Jan-84	8.90	Jan-88	5.81	Jan-92	3.80	Jan-96	5.00	Jan-00	5.32	Jan-04	0.88
Feb-84	9.09	Feb-88	5.66	Feb-92	3.84	Feb-96	4.83	Feb-00	5.55	Feb-04	0.93
Mar-84	9.52	Mar-88	5.70	Mar-92	4.04	Mar-96	4.96	Mar-00	5.69	Mar-04	0.94
Apr-84	9.69	Apr-88	5.91	Apr-92	3.75	Apr-96	4.95	Apr-00	5.66	Apr-04	0.94
May-84	9.83	May-88	6.26	May-92	3.63	May-96	5.02	May-00	5.79	May-04	1.02
Jun-84	9.87	Jun-88	6.46	Jun-92	3.66	Jun-96	5.09	Jun-00	5.69	Jun-04	1.27
Jul-84	10.12	Jul-88	6.73	Jul-92	3.21	Jul-96	5.15	Jul-00	5.96	Jul-04	1.33
Aug-84	10.47	Aug-88	7.06	Aug-92	3.13	Aug-96	5.05	Aug-00	6.09	Aug-04	1.48
Sep-84	10.37	Sep-88	7.24	Sep-92	2.91	Sep-96	5.09	Sep-00	6.00	Sep-04	1.65
Oct-84	9.74	Oct-88	7.35	Oct-92	2.86	Oct-96	4.99	Oct-00	6.11	Oct-04	1.76
Nov-84	8.61	Nov-88	7.76	Nov-92	3.13	Nov-96	5.03	Nov-00	6.17	Nov-04	2.07
Dec-84	8.06	Dec-88	8.07	Dec-92	3.22	Dec-96	4.91	Dec-00	5.77	Dec-04	2.19
Jan-85	7.76	Jan-89	8.27	Jan-93	3.00	Jan-97	5.03	Jan-01	5.15	Jan-05	2.33
Feb-85	8.27	Feb-89	8.53	Feb-93	2.93	Feb-97	5.01	Feb-01	4.88	Feb-05	2.54
Mar-85	8.52	Mar-89	8.82	Mar-93	2.95	Mar-97	5.14	Mar-01	4.42	Mar-05	2.74
Apr-85	7.95	Apr-89	8.65	Apr-93	2.87	Apr-97	5.16	Apr-01	3.87	Apr-05	2.78
May-85	7.48	May-89	8.43	May-93	2.96	May-97	5.05	May-01	3.62	May-05	2.84
Jun-85	6.95	Jun-89	8.15	Jun-93	3.07	Jun-97	4.93	Jun-01	3.49	Jun-05	2.97
Jul-85	7.08	Jul-89	7.88	Jul-93	3.04	Jul-97	5.05	Jul-01	3.51	Jul-05	3.22
Aug-85	7.14	Aug-89	7.90	Aug-93	3.02	Aug-97	5.14	Aug-01	3.36	Aug-05	3.44
Sep-85	7.10	Sep-89	7.75	Sep-93	2.95	Sep-97	4.95	Sep-01	2.64	Sep-05	3.42
Oct-85	7.16	Oct-89	7.64	Oct-93	3.02	Oct-97	4.97	Oct-01	2.16	Oct-05	3.71
Nov-85	7.24	Nov-89	7.69	Nov-93	3.10	Nov-97	5.14	Nov-01	1.87	Nov-05	3.88
Dec-85	7.10	Dec-89	7.63	Dec-93	3.06	Dec-97	5.16	Dec-01	1.69	Dec-05	3.89
Jan-86	7.07	Jan-90	7.64	Jan-94	2.98	Jan-98	5.04	Jan-02	1.65	Jan-06	4.24
Feb-86	7.06	Feb-90	7.74	Feb-94	3.25	Feb-98	5.09	Feb-02	1.73	Feb-06	4.43
Mar-86	6.56	Mar-90	7.90	Mar-94	3.50	Mar-98	5.03	Mar-02	1.79	Mar-06	4.51
Apr-86	6.06	Apr-90	7.77	Apr-94	3.68	Apr-98	4.95	Apr-02	1.72	Apr-06	4.60
May-86	6.15	May-90	7.74	May-94	4.14	May-98	5.00	May-02	1.73	May-06	4.72
Jun-86	6.21	Jun-90	7.73	Jun-94	4.14	Jun-98	4.98	Jun-02	1.70	Jun-06	4.79
Jul-86	5.83	Jul-90	7.62	Jul-94	4.33	Jul-98	4.96	Jul-02	1.68	Jul-06	4.95
Aug-86	5.53	Aug-90	7.45	Aug-94	4.48	Aug-98	4.90	Aug-02	1.62	Aug-06	4.96
Sep-86	5.21	Sep-90	7.36	Sep-94	4.62	Sep-98	4.61	Sep-02	1.63	Sep-06	4.81
Oct-86	5.18	Oct-90	7.17	Oct-94	4.95	Oct-98	3.96	Oct-02	1.58	Oct-06	4.92
Nov-86	5.35	Nov-90	7.06	Nov-94	5.29	Nov-98	4.41	Nov-02	1.23	Nov-06	4.94
Dec-86	5.53	Dec-90	6.74	Dec-94	5.60	Dec-98	4.39	Dec-02	1.19	Dec-06	4.85
Jan-87	5.43	Jan-91	6.22	Jan-95	5.71	Jan-99	4.34	Jan-03	1.17	Jan-07	4.98
Feb-87	5.59	Feb-91	5.94	Feb-95	5.77	Feb-99	4.44	Feb-03	1.17	Feb-07	5.03
Mar-87	5.59	Mar-91	5.91	Mar-95	5.73	Mar-99	4.44	Mar-03	1.13	Mar-07	4.94
Apr-87	5.64	Apr-91	5.65	Apr-95	5.65	Apr-99	4.29	Apr-03	1.13	Apr-07	4.87
May-87	5.66	May-91	5.46	May-95	5.67	May-99	4.50	May-03	1.07	May-07	4.73
Jun-87	5.67	Jun-91	5.57	Jun-95	5.47	Jun-99	4.57	Jun-03	0.92	Jun-07	4.61
Jul-87	5.69	Jul-91	5.58	Jul-95	5.42	Jul-99	4.55	Jul-03	0.90	Jul-07	4.82
Aug-87	6.04	Aug-91	5.33	Aug-95	5.40	Aug-99	4.72	Aug-03	0.95	Aug-07	4.20
Sep-87	6.40	Sep-91	5.22	Sep-95	5.28	Sep-99	4.68	Sep-03	0.94	Sep-07	3.89
Oct-87	6.13	Oct-91	4.99	Oct-95	5.28	Oct-99	4.86	Oct-03	0.92	Oct-07	3.90
Nov-87	5.69	Nov-91	4.56	Nov-95	5.36	Nov-99	5.07	Nov-03	0.93	Nov-07	3.27
Dec-87	5.77	Dec-91	4.07	Dec-95	5.14	Dec-99	5.20	Dec-03	0.90	Dec-07	3.00

Source: Board of Governors of the Federal Reserve System

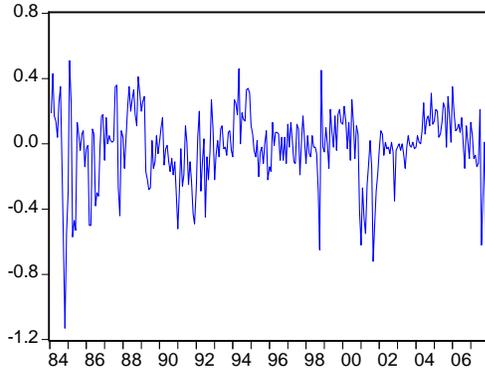


Figure 2. Differenced 3-month Treasury Bill Rate

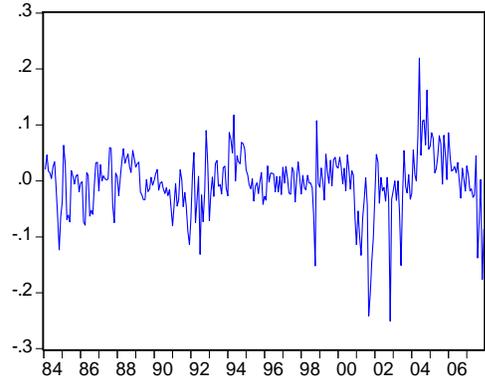


Figure 3. Differenced, in logs, 3-month Treasury Bill Rate

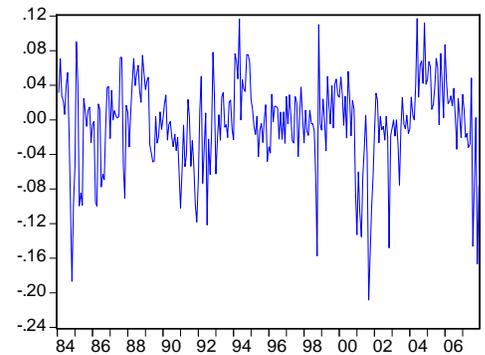


Figure 4. Differenced, in square roots, 3-month Treasury Bill Rates

Figures 3 and 4 contain the graphs of the series of differenced logs ( $\ln y_t - \ln y_{t-1} = (1-L)\ln y_t$ ) and differenced square roots ( $\sqrt{y_t} - \sqrt{y_{t-1}} = (1-L)\sqrt{y_t}$ ) respectively. The ACF and the PACF of the three transformed series are presented in Table 2. The simple plots as well as the ACF graphs suggest that the transformed series are all stationary.<sup>4</sup>

<sup>4</sup> Dickey-Fuller tests for a unit root, although not included in the paper, corroborate the findings of informal graphical tests.

Table 2. Autocorrelations and Partial Autocorrelations of Transformed Series

Lag	$(1-L)Tbills$		$(1-L)\ln(Tbills)$		$(1-L)\sqrt{Tbills}$	
	AC	PAC	AC	PAC	AC	PAC
1	0.509*	0.509*	0.502*	0.502*	0.501*	0.501*
2	0.200*	-0.079	0.285*	0.044	0.245*	-0.008
3	0.159*	0.122	0.314*	0.208*	0.245*	0.168*
4	0.129*	0.010	0.236*	-0.001	0.188*	0.004
5	0.164*	0.120	0.224*	0.100	0.199*	0.115
6	0.228*	0.121	0.298*	0.151	0.267*	0.139*
7	0.162*	-0.026	0.295*	0.086	0.224*	0.020
8	0.079	-0.015	0.197*	-0.039	0.160*	0.005
9	0.094	0.052	0.195*	0.043	0.170*	0.051
10	0.012	-0.111	0.104	-0.114	0.057	-0.126
11	0.025	0.054	0.092	0.030	0.056	0.040
12	0.019	-0.069	0.076	-0.071	0.051	-0.068
13	-0.108	-0.151	0.031	-0.052	-0.053	-0.124
14	-0.049	0.097	0.054	0.015	-0.010	0.048
15	0.095	0.112	0.025	-0.048	0.069	0.051
16	0.081	0.000	0.017	0.019	0.051	0.024
17	0.037	0.008	0.013	0.001	0.024	0.002
18	0.034	0.018	0.009	0.005	0.023	0.005
19	-0.010	0.006	-0.062	-0.070	-0.032	-0.017
20	0.027	0.049	0.014	0.108	0.021	0.078

Notes. AC and PAC denote autocorrelations and partial autocorrelations respectively and their significance is compared with  $\pm 2/\sqrt{T} = \pm 0.1231$ . \* denotes significance at the 5% level.

The general specification of an  $ARMA(p, q)$  model with mean zero, where  $p$  and  $q$  represent the order of the autoregressive and moving average parts of the series, is given by:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (1)$$

$$= (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) y_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t$$

where  $\forall i = 1, 2, \dots, p$  and  $\forall j = 1, 2, \dots, q$   
 $-1 < \varphi_i < 1$  and  $\varphi_1 + \varphi_2 + \dots + \varphi_p \neq 1$   
 $-1 < \theta_j < 1$  and  $\theta_1 + \theta_2 + \dots + \theta_p \neq 1$

The error term,  $\varepsilon_t$  is white noise with:

$$E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2, \text{Cov}(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s \quad (2)$$

The sample autocorrelations and partial autocorrelations are compared with their theoretical counterparts to identify the parameters  $p$  and  $q$ . In general, an autoregressive (moving average) model is inferred if the sample autocorrelations (partial autocorrelations) decay toward zero and the sample autocorrelations (partial autocorrelations) cut off. Let  $\rho_s$  represent the autocorrelation between  $y_t$  and its  $s$ -period lagged value  $y_{t-s}$  and let  $\phi_{ss}$  be the corresponding partial autocorrelation. In Table 3,

we present the criteria for choosing  $p$  and  $q$  based on the values of the sample autocorrelations and partial autocorrelations.

**Table 3.** Identification of ARMA Models

Model	ACF	PACF
White Noise	$\rho_s = 0 \quad \forall s \geq 1$	$\phi_{ss} = 0 \quad \forall s \geq 1$
MA(1)	$\rho_1 \neq 0, \rho_s = 0 \quad \forall s > 1$	Direct or oscillatory decay toward zero
MA(q)	$\rho_1 \neq 0, \dots, \rho_q \neq 0, \rho_s = 0 \quad \forall s > q$	Direct or oscillatory decay toward zero
AR(1)	Direct or oscillatory decay toward zero	$\phi_{11} \neq 0, \phi_{ss} = 0 \quad \forall s > 1$
AR(p)	Direct or oscillatory decay toward zero	$\phi_{11} \neq 0, \dots, \phi_{pp} \neq 0, \phi_{ss} = 0 \quad \forall s > p$
AR-MA(1,1)	Decay toward zero beginning from $\rho_1$	Decay toward zero beginning from $\phi_{11}$
AR-MA(p,q)	Decay toward zero beginning from $\rho_q$	Decay toward zero beginning from $\phi_{pp}$

In most applications, the identification may not lead to a single model and hence several models must be examined with and without pre-transformations. As mentioned earlier, since the underlying series is not stationary,  $ARMA(p, q)$  models are identified with  $y_t$  replaced with  $(1-L)y_t$ ,  $(1-L)\ln y_t$ , or  $(1-L)\sqrt{y_t}$ . We consider only three models for illustration. Model 1 is a constrained ARMA(6, 1) model, identified from the ACF and PACF graphs of a differenced series,  $(1-L)y_t$ . Models 2 and 3 are both constrained AR(6) models based on the ACF and PACF of differenced series with a log,  $(1-L)\ln y_t$ , and a square root,  $(1-L)\sqrt{y_t}$ , pre-transformation respectively. The models, and their corresponding estimation commands in Eviews, are given below.

Model1:

$$(1 - \phi_1 L - \phi_6 L^6)(1 - L)Tbill_t = (1 + \theta_1 L)\varepsilon_t \quad [Eviews : d(Tbill) MA(1) AR(6)] \quad (3)$$

Model2

$$(1 - \phi_1 L - \phi_3 L^3 - \phi_6 L^6)(1 - L)\ln(Tbill_t) = \varepsilon_t \quad [Eviews : d(\log(Tbill)) AR(1) AR(3) AR(6)] \quad (4)$$

Model3

$$(1 - \phi_1 L - \phi_6 L^6)(1 - L)\sqrt{Tbill_t} = \varepsilon_t \quad [Eviews : d(sqr(Tbill)) AR(1) AR(6)] \quad (5)$$

The above competing models are estimated with the data from 1984:01 - 2005:12. The last 24 observations from 2006 and 2007 are left out to analyze out of sample performance of these models, which we will discuss later. In Table IV, we present the estimation results of the three models. All parameters of the models are statistically significant at a 5 percent level.

**Table 4.** Parameter Estimates of Competing ARMA Models

Parameter Estimates	Model 1	Model 2	Model 3
$\phi_1$	0.3148* (0.1080)	0.4235* (0.0551)	0.4638* (0.0541)
$\phi_3$	NA	0.1469* (0.0567)	NA
$\phi_6$	0.1749* (0.0589)	0.1651* (0.0561)	0.1831* (0.0545)
$\theta_1$	0.2293* (0.1128)	NA	NA

Notes. Standard errors are in parentheses below the parameter estimates; \* denotes significance at the 5% level; Estimation period is 1984:01 - 2005:12. Model 1:  $(1 - \phi_1 L - \phi_6 L^6)(1 - L)Tbill_t = (1 + \theta_1 L)\varepsilon_t$ ; Model 2:

$$(1 - \phi_1 L - \phi_3 L^3 + \phi_6 L^6)(1 - L)\ln(Tbill_t) = \varepsilon_t; \text{ Model 3:}$$

$$(1 - \phi_1 L - \phi_6 L^6)(1 - L)\sqrt{Tbill_t} = \varepsilon_t.$$

Apart from the statistical significance of the parameters, an important test is that the residuals,  $e_t = y_t - \hat{y}_t$ , of the estimated model are white noise. In other words, for a model to be acceptable, no discernable information should be left back in the residuals. A convenient test for white noise is the Ljung-Box (1978) test, also called the Q test, which is implemented as:

$$Q = n(n+2) \sum_{i=1}^k \frac{1}{(n-i)} \hat{\rho}_i^2 \quad (6)$$

where  $\hat{\rho}_i$  is the estimated autocorrelation of the residual at lag  $i$  and  $n^3$  is the number of residuals used for the test.  $Q$  has an asymptotic chi-square distribution with degrees of freedom given by  $k-p-q$ . The choice of  $k$  is somewhat ad hoc, and it is common to perform this test for various reasonable values of  $k$ . A model is acceptable only when the null hypothesis that the residuals are white noise is not rejected. It is worth pointing out that the Q test is implemented on the residuals and therefore the degrees of freedom of the test must be adjusted for the number of estimated parameters,  $p + q$ . Eviews makes this adjustment in the  $p$ -values of the Q test if it is performed as a residual test within estimation.<sup>6</sup> It is common for students to ignore this important fact and mistakenly perform the Q-test on the residuals outside the model. In fact, Diebold makes this mistake and consequently miscalculates  $p$ -values throughout the book.<sup>7</sup>

<sup>5</sup> The number of residuals may be smaller than the original sample size due to differencing and the order of the autoregression.

<sup>6</sup> For the Q-test of the residuals in Eviews that is adjusted for the degrees of freedom, choose 'Views' after estimation and under 'Residuals Tests', compute 'Correlogram - Q statistics'.

<sup>7</sup> On page 203 of Chapter 10, "Putting it All Together", the reported  $p$ -values are spuriously large. As a consequence his selected model should have actually been discarded making the remaining discussion in this important chapter quite futile.

The top part of Table 5 presents the p-values of the Q test using  $k$  equal to 6, 12, and 18. Since the p-values are consistently above the chosen significance level of 5 percent, the residuals from Models 1, 2, and 3 are found to be white noise. Models that pass the white noise test of the residuals are then subjected to a variety of model selection criteria to select the most appropriate forecasting model.

**Table 5.** Model Selection Criteria Applied to Competing Models

Model Selection Criteria	Model 1	Model 2	Model 3
p-value for the Q test (k=6)	0.493*	0.763*	0.239*
p-value for the Q test (k=12)	0.515*	0.593*	0.339*
p-value for the Q test (k=18)	0.109*	0.815*	0.248*
AIC of the residuals	0.0374@	0.0381	0.0375
SIC of the residuals	0.0389	0.0398	0.0386@
RMSE of forecast Errors (Dynamic)	0.6820	0.6499	0.6203@
RMSE of forecast Errors (Static)	0.2300	0.2082@	0.2243

Notes. \* denotes that the residuals are white noise at the 5% level; @ denotes the best model selection measure. AIC and SIC are based on the residuals from 1984:08 - 2005:12; RMSE is based on the forecast errors from 2006:01 - 2007:12.

Model 1:  $(1 - \phi_1 L - \phi_6 L^6)(1 - L)Tbill_t = (1 + \theta_1 L)\varepsilon_t$ ;

Model 2  $(1 - \phi_1 L - \phi_3 L^3 + \phi_6 L^6)(1 - L)\ln(Tbill_t) = \varepsilon_t$ ;

Model 3:  $(1 - \phi_1 L - \phi_6 L^6)(1 - L)\sqrt{Tbill_t} = \varepsilon_t$ .

### Model Selection Criteria

Competing models are compared on the basis of in-sample predictability and out-of-sample forecasting ability. The comparison of in-sample predictability of the estimated model is based on the residuals. Here is where the confusion lies. The residuals reported by statistical packages are not comparable across models if they are based on different pre-transformations. The default residuals computed by statistical packages are  $(1 - L)(y_t - \hat{y}_t)$  for Model 1,  $(1 - L)(\ln y_t - \ln \hat{y}_t)$  for Model 2 and  $(1 - L)(\sqrt{y_t} - \sqrt{\hat{y}_t})$  for Model 3. In order to compute model selection measures, it is important to use consistently defined residuals,  $e_t = y_t - \hat{y}_t$ , where  $\hat{y}_t$  can easily be obtained for any model.<sup>8</sup> This issue is not raised in the above mentioned textbooks.

<sup>8</sup> In Eviews consistently defined residuals, Resid, are easily computed by choosing 'Forecast' from the menu and specifying the estimation range with 'Static forecast'.

Two commonly used model selection criteria, based on the residuals, are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are derived as:<sup>9</sup>

$$AIC = \exp\left(\frac{2k}{n}\right) \sum \frac{e_t^2}{n}; \quad SIC = n^{k/n} \sum \frac{e_t^2}{n} \quad (7)$$

The AIC and SIC are based on the residual sum of squares and a penalty for the number of parameters used in the estimated model. They both measure the information content in the residuals; the smaller the value, the better is the fit. Since model selection measures are used to compare models across transformations, one must first compute the residuals manually to derive AIC and SIC.

So far we have discussed model comparison on the basis of their in-sample predictability. The AIC and SIC essentially compare competing models on the basis of how well the estimated models fit the sample data. While these criteria impose a penalty for over-fitting, they are imperfect and can be misleading. Sometimes a well fitting model may not perform so well in making out-of-sample forecasts. Therefore, it is important to choose between competing models on the basis of both in-sample predictability as well as out-of sample forecasting ability.

In order to analyze out of sample performance of the models, we generally leave out the last few observations from estimation. If the sample consists of  $T$  observations, we estimate the model with  $T_1$  and leave the remaining  $T_2$  for analyzing out of sample performance, where  $T = T_1 + T_2$ .<sup>10</sup> Alternative models are compared on the basis of their in-sample predictability with a shortened span of data and also their forecasting ability of the hold-back period. Out-of-sample performance is compared on the basis of the forecast errors,  $f_{t^*} = y_{t^*} - \hat{y}_{t^*}$  for  $t^* = T_1 + 1, T_1 + 2, \dots, T$ . Dynamic as well as the static forecasts for  $T_2$  observations,  $\hat{y}_{t^*}$ , are based on the parameter estimates obtained from the reduced sample of  $T_1$  observations.<sup>11</sup> All dynamic forecasts are conditional on

<sup>9</sup> There are other monotonic transformations, including the ones reported by Eviews that can also be used for model comparison.

<sup>10</sup> The definition of the holdback period is somewhat ad hoc. Further, a large sample is needed to have reliable estimates in the in-sample period and the forecast analysis in the out-of-sample period.

<sup>11</sup> The ideal method, though time consuming, is to make forecasts on the basis of recursive estimation. In other words, we first estimate the model with  $T_1$  observations to make a one-period ahead

the information at  $T_1$ . The static method conditions on the actual value of the variable rather than its fitted value to forecast beyond one period. Therefore, while a first period forecast beyond the estimation sample is identical, subsequent forecasts can differ significantly between the dynamic and static methods. The static and dynamic forecasts are easily computed in Eviews by choosing 'Forecast' and specifying the forecast range with a static or a dynamic method. With both types of forecasts, the preferred model will have the smallest root mean square error (RMSE) of the forecast errors where

$$RMSE(f) = \sqrt{\frac{\sum f_{t^*}^2}{T_2}}. \quad (8)$$

Surprisingly these simple model selection criteria have not been implemented in standard textbooks.

The lower portion of Table V contains the AIC and SIC of the consistently defined residuals, and the RMSE of the forecast errors with static and dynamic methods for each of the estimated models, reported in Table IV. As is often the case, the four model selection criteria, namely AIC, SIC, RMSE (dynamic), and RMSE (static), do not lead to a clear winner. However, we choose Model 3 since it outperforms the other two models in terms of SIC of the residuals and RMSE (dynamic) of the forecast errors. Furthermore, Model 3 beats Model 1 and Model 2 individually in three out of four model selection criteria. Interestingly, models identified on the square root pre-transformation outperforms all other models. As a final step the chosen Model 3 is re-estimated with the entire data from 1984 – 2007 to make a forecast of 2.9188 percent for January, 2008.

It is worth noting that the above three models used in this paper are for illustration only. There may be other suitable models with transformations including detrending. Since economic theory provides no guidance on transformations or ARMA parameters, students and practitioners should be encouraged to identify all possible suitable models for comparison. While the future is uncertain, practitioners must use the available sample information to make the best possible forecast.

## Conclusion

Most academics and practitioners are in favor of imparting modern forecasting tools to business and economics students. Despite the obvious importance, many universities do not offer a specialized course in forecasting. Fur-

ther, while statistical packages such as Eviews have greatly simplified estimation and analysis of the ARMA models, popular undergraduate textbooks do not provide a comprehensive treatment of the diagnostic analysis of these models. These books lack the emphasis on consistently defined residuals to compare in-sample predictability of competing models with and without pre-transformations. Further, since forecasting is essentially an out-of-sample problem, the important analysis of the out of sample forecast errors based on static and dynamic methods has also been left out. This paper succinctly outlines this important diagnostic step of ARMA modeling, using monthly data on 3-month Treasury bills rate for illustration. The easy to follow steps of ARMA modeling with Eviews included in the paper can be invaluable to a business student or a practitioner.

Correspondence: sjaggia@calpoly.edu

## REFERENCES

- Diebold, F. X. 2007. *Elements of Forecasting*. Thomson and South-Western. Fourth Edition.
- Enders, W. 2005. *Applied Econometric Time Series*. John Wiley & Sons Inc. Second Edition.
- Hanke, J.E. and Wichern, D.W.. 2009. *Business Forecasting*. Pearson Prentice Hall. Ninth Edition.
- Studenmund, A. H. 2011. *Using Econometrics: A Practical Guide*. Pearson Addison Wesley. Sixth Edition.
- Wooldridge, J.M. 2009. *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning. Fourth Edition.

---

forecast  $\hat{y}_{T_1+1}$ . The model is then re-estimated with  $T_1 + 1$  observations to forecast  $\hat{y}_{T_1+2}$  etc.