Midterm Exam Answer Key
Exam Given March 4, 2010

<u>Section I.</u> *(30 points) True/False. Please indicate whether each of the following statements is true (T) or false (F) by circling the appropriate letter.*

T F 1. The q-statistic tests for higher-order serial correlation.

T F 2. An MA(q) series is invertible if it can be written only as a function of lagged error terms.

T F 3. To correctly identify cyclical properties it is important to model trend and seasonal patterns.

T F 4. Smaller models are typically better for forecasting than larger models even though they may explain less of the historical data movements.

T F 5. Ordinary least squares finds the model parameters that minimize the mean absolute error.

T F 6. We often omit from our estimation period data at the end of the sample, because that data is unreliable.

T F 7. Both the unconditional and conditional mean of an MA series are zero.

T F 8. The Akaike criterion tends to pick larger models than the Schwarz criterion

T F 9. A white noise process has a zero mean, constant variance, and autocovariances that are not a function of calendar time.

T F 10. A good forecasting model should explain at least 95% of the variation in a series.

T F 11. A seasonally adjusted series includes cyclical movements.

T F 12. It is possible to compare SIC criteria for log-linear and linear models.

T F 13. The partial autocorrelation function gives the correlation at each displacement taking as given the correlations at other displacements.

T F 14. The usual approach to modeling seasonality is to include a constant and a seasonal dummy for each period of the year.

T F 15. Even a good forecasting model will fail to predict the effects of rare events.

Section II Analytical. Show all your work on this page.
Consider the following time series process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta\varepsilon_{t-1}$$
$$where\ \varepsilon_t = WN(0,\sigma^2)$$

1. (10 points) *Express the time series in lag operator form.*

$$y_t = \phi_1 L y_t + \phi_2 L^2 y_t + \varepsilon_t + \theta L \varepsilon_t$$
$$(1-\phi_1 L - \phi_2 L^2) y_t = (1+\theta L)\varepsilon_t$$

2. (10 points) *Rewrite the equation in moving average form. What condition must hold in order for you to do this?*

We want y_t as a function of lagged ε_t. The easiest way to do this is to divide the above expression in lag operator form through by the LHS lag polynomial, giving:

$$y_t = \frac{(1+\theta L)}{(1-\phi_1 L - \phi_2 L^2)}\varepsilon_t$$

This expression is defined as long as the condition for stationarity is met, i.e. the two roots of the denominator polynomial, $\Phi(L)$ are within the unit circle. (Note: only in the AR(1) case does this simplify to $|\phi| < 1$, although sum(ϕ_i) < 1 is a necessary condition.)

You can also get this by substituting backward for lagged y's in the original expression:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta\varepsilon_{t-1}$$

This means that $y_{t-1} = \phi_1 y_{t-2} + \phi_2 y_{t-3} + \varepsilon_{t-1} + \theta\varepsilon_{t-2}$, *and*

$y_{t-2} = \phi_1 y_{t-3} + \phi_2 y_{t-4} + \varepsilon_{t-2} + \theta\varepsilon_{t-3}$. Substituting back into original y_t exppression gives

$$y_t = \phi_1[\phi_1(\phi_1 y_{t-3} + \phi_2 y_{t-4} + \varepsilon_{t-2} + \theta\varepsilon_{t-3}) + \phi_2 y_{t-3} + \varepsilon_{t-1} + \theta\varepsilon_{t-2}] + \phi_2(\phi_1 y_{t-3} + \phi_2 y_{t-4} + \varepsilon_{t-2} + \theta\varepsilon_{t-3}) + \varepsilon_t + \theta\varepsilon_{t-1}$$

$= \varepsilon_t + (\phi_1 + \theta)\varepsilon_{t-1} + (\phi_1\theta + \phi_1^2 + \phi_2)\varepsilon_{t-2} + (\phi_1^2\theta + \phi_2\theta)\varepsilon_{t-3}$ *plus* y_{t-3} and y_{t-4} terms.

(If I did that right). And so on.

3. (10 points) *What properties would you expect the autocorrelation and partial autocorrelation functions for this time series to have if*
   a. *ϕ_1 and ϕ_2 are zero? Explain.*
      In this case, we have an MA(1) process, for which the autocorrelation function (correlogram) will have a single spike of size θ at the lag 1 displacement, and the partial autocorrelation function will display damped oscillation if θ>0 and one-sided damping if θ<0. (See pp. 139-43 for the math.)
   b. *θ is zero? Explain.*
      In this case we have an AR(2) process, for which the correlogram will show gradual one-sided damping, or damped oscillation if the process has complex roots. The partial autocorrelation function will have spikes at the 1st and 2nd lag displacements only. (See pp. 146-151.)

You can confirm these statements by simulating the processes using eviews series commands.

Section III. (40 points) Application. Three candidate models of oil prices are reported in the three pages that follow. Review these model results and answer the following questions. TIME is a time trend and DQ1 to DQ3 are quarterly seasonal dummies.

1. *(10 points) Which model would you choose based on in-sample estimation results? Why? Are there any economic reasons to prefer one specification over the others?*

	Adj. R-squared	AIC	SIC	Hannan-Quinn
	(big is better)	(small is better)	(small is better)	(small is better)
Model 1	0.51	7.72	7.82	7.76
Model 2	0.52	7.71	7.82	7.75
Model 3	0.55	7.64	7.73	7.67

Any of our three standard selection criteria, minimizing s^2 (mean squared error adjusted for degrees of freedom; equivalent to maximizing adjusted R-squared), AIC and SIC favor the same model, model 3, the exponential model. (The Hannan-Quinn criteria also picks this model.)

There is no strong economic reason to favor a particular model. There seems to be some nonlinearity, which is best captured by the exponential model. Note that this is also the model that has the smallest residuals for the most recent period—although they are still very large! The exponential functional form fits series that have a constant average growth rate over time. Whether that makes sense for oil prices is an open question, since they seem to be characterized by long periods of stability or decline punctuated by short periods of rapid appreciation.

2. *(10 points) Would you make any further simplifications to your selected model? Explain.*

The seasonal dummy variables are not significantly different from zero, indicating no significant difference between these three quarters and the omitted fourth quarter. (This could be tested formally using a Wald test.) So I would drop them and re-estimate the three candidate models and compare the results again.

3. *(10 points) Is there evidence that your selected model could be enhanced to improve its forecasting performance? What sort of dynamic model appears most likely to be useful? Explain.*

Yes, there is clearly a lot of unmodeled cyclical behavior, evident in the strong persistence of the residuals in the actual/forecast/residuals graphs. This looks like it may be AR(1) behavior (based on gradually damping autocorrelation and on the large partial autocorrelation at lag 1. But we would need to explore alternative ARMA models to try to fit the cyclical patterns in the data.

4. *(10 points) Is there evidence of any problems with your selected model that could affect the reliability of the parameter estimates? Explain.*

The residuals are serially correlated, evident in their strong persistence, in the very small Durbin-Watson statistic, and most formally in the clear rejection of white noise errors by the Q-Stats for each model. Also, the errors are not normally distributed, indicated by the low P-value for the Jarque-Bera statistic (they are kurtotic—more peaked—compared to the normal distribution). These departures from white noise errors imply that the OLS estimates may not be reliable. And modeling the serial correlation will also improve the forecast performance.