A tutorial note on the properties of ARIMA optimal forecasts

Galimberti, Jaqueson K.

The University of Manchester

10 January 2012

A tutorial note on the properties of ARIMA optimal forecasts*

Jaqueson K. Galimberti†

First draft: January 10, 2012.

This version: August 20, 2012.

Abstract

Assuming an $ARIMA(p,I,q)$ model represents the data, I show how optimal forecasts can be computed and derive general expressions for its main properties of interest. Namely, I present stepwise derivations of expressions for the variances of forecast errors, and the covariances between them at arbitrary forecasting horizons. Matricial forms for these expressions are also presented to facilitate computational implementation.

1 Preliminaries

Lets say that a time series of observations y_t was found (or estimated) to be represented by the $ARIMA(p,I,q)$ model given by

$$\phi(L)(1-L)^I y_t = \theta(L) v_t, \tag{1}$$

where v_t is a white noise sequence with given (or estimated) variance σ_v^2, $\phi(L)$ and $\theta(L)$ are lag polynomials given by

$$\phi(L) = 1 - \sum_{j=1}^{p} a_j L^j,\ \theta(L) = 1 - \sum_{j=1}^{q} b_j L^j, \tag{2}$$

which are taken as the p-order stationary autoregressive operator and the q-order invertible moving average operator, respectivelly. As usual, to ensure stationarity and invertibility the roots of

$$1 - \sum_{j=1}^{p} a_j x^j = 0 \; and \; 1 - \sum_{j=1}^{q} b_j x^j = 0 \tag{3}$$

must lie outside the (complex) unit circle, respectivelly.

Although our main purpose here is to analyze the properties of the forecast errors obtained from the $ARIMA(p,I,q)$ model, to facilitate our derivations we neglect the difference operator at some moments by using a simple change of variable given by

$$x_t = (1-L)^I y_t, \tag{4}$$

*This is written for didactic purposes, supposedly a handy reference. Famous textbook references on the topic are found in Hamilton (1994); Box et al. (2008). Although not aimed for publication, citations, comments, and suggestions are welcome.

†Email: jakaga2002@yahoo.com.br. Website: https://sites.google.com/site/jkgeconoeng/.

such that (1) reduces to an $ARMA\,(p,q)$ form given by

$$\phi\,(L)\,x_t = \theta\,(L)\,v_t. \tag{5}$$

In what follows we first show how optimal forecasts for y_{t+h}, as given by the conditional expectation operator $E_{t-1}\,[y_{t+h}]$, can be computed from the model specification in (4)-(5). Next, in section §3 we obtain expressions for the forecast errors and use these to compute conditional variances and covariances of these errors for varying forecasting horizons h.

2 Optimal Forecasts

2.1 For the differenced series

Focusing on the $ARMA\,(p,q)$ form of (5), notice that from (2) we can represent x_t as determined by its own lagged values and the sequence of innovations,

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} - \sum_{j=1}^{q} b_j v_{t-j} + v_t, \tag{6}$$

which in turn can be easily extended for any desired horizon h,

$$x_{t+h} = \sum_{j=1}^{p} a_j x_{t+h-j} - \sum_{j=1}^{q} b_j v_{t+h-j} + v_{t+h}. \tag{7}$$

For the computation of forecasts and its errors it is helpful to decompose the summations in (7) into the pieces of information that are known at the period at which the forecast is being made, here $t-1$, and those not yet known, which leads to

$$\begin{aligned} x_{t+h} \quad = \quad & \sum_{j=1}^{h} a_j x_{t+h-j} - \sum_{j=1}^{h} b_j v_{t+h-j} + v_{t+h} + \\ & + \sum_{j=h+1}^{p} a_j x_{t+h-j} - \sum_{j=h+1}^{q} b_j v_{t+h-j}, \end{aligned} \tag{8}$$

where the first line contains the information not yet known to the forecaster, and the second line contains only information already observed. Optimal forecasts can then be obtained using the conditional expectation operator, $E_{t-1}\,[\bullet]$, which applied to (8) leads to

$$E_{t-1}\,[x_{t+h}] = \sum_{j=1}^{h} a_j E_{t-1}\left[x_{t+h-j}\right] + \sum_{j=h+1}^{p} a_j x_{t+h-j} - \sum_{j=h+1}^{q} b_j v_{t+h-j}, \tag{9}$$

from which forecasts for any horizon h can be computed recursively based solely on information available on period $t-1$.

2.2 For the integrated series

Once the forecasts for the differenced series x_t were computed, to obtain the forecasts for the integrated series y_t first notice that substituting (4) back into $E_{t-1}[x_{t+h}]$ we obtain

$$E_{t-1}[x_{t+h}] = E_{t-1}\left[(1-L)^I y_{t+h}\right]. \tag{10}$$

The issue now is to find a way to disentangle $E_{t-1}[y_{t+h}]$ from the RHS of (10). To do this notice that from the binomial theorem of elementary algebra we have that the difference operator can be expressed as an Ith order polynomial in the lag operator as given by

$$(1-L)^I = \sum_{j=0}^{I} \binom{I}{j} (-1)^j L^j, \tag{11}$$

where $\binom{I}{j}$ denotes a binomial coefficient, which in factorial form can be computed as $\frac{I!}{j!(I-j)!}$. Substituting (11) into (10) we find that

$$\begin{aligned} E_{t-1}[x_{t+h}] &= E_{t-1}\left[\sum_{j=0}^{I} \binom{I}{j} (-1)^j y_{t+h-j}\right], & (12)\\ &= E_{t-1}[y_{t+h}] + E_{t-1}\left[\sum_{j=1}^{I} \binom{I}{j} (-1)^j y_{t+h-j}\right], & (13) \end{aligned}$$

where the last equation is found by separating the first element of the summation and solving for the coefficient given $j = 0$. We can decompose the remaining summation even further by noticing that $E_{t-1}\left[y_{t+h-j}\right] = y_{t+h-j}$ for any $j > h$, given that the value inside of the expectation brackets was already observed for these cases. Thus, decomposing the summation in (13) and rearranging terms we arrive at

$$\begin{aligned} E_{t-1}[y_{t+h}] &= E_{t-1}[x_{t+h}] - \sum_{j=1}^{\min\{h,I\}} \binom{I}{j} (-1)^j E_{t-1}\left[y_{t+h-j}\right] + \\ &\quad - \sum_{j=h+1}^{I} \binom{I}{j} (-1)^j y_{t+h-j}, & (14) \end{aligned}$$

which provides a recursive formula for computation of forecasts for the integrated series at any forecasting horizon using only the observations of the integrated series known at period $t-1$ and the forecasts previously computed for the differenced series, as given by (9).

3 Forecast Errors and Its Properties

3.1 For the differenced series

Letting $e(x_{t+h})$ denote the error of the forecast made using information from period $t-1$ for the value of the differenced series at period $t+h$, we have that

$$\begin{aligned} e(x_{t+h}) &= x_{t+h} - E_{t-1}[x_{t+h}], \quad (15) \\ &= \sum_{j=1}^{h} \left\{ a_j \left(x_{t+h-j} - E_{t-1}\left[x_{t+h-j}\right] \right) - b_j v_{t+h-j} \right\} + v_{t+h}, \quad (16) \end{aligned}$$

where (16) is obtained by simple substitution of (8) and (9) into the definition of the forecast error in (15). But notice that, from (15), $e\left(x_{t+h-j}\right) = x_{t+h-j} - E_{t-1}\left[x_{t+h-j}\right]$ for any $j = 1, ..., h$. Substituting this into (16) we arrive at a recursive formula for the forecast error,

$$e(x_{t+h}) = \sum_{j=1}^{h} \left\{ a_j e\left(x_{t+h-j}\right) - b_j v_{t+h-j} \right\} + v_{t+h}. \quad (17)$$

It is important to remember that though the forecasts errors for any horizon h can be computed recursively from (17), these measures, obviously, cannot be computed numerically before the actual realizations of the forecasted series have been observed. Our interest here is of course on the symbolic expressions for these forecast errors so that we can compute conditional variances and covariances for them. Before going through that, notice that there is an even simpler recursive form for the forecast errors given by

$$e(x_{t+h}) = \sum_{j=0}^{h} \delta_j v_{t+h-j}, \quad (18)$$

with

$$\delta_0 = 1 \textit{ and } \delta_j = \sum_{k=1}^{j} a_k \delta_{j-k} - b_j. \quad (19)$$

Now, letting $\sigma^2_{e(x_{t+h})}$ denote the conditional variance of the forecast error $e(x_{t+h})$ we have that from its definition

$$\begin{aligned} \sigma^2_{e(x_{t+h})} &= E_{t-1}\left[\left(e(x_{t+h}) - E_{t-1}\left[e(x_{t+h})\right] \right)^2 \right], \quad (20) \\ &= E_{t-1}\left[e(x_{t+h})^2 \right], \quad (21) \\ &= E_{t-1}\left[\left(\sum_{j=0}^{h} \delta_j v_{t+h-j} \right)^2 \right], \quad (22) \end{aligned}$$

where the simplification in (21) is obtained by direct application of the expectation operator to (15) (or 18), and (22) substitutes (18) into (21). Further, from the assumption that $v_t \sim IID\left(0, \sigma_v^2\right)$ we have that the expectation of the cross-products between the terms of the summation in (22) are null, i.e., $E_{t-1}[v_t v_s] = 0 \forall t \neq s$, while the expectation of the squared terms are the variance of v_t, i.e., $E_{t-1}\left[v_t^2\right]=\sigma_v^2$. Thus the conditional variance of the forecast errors for any horizon h are given by

$$\sigma^2_{e(x_{t+h})}=\sigma_v^2\sum_{j=0}^{h}\delta_j^2. \tag{23}$$

Conditional covariances of the forecast errors can be computed in a similar way. Letting $\sigma_{e(x_{t+h},x_{t+h-l})}$ denote the conditional covariance between the forecast errors $e\left(x_{t+h}\right)$ and $e\left(x_{t+h-l}\right)$, for any $l=1,...,h$, we have from definition that

$$\begin{aligned}\sigma_{e(x_{t+h},x_{t+h-l})} &= E_{t-1}\left[e\left(x_{t+h}\right)e\left(x_{t+h-l}\right)\right], && (24)\\ &= E_{t-1}\left[\left(\sum_{j=0}^{h}\delta_j v_{t+h-j}\right)\left(\sum_{j=0}^{h-l}\delta_j v_{t+h-l-j}\right)\right]. && (25)\end{aligned}$$

Now, it is evident from (25) that the product of the two summations lead to a total of $(h+1)(h-l+1)$ terms. However, only the $(h-l+1)$ terms of the second summation will lead to squared terms of v, such that following the same reasoning as for the computation of the variance, from the assumption that v_t is *IID* we have that (25) reduces to

$$\sigma_{e(x_{t+h},x_{t+h-l})}=\sigma_v^2\sum_{j=0}^{h-l}\delta_j\delta_{l+j}. \tag{26}$$

Notice that for $l>h$ the conditional covariance is null given that $e\left(x_{t+h-l}\right)=0$ for this case.

Finally, notice that (23) results as a special case of (26) when $l=0$. It may be useful for computation to consider a matricial form for these expressions, resulting in a matrix of variances/covariances of the forecast errors. Letting the column vector with $h+1$ forecast errors on the differenced series be denoted by $\mathbf{e}\left(\mathbf{x_h}\right)'=\begin{bmatrix} e\left(x_t\right) & e\left(x_{t+1}\right) & ... & e\left(x_{t+h}\right)\end{bmatrix}$, the forecast errors (co-)variance matrix is defined by $\Sigma_{\mathbf{x}}=E_{t-1}\left[\mathbf{e}\left(\mathbf{x_h}\right)\mathbf{e}\left(\mathbf{x_h}\right)'\right]$ and from (26) is given by

$$\Sigma_{\mathbf{x}}=\Delta_\delta\Delta_\delta'\sigma_v^2, \tag{27}$$

where Δ_δ is a lower triangular matrix formed with the δ_j coefficients from (19). Specifically,

$$\Delta_\delta=\begin{bmatrix}\delta_0 & 0 & \cdots & \cdots & 0\\ \delta_1 & \delta_0 & 0 & \cdots & 0\\ \delta_2 & \delta_1 & \delta_0 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \delta_h & \delta_{h-1} & \delta_{h-2} & \cdots & \delta_0\end{bmatrix}. \tag{28}$$

3.2 For the integrated series

To obtain similar measures for the integrated series we are going to first establish the relationship between the forecast error for the differenced series and that for the integrated series, so that we can then make use of the expressions we have already obtained in the previous subsection. To start notice that from (4) and (11) we have that

$$x_{t+h} = \sum_{j=0}^{I} \binom{I}{j} (-1)^j y_{t+h-j}, \tag{29}$$

which substituted into the definition of the differenced series forecast error, (15), results in

$$\begin{aligned} e(x_{t+h}) &= \sum_{j=0}^{I} \binom{I}{j} (-1)^j y_{t+h-j} - \sum_{j=h+1}^{I} \binom{I}{j} (-1)^j y_{t+h-j} + \\ &\quad - \sum_{j=0}^{h} \binom{I}{j} (-1)^j E_{t-1}\left[y_{t+h-j}\right], \qquad (30) \\ &= \sum_{j=0}^{h} \binom{I}{j} (-1)^j \left(y_{t+h-j} - E_{t-1}\left[y_{t+h-j}\right]\right), \qquad (31) \end{aligned}$$

where in (30) we are already incorporating a decomposition of the expected value of the summation in (29) into the portion of terms already known at the time of the forecast and those not yet known, just as we did in (14), and in (31) we are just simplifying the summations in (30) by removing the terms that cancel each other in the first two summations and then aggregating the remainder with the last summation of expected values.

Now, if we apply the definition of the forecast error function to the integrated series forecasts we have that

$$e(y_{t+h}) = y_{t+h} - E_{t-1}[y_{t+h}], \tag{32}$$

which for any $j = 0, \ldots, h$ can be equivalently put as

$$e\left(y_{t+h-j}\right) = y_{t+h-j} - E_{t-1}\left[y_{t+h-j}\right]. \tag{33}$$

Substituting (33) into (31) we arrive at a recursive formula for the integrated series forecast error, i.e.,

$$\begin{aligned} e(x_{t+h}) &= \sum_{j=0}^{h} \binom{I}{j} (-1)^j e\left(y_{t+h-j}\right), \qquad (34) \\ &= e(y_{t+h}) + \sum_{j=1}^{h} \binom{I}{j} (-1)^j e\left(y_{t+h-j}\right), \qquad (35) \\ e(y_{t+h}) &= e(x_{t+h}) - \sum_{j=1}^{h} \binom{I}{j} (-1)^j e\left(y_{t+h-j}\right), \qquad (36) \end{aligned}$$

where (35) is obtained by releasing the first term of the summation in (34), and the recursive formula in (36) is obtained by rearranging the terms in (35). Although (36) already provides a way to obtain the forecast errors for the integrated series from those calculated for the differenced series in (18), we can go further and simplify it to be determined solely from the forecast errors for the differenced series. Repeating the recursive substitutions in (36) a few times, we find that this formula can be generalized into

$$e\left(y_{t+h}\right)=\sum_{j=0}^{h}\beta_j e\left(x_{t+h-j}\right), \tag{37}$$

with

$$\beta_j=\frac{(I+j-1)!}{j!\,(I-1)!}, \tag{38}$$

which therefore establish a direct relationship between the forecast errors for the differenced series and the forecast errors for the integrated series, where the (non-negative integer) order of integration is arbitrary.

To obtain an expression for the conditional variance of forecast errors for the integrated series we can use (37)-(38) with the definition of this measure, as in (21),

$$\sigma^2_{e(y_{t+h})} = E_{t-1}\left[e\left(y_{t+h}\right)^2\right], \tag{39}$$

$$= E_{t-1}\left[\left(\sum_{j=0}^{h}\beta_j e\left(x_{t+h-j}\right)\right)^2\right], \tag{40}$$

where the square of the summation can be expressed as a squared multinomial

$$\sigma^2_{e(y_{t+h})} = \sum_{j=0}^{h}\beta_j^2 E_{t-1}\left[e\left(x_{t+h-j}\right)^2\right]+ +2\sum_{j=0}^{h}\sum_{l=j+1}^{h}\beta_j\beta_l E_{t-1}\left[e\left(x_{t+h-j}\right)e\left(x_{t+h-l}\right)\right], \tag{41}$$

with the expectation operator already solved for the parameters. But notice that the two expectations left to be solved in (41) correspond to the definitions of the conditional variance and covariance, given by (21) and (24) respectivelly, though we need to adjust the period of the forecast. Doing that we find that

$$\sigma^2_{e\left(x_{t+h-j}\right)} = \sigma_v^2\sum_{i=0}^{h-j}\delta_i^2, \tag{42}$$

$$\sigma_{e\left(x_{t+h-j},x_{t+h-l}\right)} = \sigma_v^2\sum_{i=0}^{h-l}\delta_{h-j-i}\delta_{h-l-i}. \tag{43}$$

Substituting (42) and (43) into (41) we arrive at a formula for the conditional variance of the forecast error for the integrated series

$$\sigma^2_{e(y_{t+h})}=\sum_{j=0}^{h}\beta_j^2\sigma^2_{e\left(x_{t+h-j}\right)}+2\sum_{j=0}^{h-1}\sum_{l=j+1}^{h}\beta_j\beta_l\sigma_{e\left(x_{t+h-j},x_{t+h-l}\right)}, \tag{44}$$

which can be indirectly computed from the given (or estimated) variance σ_v^2, and the autoregressive and moving average parameters of the $ARIMA\left(p,I,q\right)$ model in (1)-(2), using (42), (43), and the definitions of the δ and β parameters given in (19) and (38), respectivelly.

We can follow a similar approach to obtain an expression for the conditional covariance between forecast errors on the integrated series. From the definition of the conditional covariance, (24), we have that

$$\begin{aligned}\sigma_{e(y_{t+h},y_{t+h-l})} &= E_{t-1}\left[e\left(y_{t+h}\right)e\left(y_{t+h-l}\right)\right], &(45)\\ &= E_{t-1}\left[\left(\sum_{i=0}^{h}\beta_i e\left(x_{t+h-i}\right)\right)\left(\sum_{i=0}^{h-l}\beta_i e\left(x_{t+h-l-i}\right)\right)\right], &(46)\end{aligned}$$

where in (46) we have substituted (37) adjusting the indexes. The product of the two summations obviously will have $(h+1)(h-l+1)$ terms, which can be decomposed into $(h-l+1)$ squared terms and $h(h-l+1)$ cross-product terms. Such a decomposition is given by

$$\begin{aligned}\sigma_{e(y_{t+h},y_{t+h-l})} &= \sum_{i=0}^{h-l}\beta_i\beta_{l+i}E_{t-1}\left[e\left(x_{t+h-l-i}\right)^2\right]+\\ &+ \sum_{i=0}^{h-l}\sum_{k=0}^{l+i-1}\beta_i\beta_k E_{t-1}\left[e\left(x_{t+h-l-i}\right)e\left(x_{t+h-k}\right)\right]+\\ &+ \sum_{i=0}^{h-l}\sum_{k=l+i+1}^{h}\beta_i\beta_k E_{t-1}\left[e\left(x_{t+h-l-i}\right)e\left(x_{t+h-k}\right)\right]. &(47)\end{aligned}$$

Again, we can identify the expressions under expectations with the definitions of the conditional variance and covariance for the forecast error on the differenced variable. Namely, we have that

$$\begin{aligned}\sigma^2_{e(x_{t+h-l-i})} &= \sigma_v^2\sum_{m=0}^{h-l-i}\delta_m^2, &(48)\\ \sigma_{e(x_{t+h-l-i},x_{t+h-k})} &= \sigma_v^2\sum_{m=0}^{\min\{h-l-i,h-k\}}\delta_{h-l-i-m}\delta_{h-k-m}, &(49)\end{aligned}$$

where the minimization determining the last value for m into the summation in (49) will always pick the first option $(h-l-i)$ under the conditions implied by the first double summation in (47), and the second option $(h-k)$ under the second double summation in (47). Substituting (48) and (49) into (47) we obtain our final expression for the conditional covariance of the forecast errors on the integrated variable, i.e.,

$$\begin{aligned}\sigma_{e(y_{t+h},y_{t+h-l})} &= \sum_{i=0}^{h-l}\beta_i\beta_{l+i}\sigma^2_{e(x_{t+h-l-i})}+\\ &+ \sum_{i=0}^{h-l}\sum_{k=0}^{l+i-1}\beta_i\beta_k\sigma_{e(x_{t+h-l-i},x_{t+h-k})}+\\ &+ \sum_{i=0}^{h-l}\sum_{k=l+i+1}^{h}\beta_i\beta_k\sigma_{e(x_{t+h-l-i},x_{t+h-k})}. &(50)\end{aligned}$$

As with the differenced series, notice that (44) results as a special case of (50) when $l=0$. Again, we can turn these expressions into an unique matricial form to obtain the matrix of variances/covariances of forecast errors on the integrated series. Following previous notation, this matrix is defined as $\Sigma_{\mathbf{y}} = E_{t-1}\left[\mathbf{e}\left(\mathbf{y_h}\right)\mathbf{e}\left(\mathbf{y_h}\right)'\right]$ and from

(50) can be computed by

$$\Sigma_{\mathbf{y}} = \Delta_\beta \Sigma_{\mathbf{x}} \Delta'_\beta, \tag{51}$$

where Δ_β is a lower triangular matrix constructed with the β_j coefficients from (38) in the same way as (28), i.e.,

$$\Delta_\beta = \begin{bmatrix} \beta_0 & 0 & \cdots & \cdots & 0 \\ \beta_1 & \beta_0 & 0 & \cdots & 0 \\ \beta_2 & \beta_1 & \beta_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_h & \beta_{h-1} & \beta_{h-2} & \cdots & \beta_0 \end{bmatrix}. \tag{52}$$

Acknowledgments

I thank to the assistance provided by Joyce S. de Oliveira in the verification of some of the tough expressions above. Any remaining errors are of my entire responsibility, but let me know if you find any, please.

References

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2008). *Time Series Analysis: Forecasting and Control* (4th ed.). Wiley.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.