Predictability in Predictive Systems

Donald Robertson Stephen Wright

September 2009

1. Introduction

We establish upper and lower bounds on the predictability of a time series generated by a set of p $AR(1)$ state variables so that the reduced form is an $ARMA(p,p)$.

2. The Predictive System

Consider the following system of equations

$$\begin{aligned} y_t &= \beta_1 x_{1t-1} + \beta_2 x_{2t-1} + \ldots + \beta_p x_{pt-1} + u_t \\ x_{1t} &= \lambda_1 x_{1t-1} + v_{1t} \\ x_{2t} &= \lambda_2 x_{2t-1} + v_{2t} \\ &\vdots \\ x_{pt} &= \lambda_p x_{pt-1} + v_{pt} \end{aligned}$$

or concisely

$$\begin{aligned} y_t &= \beta' x_{t-1} + u_t \\ x_t &= \Lambda x_{t-1} + v_t \end{aligned}$$

where $x_t = (x_{1t}, \ldots x_{pt})'$ is a $p \times 1$ vector of predictor variables, $\beta = (\beta_1, \ldots \beta_p)'$ a $p \times 1$ vector of coefficients, $\Lambda = diag\{\lambda_1, \ldots, \lambda_p\}$ a $p \times p$ diagonal matrix and we assume $|\lambda_i| < 1$ for all i, $v_t = (v_{1t}, \ldots v_{pt})'$ and $(u_t, v_t')'$ is a $(p+1) \times 1$ vector of white noise errors with covariance matrix Ω where

$$\Omega = \begin{pmatrix} \sigma_u^2 & \sigma_{u1}^2 & & \sigma_{up}^2 \\ \sigma_{u1}^2 & \sigma_{11}^2 & \cdots & \sigma_{1p}^2 \\ & \vdots & \ddots & \\ \sigma_{up}^2 & \sigma_{1p}^2 & & \sigma_{pp}^2 \end{pmatrix}$$

Denote the R^2 in this predictive regression by $R_x^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$.

The reduced form for y_t is an $ARMA(p,p)$

$$\begin{aligned} y_t &= \beta'(I - \Lambda L)^{-1} v_{t-1} + u_t \\ &= (\beta_1, \ldots \beta_p) \begin{pmatrix} 1-\lambda_1 L & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & & \\ 0 & & & 1-\lambda_p L \end{pmatrix}^{-1} \begin{pmatrix} v_{1t-1} \\ v_{2t-1} \\ \\ v_{pt-1} \end{pmatrix} + u_t \end{aligned}$$

where L is the backward shift operator.

This gives

$$\prod_{i=1}^{p}(1-\lambda_i L)y_t = \sum_{j=1}^{p}\beta_j\left(\frac{\prod_{i=1}^{p}(1-\lambda_i L)}{(1-\lambda_j L)}\right)v_{jt-1} + \prod_{i=1}^{p}(1-\lambda_i L)u_t$$

which can usefully be written in an invertible $ARMA$ form as

$$y_t = \frac{\prod_{i=1}^{p}(1-\theta_i L)}{\prod_{i=1}^{p}(1-\lambda_i L)}\varepsilon_t$$

for some white noise process ε_t where the MA coefficients are derived by matching the autocorrelation structure on the right hand side and we choose these coefficients to satisfy $|\theta_i| < 1$ for all i. Note that the AR coefficients match the diagonal elements of Λ. We call this the fundamental $ARMA$ representation. Denote the R^2 that one would obtain in this $ARMA$ by $R_\varepsilon^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}$. In principle the set $\{\lambda_i, \theta_i\}$ can be derived from observations on y_t alone.

An alternative non-invertible $ARMA$ representations replaces each θ_i by its inverse.

$$y_t = \frac{\prod_{i=1}^{p}(1-\theta_i^{-1} L)}{\prod_{i=1}^{p}(1-\lambda_i L)}\eta_t$$

We call this the major-non-fundamental $ARMA$ representation (note that there are a number of other non-fundamental representations). Note that $\sigma_\eta^2 < \sigma_\varepsilon^2$ (see Hamilton (1994) pages 66-68 for discussion). Denote the R^2 that one would obtain in this $ARMA$ by $R_\eta^2 = 1 - \frac{\sigma_\eta^2}{\sigma_y^2}$ (note that since the η_t cannot be obtained from the history of the process y_t one could never in fact run this regression but that does not prevent us calculating its properties).

3. Fundamental and Non-Fundamental Predictive Systems

It is possible to write the fundamental $ARMA$ model as a system of the same form as the original predictive system

$$\begin{aligned}
y_t &= \beta_1^\varepsilon x_{1t-1}^\varepsilon + \beta_2^\varepsilon x_{2t-1}^\varepsilon + \ldots + \beta_p^\varepsilon x_{pt-1}^\varepsilon + \varepsilon_t \\
x_{1t}^\varepsilon &= \lambda_1 x_{1t-1}^\varepsilon + \varepsilon_t \\
x_{2t}^\varepsilon &= \lambda_2 x_{2t-1}^\varepsilon + \varepsilon_t \\
&\vdots \\
x_{pt}^\varepsilon &= \lambda_p x_{pt-1}^\varepsilon + \varepsilon_t
\end{aligned}$$

where note that the R^2 in this predictive regression will equal R_ε^2. Note carefully that the processes x_{1t}^ε, etc. differ from the original x_{1t} etc. The coefficients β_1^ε etc differ from the original β_1 etc. and the covariance matrix here is singular since the innovations are identical. The autoregressive parameters λ_1 etc are the same as in the original formulation.

The coefficients β_i^ε can be derived from the θ_i by equating coefficients (of L^j) in the relation

$$1+\frac{\beta_1^\varepsilon L}{1-\lambda_1 L}+\ldots+\frac{\beta_p^\varepsilon L}{1-\lambda_p L}=\frac{(1-\theta_1 L)(1-\theta_2 L)\ldots(1-\theta_p L)}{(1-\lambda_1 L)(1-\lambda_2 L)\ldots(1-\lambda_p L)}$$

Likewise we can write the major-non-fundamental $ARMA$ as a system

$$\begin{aligned}
y_t &= \beta_1^\eta x_{1t-1}^\eta+\beta_2^\eta x_{2t-1}^\eta+\ldots+\beta_p^\eta x_{pt-1}^\eta+\eta_t\\
x_{1t}^\eta &= \lambda_1 x_{1t-1}^\eta+\eta_t\\
x_{2t}^\eta &= \lambda_2 x_{2t-1}^\eta+\eta_t\\
&\vdots\\
x_{pt}^\eta &= \lambda_p x_{pt-1}^\eta+\eta_t
\end{aligned}$$

where note that the R^2 in this predictive regression will equal R_η^2. Note carefully that the processes x_{1t}^η, etc. differ from the original x_{1t} etc. The coefficients β_1^η etc differ from the original β_1 etc. and the covariance matrix here is singular since the innovations are identical. The autoregressive parameters λ_1 etc are the same as in the original formulation.

The coefficients β_i^η can be derived from the θ_i by equating coefficients (of L^j) in the relation

$$1+\frac{\beta_1^\eta L}{1-\lambda_1 L}+\ldots+\frac{\beta_p^\eta L}{1-\lambda_p L}=\frac{\left(1-\theta_1^{-1} L\right)\left(1-\theta_2^{-1} L\right)\ldots\left(1-\theta_p^{-1} L\right)}{(1-\lambda_1 L)(1-\lambda_2 L)\ldots(1-\lambda_p L)}$$

4. Bounds on the Predictive R^2

We have the following:

Proposition

(i) In the regression

$$y_t=\gamma_1 x_{1t-1}+\ldots+\gamma_p x_{pt-1}+\delta_1 x_{1t-1}^\varepsilon+\ldots+\delta_p x_{pt-1}^\varepsilon+\xi_t$$

then $\delta_1=\ldots=\delta_p=0$.

(ii) In the regression

$$y_t=\tilde{\gamma}_1 x_{1t-1}+\ldots+\tilde{\gamma}_p x_{pt-1}+\tilde{\delta}_1 x_{1t-1}^\eta+\ldots+\tilde{\delta}_p x_{pt-1}^\eta+\zeta_t$$

then $\tilde{\gamma}_1=\ldots=\tilde{\gamma}_p=0$.

Proof

First note that in the regression

$$y_t=\gamma_1 x_{1t-1}+\ldots+\gamma_p x_{pt-1}+\delta_1 x_{1t-1}^\varepsilon+\ldots+\delta_p x_{pt-1}^\varepsilon+\xi_t$$

we will have $\delta_1 = \ldots = \delta_p = 0$ if we establish that $x^{\varepsilon}_{1t-1}, \ldots, x^{\varepsilon}_{pt-1}$ are orthogonal to u_t , since $u_t = y_t - \beta_1 x_{1t-1} + \beta_2 x_{2t-1} + \ldots + \beta_p x_{pt-1}$.

We have

$$\begin{aligned} y_t = \frac{\prod_{i=1}^{p}(1-\theta_i L)}{\prod_{i=1}^{p}(1-\lambda_i L)}\varepsilon_t &= \beta_1 x_{1t-1} + \beta_2 x_{2t-1} + \ldots + \beta_p x_{pt-1} + u_t \\ &= \beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_p \frac{v_{pt-1}}{1-\lambda_p L} + u_t \end{aligned}$$

So we may write

$$x^{\varepsilon}_{jt-1} = \frac{\varepsilon_{t-1}}{1-\lambda_j L} = \left(\frac{L}{1-\lambda_j L}\right) \frac{\prod_{i=1}^{p}(1-\lambda_i L)}{\prod_{i=1}^{p}(1-\theta_i L)} \left(\beta_1 \frac{L v_{1t}}{1-\lambda_1 L} + \beta_2 \frac{L v_{2t}}{1-\lambda_2 L} + \ldots + \beta_p \frac{L v_{pt}}{1-\lambda_p L} + u_t\right)$$

Given the assumed white noise property of $(u_t, v_t^{'})^{'}$, u_t will be orthogonal to

$$x^{\varepsilon}_{jt-1} = \left(\frac{L}{1-\lambda_j L}\right) \frac{\prod_{i=1}^{p}(1-\lambda_i L)}{\prod_{i=1}^{p}(1-\theta_i L)} \left(\beta_1 \frac{L v_{1t}}{1-\lambda_1 L} + \beta_2 \frac{L v_{2t}}{1-\lambda_2 L} + \ldots + \beta_p \frac{L v_{pt}}{1-\lambda_p L} + u_t\right)$$

since this expression involves only u's and v's dated $t-1$ and earlier. This establishes the first part of the Proposition.

Secondly note that in the regression

$$y_t = \tilde{\gamma}_1 x_{1t-1} + \ldots + \tilde{\gamma}_p x_{pt-1} + \tilde{\delta}_1 x^{\eta}_{1t-1} + \ldots + \tilde{\delta}_p x^{\eta}_{pt-1} + \zeta_t$$

we will have $\tilde{\gamma}_1 = \ldots = \tilde{\gamma}_p = 0$ if we establish that $x_{1t-1}, \ldots, x_{pt-1}$ are orthogonal to η_t, since $\eta_t = y_t - \beta^{\eta}_1 x^{\eta}_{1t-1} + \beta^{\eta}_2 x^{\eta}_{2t-1} + \ldots + \beta^{\eta}_p x^{\eta}_{pt-1}$

We have

$$\begin{aligned} y_t = \frac{\prod_{i=1}^{p}(1-\theta_i^{-1} L)}{\prod_{i=1}^{p}(1-\lambda_i L)}\eta_t &= \beta_1 x_{1t-1} + \beta_2 x_{2t-1} + \ldots + \beta_p x_{pt-1} + u_t \\ &= \beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_p \frac{v_{pt-1}}{1-\lambda_p L} + u_t \end{aligned}$$

So

$$\eta_t = \frac{\prod_{i=1}^{p}(1-\lambda_i L)}{\prod_{i=1}^{p}(1-\theta_i^{-1} L)} \left(\beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_p \frac{v_{pt-1}}{1-\lambda_p L} + u_t\right)$$

We can write

$$\frac{1}{1-\theta_i^{-1} L} = \frac{-\theta_i F}{1-\theta_i F}$$

where F is the forward shift operator.

So

$$\eta_t = F^p \left(\prod_{i=1}^{p} \theta_i\right) \left(\frac{\prod_{i=1}^{p}(1-\lambda_i L)}{\prod_{i=1}^{p}(1-\theta_i F)}\right) \left(\beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_p \frac{v_{pt-1}}{1-\lambda_p L} + u_t\right)$$

Now

$$F^p \frac{\prod_{i=1}^p (1-\lambda_i L)}{\prod_{i=1}^p (1-\theta_i F)} \frac{v_{kt-1}}{(1-\lambda_k L)} = v_{kt} + c_1 v_{kt+1} + c_2 v_{kt+2+} \dots$$

for some $c_1, c_2, \dots$ and

$$F^p \frac{\prod_{i=1}^p (1-\lambda_i L)}{\prod_{i=1}^p (1-\theta_i F)} u_t = u_t + b_1 u_{t+1} + b_2 u_{t+2} \dots$$

for some $b_1, b_2, \dots$. So η_t involves only forward values of u_t and v_t and will thus be orthogonal to any $\frac{v_{jt-1}}{1-\lambda_j L}$ by the assumed white noise properties of $(u_t, v_t^{'})^{'}$. This establishes the second part of the Proposition.

Since least squares seeks to minimises the variance of the residual we immediately obtain

Corollary

$$R_\varepsilon^2 \leq R_x^2 \leq R_\eta^2$$

5. Conclusion

In a predictive system with an $ARMA(p,p)$ reduced form the fundamental and non-fundamental $ARMA$ representations place bounds on the predictability of any valid representation of the data (ie any representation consistent with the univariate $ARMA$). Since both fundamental and non-fundamental representations depend only on the set of coefficients $\{\lambda_i, \theta_i\}$, knowledge of the univariate properties of the series allows one to infer the possible degree of predictability. Time series properties that place bounds on the set $\{\lambda_i, \theta_i\}$, for instance the univariate R^2 or the variance ratio as defined by Cochrane (1988), can therefore also give information about predictability.

References

[1] Cochrane, J.H (1988) "How big is the random walk in GDP?. "Journal of Political Economy, v. 96, n. 5, p. 893-92

[2] Hamilton J D (1994)Time Series Analysis 1994 Princeton University Press