Whittle's recursive solution to the multivariate Yule-Walker equations and the prediction of discounted future values

Granville Tunnicliffe Wilson*

October 17, 2017

1 Optimality of the Yule-Walker equations solution.

In chapter 2 of Tunnicliffe Wilson et al. (2015) we introduced the autoregressive linear prediction equation for the current value x_t of a multivariate stationary time series in terms of the previous values up to lag p:

$$\begin{aligned} x_t &= \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + e_t \\ &= \hat{x}_t + e_t. \end{aligned} \tag{1}$$

The Yule-Walker equations determine the matrix coefficients $\Phi_1, \ldots, \Phi_p$ so that the error term e_t is uncorrelated with the predicting terms, i.e.

$$\mathrm{Cov}(e_t, x_{t-k}) = 0 \quad \text{for} \quad k = 1, \ldots, p. \tag{2}$$

Taking the covariance of (1) with x_{t-k} for this range of lags then gives directly the equations associated with the names of Yule and Walker as:

$$\Gamma_k = \Phi_1 \Gamma_{k-1} + \Phi_2 \Gamma_{k-2} + \cdots + \Phi_p \Gamma_{k-p} \quad \text{for} \quad k = 1, \ldots, p. \tag{3}$$

In block matrix form these are

$$\begin{pmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_p \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \end{pmatrix} \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{(p-1)} \\ \Gamma_{-1} & \Gamma_0 & \cdots & \Gamma_{(p-2)} \\ \vdots & \ddots & \ddots & \vdots \\ \Gamma_{-(p-1)} & \Gamma_{-(p-2)} & \cdots & \Gamma_0 \end{pmatrix}. \tag{4}$$

The block matrix on the RHS is the covariance matrix of the succession of time series values x_{t-1}, x_{t-2}, ..., x_{t-p} and is therefore positive definite under the reasonable assumption that no one element of these terms can be exactly predicted from others. The equations therefore always have a unique solution which satisfies the condition (2).

*Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.

Denote the coefficients in the solution of (3) as $\hat{\Phi}_k$, with the predictor $\hat{x}_t$ in (1) having corresponding error $\hat{e}_t$. Now consider any other linear predictor $\tilde{x}_t$ of the same form with coefficients $\tilde{\Phi}_k$ and error $\tilde{e}_t$. It is simple to demonstrate that

$$\operatorname{Var} \tilde{e}_t \geq \operatorname{Var} \hat{e}_t \tag{5}$$

with equality only if $\tilde{x}_t$ and $\hat{x}_t$ are identical, i.e. $\tilde{\Phi}_k = \hat{\Phi}_k$ for $k = 1, \ldots, p$. This follows because

$$\tilde{e}_t = x_t - \tilde{x}_t = \hat{e}_t + (\hat{x}_t - \tilde{x}_t) \tag{6}$$

and on the RHS $\hat{e}_t$ is uncorrelated with the bracketed term which is a linear combination of the lagged values in the predictors. Therefore

$$\operatorname{Var} \tilde{e}_t = \operatorname{Var} \hat{e}_t + \operatorname{Var}(\hat{x}_t - \tilde{x}_t) \geq \operatorname{Var} \hat{e}_t. \tag{7}$$

This also implies that for any vector α,

$$\alpha'(\operatorname{Var} \tilde{e}_t)\, \alpha \geq \alpha'(\operatorname{Var} \hat{e}_t)\, \alpha, \tag{8}$$

and

$$\operatorname{trace} \operatorname{Var} \tilde{e}_t \geq \operatorname{trace} \operatorname{Var} \hat{e}_t, \tag{9}$$

and

$$\det \operatorname{Var} \tilde{e}_t \geq \det \operatorname{Var} \hat{e}_t, \tag{10}$$

with the last two inequalities being strict if $\tilde{x}_t$ and $\hat{x}_t$ are distinct.

2 The Whittle recursions

Whittle (1963) sets out recursions by which the multivariate Yule-Walker equations may be solved in order p^2 operations, rather then p^3, which would be the case using a general equation solver. The recursive scheme has the advantage that it supplies the solutions for all orders p up to a specified maximum. We will briefly reprise this methodology using the notation of the book. As mentioned in section 2.10 of the book these recursions also generate the coefficients in the backward predictor of the form

$$x_t = \tilde{\Phi}_1 x_{t+1} + \tilde{\Phi}_2 x_{t+2} + \cdots + \tilde{\Phi}_p x_{t+p} + \tilde{e}_t. \tag{11}$$

Given both predictors (1) and (11) of order p, the predictors of order $p+1$ are generated as follows. The forward predictor of order p is extended to order $p+1$ by including x_{t-p-1} as a predictor. This can equivalently be achieved by including, instead, the quantity obtained by orthogonalizing x_{t-p-1} with respect to the existing predictors x_{t-1}, x_{t-2},..., x_{t-p}. But this is just the backward error in predicting x_{t-p-1}.

$$\tilde{e}_{t-p-1} = x_{t-p-1} - \tilde{\Phi}_1 x_{t-p} - \tilde{\Phi}_2 x_{t-p+1} - \cdots - \tilde{\Phi}_p x_{t-1} + \tilde{e}_t. \tag{12}$$

Because of the orthogonality, the new forward coefficient Φ_{p+1}, of both the orthogonalized and original variable x_{t-p-1}, is given simply by the solution of

$$\operatorname{Cov}(x_t, \tilde{e}_{t-p-1}) = \Phi_{p+1} \operatorname{Var}(\tilde{e}_{t-p-1}). \tag{13}$$

Hence

$$\Phi_{p+1} = \left(\Gamma_{p+1} - \tilde{\Phi}_1\Gamma_p - \tilde{\Phi}_2\Gamma_{p-1} - \cdots - \tilde{\Phi}_p\Gamma_1\right)\tilde{V}_e^{-1} \tag{14}$$

where $\tilde{V}_e = \mathrm{Var}\,(\tilde{e}_{t-p-1}) = \mathrm{Var}\,(\tilde{e}_t)$. The predictor of lag $p+1$ is then updated to

$$\hat{x}_t \leftarrow \hat{x}_t + \Phi_{p+1}\tilde{e}_{t-p-1} \tag{15}$$

with, on substituting for the two terms in this expression, the updated coefficients

$$\Phi_k \leftarrow \Phi_k - \Phi_{p+1}\tilde{\Phi}_{p+1-k} \quad \text{for} \quad k = 1 \ldots p. \tag{16}$$

From (15) the error of the lag $p+1$ prediction error is updated to

$$e_t \leftarrow e_t + \Phi_{p+1}\tilde{e}_{t-p-1} \tag{17}$$

and because the updated e_t is uncorrelated with $\tilde{e}_{t-p-1}$ its variance V_e is updated to

$$V_e \leftarrow V_e - \Phi_{p+1}\tilde{V}_e\Phi'_{p+1} \tag{18}$$

In the same manner the backward predictor is updated as

$$\tilde{\Phi}_{p+1} = \left(\Gamma'_{p+1} - \Phi_1\Gamma'_p - \Phi_2\Gamma'_{p-1} - \cdots - \Phi_p\Gamma'_1\right)\tilde{V}_e^{-1} \tag{19}$$

where we have written Γ'_k for Γ_{-k},

$$\tilde{\Phi}_k \leftarrow \tilde{\Phi}_k - \tilde{\Phi}_{p+1}\Phi_{p+1-k} \quad \text{for} \quad k = 1 \ldots p \tag{20}$$

and

$$\tilde{V}_e \leftarrow \tilde{V}_e - \tilde{\Phi}_{p+1}V_e\tilde{\Phi}'_{p+1}. \tag{21}$$

Of interest from (13) and the corresponding backward expression is the equality

$$\Phi_{p+1}\,\tilde{V}_e = \mathrm{Cov}\,(x_t, \tilde{e}_{t-p-1}) = \mathrm{Cov}\,(e_t, \tilde{e}_{t-p-1}) = \mathrm{Cov}\,(x_{t-p-1}, e_t)' = V_e\,\tilde{\Phi}'_{p+1} \tag{22}$$

from which (18) can be rewritten

$$V_e \leftarrow \left(I - \Phi_{p+1}\tilde{\Phi}_{p+1}\right)V_e \tag{23}$$

and (21) as

$$\tilde{V}_e \leftarrow \left(I - \tilde{\Phi}_{p+1}\Phi_{p+1}\right)\tilde{V}_e. \tag{24}$$

The determinants of the variances in these last two equations both reduce therefore by the common determinant of $(I - \Phi_{p+1}\tilde{\Phi}_{p+1})$ and $(I - \tilde{\Phi}_{p+1}\Phi_{p+1})$ which can be taken as a generalized measure of the improvement in predictability on increasing the order from p to $p+1$, for which an example is plotted in Figure 2.9 of the book.

The recursions start from the null lag $p = 0$ with $V_e = \tilde{V}_e = V_x$ and appropriate simplification of the equations for the step from $p = 0$ to $p = 1$. A consequence of (23) and (24) is that $\det V_e$ and $\det \tilde{V}_e$ are identical for all orders of the predictor.

3 Extension to prediction of a further series

Suppose now we have another multivariate stationary time series y_t having known cross-covariances with x_t:

$$\mathrm{Cov}(y_t, x_{t-k}) = \Gamma_{y\,x,k}. \tag{25}$$

Then the recursive method may be extended to determine the coefficients of the linear predictor of y_t given x_{t-1}, x_{t-2}, ..., x_{t-p} that is optimal in the sense described in section 1:

$$\begin{aligned} y_t &= \xi_1 x_{t-1} + \xi_2 x_{t-2} + \cdots + \xi_p x_{t-p} + f_t \\ &= \hat{y}_t + f_t. \end{aligned} \tag{26}$$

This equation is extended to order $p+1$ by again including as the new variable the quantity $\tilde{e}_{t-p-1}$ as in (12), rather than x_{t-p-1}, and similarly, the new coefficient ξ_{p+1} is given by the solution of

$$\mathrm{Cov}\left(y_t, \tilde{e}_{t-p-1}\right) = \xi_{p+1}\mathrm{Var}\left(\tilde{e}_{t-p-1}\right). \tag{27}$$

Hence

$$\xi_{p+1} = \left(\Gamma_{y\,x,p+1} - \tilde{\Phi}_1\Gamma_{y\,x,p} - \tilde{\Phi}_2\Gamma_{y\,x,p-1} - \cdots - \tilde{\Phi}_p\Gamma_{y\,x,1}\right)\tilde{V}_e^{-1} \tag{28}$$

The predictor of lag $p+1$ is then updated to

$$\hat{y}_t \leftarrow \hat{y}_t + \xi_{p+1}\tilde{e}_{t-p-1} \tag{29}$$

with, on substituting for the two terms in this expression, the updated coefficients

$$\xi_k \leftarrow \xi_k - \xi_{p+1}\tilde{\Phi}_{p+1-k} \quad \text{for} \quad k = 1 \ldots p. \tag{30}$$

From (15) the error of the lag $p+1$ prediction error is updated to

$$f_t \leftarrow f_t + \xi_{p+1}\tilde{e}_{t-p-1} \tag{31}$$

and because the updated f_t is uncorrelated with $\tilde{e}_{t-p-1}$ its variance V_f is updated to

$$V_f \leftarrow V_f - \xi_{p+1}\tilde{V}_e\xi'_{p+1}. \tag{32}$$

The recursions start from the null lag $p = 0$ with $V_f = V_y$ and appropriate simplification of the updating equations from order p to $p+1$.

4 Prediction of exponentially weighted future values

In section 2.6 of the book an example is given of prediction at lead times $h \geq 1$ from a VAR(p) model that is determined so as to minimize the prediction error variance of an exponentially weighted, or discounted, combination of future series values:

$$x_{t,\rho} = x_{t+1} + \rho x_{t+2} + \rho^2 x_{t+3} + \cdots = \sum_{h=1}^{\infty} \rho^{h-1} x_{t+h}. \tag{33}$$

We now show how the coefficients of this model and its forecast accuracy are derived. There are three steps. First, we derive the coefficients of the linear predictor of $x_{t,\rho}$ in (33) from the most recent p values x_t, x_{t-1}, ..., x_{t-p+1}. Second, from these we determine the coefficients of a VAR(p) model whose predictions (at time t) of x_{t+1}, x_{t+2}, ..., combine to give exactly the same prediction of $x_{t,\rho}$. In fact, this is a specific case of the more general method of determining coefficients of VZAR models introduced in chapter 6 of the book. Finally, we calculate the accuracy, in the form of prediction error variances of this model, as displayed in Figure 2.6 of the book. The point is that the VAR(p) model determined in this way to minimize the prediction error variance of $x_{t,\rho}$, has improved accuracy for prediction at higher lead times using low model orders, when compared with the model determined by solution of the Yule-Walker equations to minimize the prediction error variance at lead time 1.

4.1 The predictor of a discounted sum of future values

This first step still utilizes the Whittle recursive solution, because we wish to predict the series

$$y_{t+1} = x_{t,\rho} = \left(1 - \rho B^{-1}\right)^{-1} x_{t+1} \tag{34}$$

from x_t, x_{t-1}, ..., x_{t-p+1}. We can therefore apply the method of section (3), for which we require

$$\begin{aligned} \Gamma_{y\,x,k} &= \mathrm{Cov}(y_t, x_{t-k}) \\ &= \mathrm{Cov}\left[\left(1 - \rho B^{-1}\right)^{-1} x_t, x_{t-k}\right] \\ &= \mathrm{Cov}\left[x_t, \left(1 - \rho B\right)^{-1} x_{t-k}\right], \end{aligned} \tag{35}$$

where we are using

$$\mathrm{Cov}(B^{-k}x_t, x_t) = \mathrm{Cov}(x_{t+k}, x_t) = \mathrm{Cov}(x_t, x_{t-k}) = \mathrm{Cov}(x_t, B^k x_t) \tag{36}$$

to replace the forward shift operator on the left by the back shift operator on the right in the covariance in (35).

In the example in chapter 2 we use a given VAR(2) model, $\Phi(B)x_t = e_t$ for the set of 5 Pig market series. We can represent these with the addition of the further 5 series $z_t = (1 - \rho B)^{-1}x_{t-1}$ by the block matrix operator model which sets $(1 - \rho B)z_t = x_{t-1}$:

$$\begin{pmatrix} \Phi(B) & 0 \\ -IB & (I - \rho B) \end{pmatrix} \begin{pmatrix} x_t \\ z_t \end{pmatrix} = \begin{pmatrix} e_t \\ 0 \end{pmatrix}. \tag{37}$$

The VAR(2) model covariances of these 10 series are then evaluated as described in the document `VARcovfunCalculation.pdf`. The error vector in (37) is clearly singular but this causes no difficulty for the calculation. The required covariances in (35) are then extracted from the block covariances of the model in (37) as $\Gamma_{y\,x,k} = \Gamma_{x\,z,k-1}$. The recursive method of section 3 then gives the coefficients ξ_k of the order p predictor of the discounted combination of future values of x_t as

$$x_{t,\rho} = \xi_1 x_t + \xi_2 x_{t-1} + \cdots + \xi_p x_{t-p+1} + f_{t+1} = \xi(B)x_t + f_{t+1}. \tag{38}$$

4.2 The VAR model predictor of the discounted sum of future values

We seek in this section to establish coefficients $\Phi_1, \ldots, \Phi_p$ of the VAR(p) model $\Phi(B)x_t = e_t$ which gives a prediction of the discounted future values $x_{t,\rho}$ in (33) that coincides with that given by (38).

Given present and past values of x_{t-k} for $k \geq 0$, the forecast future values x_{t+k} for $k > 0$ obtained using the VAR(p) model are generated such that

$$\Phi(B)x_{t+k} = e_{t+k} = 0 \quad \text{for} \quad k > 0. \tag{39}$$

In the following development, for simplicity of notation, we will not distinguish between the present and past known, and the future forecast values of the series x_t and e_t related in this manner. We now express the *forecast* future discounted quantity using these values as

$$\begin{aligned} x_{t,\rho} &= \frac{B^{-1}}{1-\rho B^{-1}} x_t \\ &= \frac{B^{-1}}{1-\rho B^{-1}} \Phi(B)^{-1} e_t \\ &= \left[N \frac{B^{-1}}{1-\rho B^{-1}} + \xi(B)\Phi(B)^{-1} \right] e_t \end{aligned} \tag{40}$$

Where this last expression is obtained using partial fractions in which N is a constant matrix coefficient and is given together with $\xi(B)$ by solving

$$N\Phi(B) = (B-\rho)\xi(B) \equiv I. \tag{41}$$

Because the first term in the last line of (40) is a combination of future values of e_t, it is zero. The second term is a combination of present and past values of e_t, and therefore x_t, giving the forecast value of

$$x_{t,\rho} = \xi(B)x_t. \tag{42}$$

It is evident that our choice of notation $\xi(B)$ in the partial fraction of (40) anticipated its giving the required predictor of $x_{t,\rho}$ in (42).

Thus from (41) we can always determine the required VAR(p) model as

$$\Phi(B) = N^{-1}\left[I - (B-\rho)\xi(B)\right]. \tag{43}$$

The constant N is readily determined by setting $B = 0$ as $N = I + \rho\xi_1$. From(40) we can also determine the forecast error variance implied by the VAR model as

$$\text{Var}\left(N \frac{B^{-1}}{1-\rho B^{-1}} e_t \right) = (1-\rho^2)^{-1} N V_e N', \tag{44}$$

where we are now taking the true (not forecast) future values of e_t and using the modeling assumption that these are uncorrelated with variance V_e.

4.3 The multi-step prediction variance using the derived predictor

Figure (2.6) in the book shows the prediction error variances obtained when forecasts are generated using VAR(p) models derived by the methods described in the previous sections. The application is to two of the series from the set of 5 Pig market series. The VAR(2) model assumed for the full set of these series implies that this subset cannot be represented by any finite order bivariate VAR. All forecasts obtained by VAR approximations are therefore sub-optimal, although the Figure indicates that all models of order 10 or more are effectively optimal.

To compute the actual error variance resulting from the use of an assumed VAR(p) model is straightforward. First, the implied coefficients of the predictor at the chosen lead time h are determined from the model coefficients as in:

$$\hat{x}_{t,h} = \Phi_1^h x_t + \Phi_2^h x_{t-1} + \cdots + \Phi_p^h x_{t-p+1}. \tag{45}$$

Then the variance is evaluated, of the multi-step prediction error

$$e_{t,h} = x_{t+h} - \Phi_1^h x_t - \Phi_2^h x_{t-1} - \cdots + \Phi_p^h x_{t-p+1}, \tag{46}$$

using the known true lagged covariances of the series terms on the RHS of (46).

All the computations used to implement the methods in this section for the example shown in Figure 2.6 of the book, are in the MATLAB script `pigsmodelMSTEP.m`.

5 Stationarity of the Yule-Walker solution

As expressed in section 2.5 of the book, the necessary and sufficient condition upon the coefficients of the VAR(p) model (1) for it to represent a stationarity process is expressed in terms of the matrix polynomial

$$\Phi(z) = I - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p. \tag{47}$$

A convenient statement of the condition is that $\det \Phi(z)$ has no zeros for $|z| \leq 1$. This can be directly verified algebraically. Equivalently, in the formal expansion of

$$\Phi(z)^{-1} = I + \Psi_1 x + \Psi_2 z^2 + ... = \sum_{k=0}^{\infty} \Psi_k z^k \tag{48}$$

the coefficients Ψ_k converge to zero. In that case their convergence is bounded geometrically.

An important result is that the coefficients of an approximating VAR(p) model obtained by solution of the Yule-Walker equations always satisfy the stationarity condition, provided only that the covariance matrix G_{p+1} of the succession of time series values x_t, x_{t-1}, ..., x_{t-p} is positive definite. (Substituting the sample lagged covariances defined in (3.35) of the book this condition is generally satisfied for the order $p < n$.) We now outline Whittle's demonstration of this result. Further, if the whole sequence of lagged covariances Γ_k is positive definite, i.e. G_p is positive for all p, then the VAR(p) model determined by prediction of the sum of discounted future values, as in section 4.2 above, also satisfies the stationarity condition. The proof of this will be given in the documents relating to the VZAR models of Chapter 6 of the book.

5.1 Whittle's proof of stationarity - an outline

A proof is given in Whittle (1963), by writing the model in what we term the state space form (2.72) and (2.73) of the book, reproduced here:

$$\begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \vdots \\ \cdots & 0 & I & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{49}$$

and its concisely representation

$$S_t = T S_{t-1} + E_t. \tag{50}$$

For the approximating model of order p obtained by solution of the Yule-Walker equations, because S_{t-1} and E_t are orthogonal by construction, the variance matrices V_S and V_E of S_t and E_t are related by

$$V_S = T\, V_S T' + V_E. \tag{51}$$

Whittle shows (in his own notation) that this relationship immediately determines that any eigenvalue λ of T must satisfy $|\lambda| \leq 1$. With a little more consideration he demonstrates that $|\lambda| = 1$ can only occur if there is an exact linear relationship between the elements of x_t, x_{t-1}, ..., x_{t-p}, which is precluded by our assumption that G_{p+1} is positive definite. Now the eigenvalues of T are the zeros of $\det \Phi(z^{-1})$, which establishes the result.

5.2 Subset autoregressions - a caution

Because the VAR(p) model derived from a positive definite sequence of autocovariances Γ_k by solution of the Yule-Walker equations is always stationary, it defines a process whose autocovariances agree up to lag p with those, Γ_k, from which the model was derived. The Yule-Walker equations may be re-formed as a set of linear equations which may be uniquely solved for these autocovariances given the coefficients of the derived VAR(p).

When deriving a VAR(p) model from sample autocovariances, coefficients at particular lags may not be significantly different from zero, and sometimes, sensibly, the model is re-derived by solving the Yule-Walker equations with the corresponding coefficients and equations removed. However, we mention on pages 24 and 25 of the book a caution that the subset autoregression so obtained does not necessarily share the desirable properties mentioned in the previous paragraph. The derived model is not necessarily stationary, and even if it is, the autocovariances of this model may not agree with those used to derive its coefficients. We demonstrate this with two simple univariate examples of AR(3) models.

First take the covariances given in the row 2 of Table 1, which are those of the model given in the row 3. Row 4 shows the coefficients of the subset AR(3) derived using these, assuming that the coefficient at lag two is zero (in bold), regressing only on lags 1 and 3, and discarding the second of the Yule-Walker equations. This model is in fact stationary, but its autocovariances, shown in the last row, do not coincide with those used to derive the model.

k	0	1	2	3
$\gamma(k)$	10.0	-9.0	8.0	-6.0
σ^2, ϕ_k	1.25	-0.9167	0.5	0.5833
σ^2, ϕ_k	1.5	-1.1667	**0**	0.3333
$\gamma(k)$	12.5455	-11.4545	9.5455	-6.9545

Table 1: A stationary subset model

A second example is shown in Table 2. Again, rows 2 and 3 show the autocovariances to lag 3 and coefficients of a corresponding stationary AR(3) model. Row 4 shows the derived subset autoregression with the zero coefficient (in bold again) at lag two. However, this model is not stationary. There is a zero of $\phi(z)$ at $z = 0.7668$.

k	0	1	2	3
$\gamma(k)$	10.0000	-3.5000	-7.0000	9.0000
σ^2, ϕ_k	0.1771	0.1778	-0.3182	0.9131
σ^2, ϕ_k	0.3627	0.5490	**0**	1.2843

Table 2: A non-stationary subset model

References

G. Tunnicliffe Wilson, M. Reale, and J. Haywood. *Models for dependent time series*. New York, CRC Press, 2015.

P. Whittle. On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50:129–134, 1963.