Recursive Least Squares

0. Introduction

Our model is the simple regression model. What we have in mind is a reaction function relating the policy objective (say inflation) and an instrument (say an interest rate). If a policy maker uses this equation to set interest rates, he is always confronted with the question of the stability of the estimated equation. We will learn i) how the estimated coefficients are revised as information comes in and ii) some tests of the stability of the equation.

At the same time this lecture is a good introduction for more complex models, for example TVP (time varying parameter) models with which we will deal later on. Most of what we are going to learn in this section will remain true. The material of this chapter will then be a special case.

Literature:

A. C. Harvey, Time Series Models, Harvester Wheatsheaf, 1981 (first edition), 1993 (second edition).

J. Johnston and J. DiNardo, Econometric Methods, 1997 (fourth edition).

1. Recursive Least Squares

The model:

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \cdots + \beta_K x_{tK} + \varepsilon_t \quad t = 1,\ldots,T \tag{0.1}$$

Matrix notation:

$$\underset{T\times 1}{\mathbf{Y}_T} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_T \end{bmatrix} \quad \underset{T\times k}{\mathbf{X}_T} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} & \cdots & x_{1K} \\ 1 & x_{22} & \cdots & x_{2k} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 1 & x_{t2} & \cdots & x_{tk} & \cdots & x_{tK} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 1 & x_{2T} & \cdots & x_{TK} & \cdots & x_{TK} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_t' \\ \vdots \\ \mathbf{x}_T' \end{bmatrix} \quad \underset{K\times 1}{\boldsymbol{\beta}_T} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_k \\ \\ \beta_K \end{bmatrix} \quad \underset{T\times 1}{\boldsymbol{\varepsilon}_T} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_T \end{bmatrix} \tag{0.2}$$

$$\mathbf{Y}_T = \mathbf{X}_T \boldsymbol{\beta}_T + \boldsymbol{\varepsilon}_T \tag{0.3}$$

Some usual assumptions:

$$E\left(\boldsymbol{\varepsilon}_T\right) = \mathbf{0}, \quad Var\left(\boldsymbol{\varepsilon}_T \boldsymbol{\varepsilon}_T^{'}\right) = \sigma^2 \mathbf{I}_T \tag{0.4}$$

The least squares estimator:

$$\mathbf{b}_T = \left(\mathbf{X}_T^{'} \mathbf{X}_T\right)^{\mathbf{-1}} \mathbf{X}_T^{'} \mathbf{y}_T = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{'}\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t y_t \tag{0.5}$$

The summation notation for beta gives some insights in recursions.

For K=2, T=3:

$$\mathbf{b}_3 = \left(\begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \right)^{-1} \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= \begin{pmatrix} \sum_{t=1}^{3} x_{t1}^2 & \sum_{t=1}^{3} x_{t1}x_{t2} \\ \sum_{t=1}^{3} x_{t2}x_{t1} & \sum_{t=1}^{3} x_{t2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{3} x_{t1}y_t \\ \sum_{t=1}^{3} x_{t2}y_t \end{pmatrix}$$

$$\mathbf{b}_T = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{'} \right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t y_t$$

$$= \left(\begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} + \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} \begin{bmatrix} x_{21} & x_{22} \end{bmatrix} + \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \begin{bmatrix} x_{31} & x_{32} \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} y_1 + \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} y_2 + \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} y_3 \right)$$

$$= \left(\begin{bmatrix} x_{11}^2 & x_{11}x_{12} \\ x_{12}x_{11} & x_{12}^2 \end{bmatrix} + \begin{bmatrix} x_{21}^2 & x_{21}x_{22} \\ x_{22}x_{21} & x_{22}^2 \end{bmatrix} + \begin{bmatrix} x_{31}^2 & x_{31}x_{32} \\ x_{31}x_{31} & x_{32}^2 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} x_{11}y_1 \\ x_{12}y_1 \end{bmatrix} + \begin{bmatrix} x_{21}y_2 \\ x_{22}y_2 \end{bmatrix} + \begin{bmatrix} x_{31}y_3 \\ x_{32}y_3 \end{bmatrix} \right) =$$

$$= \begin{pmatrix} \sum_{t=1}^{3} x_{t1}^2 & \sum_{t=1}^{3} x_{t1}x_{t2} \\ \sum_{t=1}^{3} x_{t2}x_{t1} & \sum_{t=1}^{3} x_{t2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{3} x_{t1}y_t \\ \sum_{t=1}^{3} x_{t2}y_t \end{pmatrix}$$

By inspection it follows that:

$$\mathbf{X}_T^{'}\mathbf{X}_T = \mathbf{X}_{T-1}^{'}\mathbf{X}_{T-1} + \mathbf{x}_T\mathbf{x}_T^{'} \tag{0.6}$$

and

$$\mathbf{X}_T^{'}\mathbf{y}_T = \mathbf{X}_{T-1}^{'}\mathbf{y}_{T-1} + \mathbf{x}_T y_T \tag{0.7}$$

This gives the idea to determine beta recursively.

Exercise

The cash inflow y_t of a firm in period t is given by $y_t = \mu + \varepsilon_t \quad \varepsilon_t \sim IID\left(0, \sigma^2\right)$. The manager looks for an estimator of μ.

i) Show that the sample mean $\overline{y}_T = \frac{\sum_{t=1}^{T} y_t}{T}$ is the Least Squares estimator of μ: $\hat{\mu} = \overline{y}$. Hint: Minimize the sum of squared residuals with respect to μ: $\min_{\mu}\left\{ S = \sum_{t=1}^{T}\left(y_t - \mu\right)^2 \right\}$

ii) Show that the estimator of μ can be written in a recursive form, so that it can be updated as the sample of observed cash inflows increases from T to T+1. Hint: Show that $\overline{y}_{T+1} = \overline{y}_T + \left(y_{T+1} - \overline{y}_T\right) / \left(T+1\right)$

Using a partial information set, various estimates **b**$_t$ t=k,…,T can be computed.

Information in t:

$$\mathbf{y}_t \; \mathbf{X}_t$$

$$\mathbf{b}_t = \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1} \mathbf{X}_t'\mathbf{y}_t$$

$$Var\left(\mathbf{b}_t\right) = \sigma^2 \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1}$$

Information in t+1:

$$\mathbf{y}_{t+1} = \begin{bmatrix} \mathbf{y}_t \\ y_{t+1} \end{bmatrix} \quad \mathbf{X}_{t+1} = \begin{bmatrix} \mathbf{X}_t \\ \mathbf{x}'_{t+1} \end{bmatrix}$$

$$\mathbf{b}_{t+1} = \left(\mathbf{X}_{t+1}'\mathbf{X}_{t+1}\right)^{-1} \mathbf{X}_{t+1}'\mathbf{y}_{t+1}$$

$$Var\left(\mathbf{b}_{t+1}\right) = \sigma^2 \left(\mathbf{X}_{t+1}'\mathbf{X}_{t+1}\right)^{-1}$$

How is $\mathbf{b}_t$ revised in function of the new information?

$$\mathbf{b}_{t+1} = \mathbf{b}_t + \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1} \mathbf{x}_{t+1} \frac{\left(y_{t+1} - \mathbf{x}_{t+1}'\mathbf{b}_t\right)}{f_t} \tag{0.8}$$

$$\left(\mathbf{X}_{t+1}'\mathbf{X}_{t+1}\right)^{-1} = \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1} - \frac{\left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1} \mathbf{x}_{t+1}\mathbf{x}_{t+1}' \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1}}{f_{t+1}} \tag{0.9}$$

$$f_{t+1} = 1 + \mathbf{x}_{t+1}' \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1} \mathbf{x}_{t+1} \tag{0.10}$$

We call these equations the updating equation for $\mathbf{b}_t$ and $\left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1}$.

The interpretation is easier if one uses a less cumbersome notation:

$$Var\left(\mathbf{b}_t\right) = \sigma^2 \left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1} = \mathbf{P}_t \tag{0.11}$$

$\mathbf{P}_t$ is the covariance matrix of the coefficient estimates. The updating equation for $\mathbf{b}_t$ becomes:

$$\mathbf{b}_{t+1} = \mathbf{b}_t + \mathbf{P}_t\mathbf{x}_{t+1}\frac{\left(y_{t+1} - \mathbf{x}_{t+1}^{'}\mathbf{b}_t\right)}{F_t} \tag{0.12}$$

The second updating equation becomes:

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \frac{\mathbf{P}_t\mathbf{x}_{t+1}\mathbf{x}_{t+1}^{'}\mathbf{P}_t}{F_{t+1}} \tag{0.13}$$

and

$$F_{t+1} = \sigma^2 f_{t+1} = \sigma^2 + \mathbf{x}_{t+1}^{'}\mathbf{P}_t\mathbf{x}_{t+1} \tag{0.14}$$

See the exercises 1 and 2 for a derivation of the expressions (0.8) and (0.9).

Intuitive Interpretation

$$\mathbf{b}_{t+1} = \mathbf{b}_t + \mathbf{P}_t \mathbf{x}_{t+1} \frac{\left(y_{t+1} - \mathbf{x}_{t+1}^{'} \mathbf{b}_t \right)}{F_t}$$

Discussion of factors:

i) The prediction error weighted by its variance.
ii) The less precise $\mathbf{b}_t$, the bigger $Var(\mathbf{b}_t) = \mathbf{P}_t$ which in turn implies that more weight is given to the forecast error (innovation).
iii) The size of $\mathbf{x}_{t+1}$ (sign!)

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \frac{\mathbf{P}_t \mathbf{x}_{t+1} \mathbf{x}_{t+1}^{'} \mathbf{P}_t}{F_{t+1}}$$

Discussion of factors:

iv) The second term can only be positive or null. Therefore the variance can only decrease with time.
v) If linear dependence there is no learning: $\mathbf{P}_t \mathbf{x}_{t+1} = \mathbf{0}$.

F will be shown to be the variance of the recursive residuals.

2. Recursion and Initialization

The updating formulae are:

$$\mathbf{b}_{t+1} = \mathbf{b}_t + \mathbf{P}_t \mathbf{x}_{t+1} \frac{\left(y_{t+1} - \mathbf{x}'_{t+1} \mathbf{b}_t \right)}{F_t}$$

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \frac{\mathbf{P}_t \mathbf{x}_{t+1} \mathbf{x}'_{t+1} \mathbf{P}_t}{F_{t+1}}$$

$$F_{t+1} = \sigma^2 + \mathbf{x}'_{t+1} \mathbf{P}_t \mathbf{x}_{t+1}$$

i) Start with $\mathbf{b_t}$ and Var($\mathbf{b_t}$)=$\mathbf{P_t}$
ii) Observe the new data: y_{t+1}, $\mathbf{x}_{t+1}$
iii) Compute $\mathbf{b_{t+1}}$ and F_{t+1}, $\mathbf{P_{t+1}}$

No matrix inversion is needed.

There are two problems left:
- How to find start values? (Intialization)
- How to deal with σ^2?

Initialization:

First approach:

Start the recursion at t=K. Use LS with the first k observations:

$$
\begin{aligned}
Var\left(\mathbf{b}_K\right) = \mathbf{P}_K &= \sigma^2\left(\mathbf{X}_K'\mathbf{X}_K\right)^{-1} = \sigma^2\left(\mathbf{X}_K\right)^{-1}\left(\mathbf{X}_K'\right)^{-1} = \sigma^2\left(\mathbf{X}_K\right)^{-1}\left(\mathbf{X}_K^{-1}\right)' \\
\mathbf{b}_K &= \left(\mathbf{X}_K'\mathbf{X}_K\right)^{\mathbf{-1}}\mathbf{X}_K'\mathbf{y}_K = \left(\mathbf{X}_K\right)^{-1}\left(\mathbf{X}_K'\right)^{-1}\mathbf{X}_K'\mathbf{y}_K = \left(\mathbf{X}_K\right)^{-1}\mathbf{y}_K
\end{aligned}
$$

A nice feature of recursive least squares is that only one matrix inversion needs to be computed, namely $\mathbf{X}_K^{-1}$ at the start.

Second approach:

Start the recursion at t=0.

Use an arbitrary value of $\mathbf{b}_0$: $\mathbf{b}_0 = 0$

Give an large uncertainty to the estimate of $\mathbf{b}_0$: $Var\left(\mathbf{b}_0\right) = \mathbf{P}_0 = \kappa I_K = 10^6 I_K$

The estimates $\mathbf{b}_1, \ldots, \mathbf{b}_{K-1}$ and the prediction errors $\tilde{v}_1, \ldots, \tilde{v}_K$ have no statistical meaning.

3. LS-Residuals

The prediction errors (recursive residuals) have interesting properties that the LS residuals do not have. Here we summarize some of the properties of the LS-residuals.

The $T \times 1$ vector of LS residuals is

$$\mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix} \tag{0.15}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\mathbf{y} = \mathbf{M}\mathbf{y} \tag{0.16}$$

The $T \times T$ Matrix **M** is orthogonal to **X**, i.e. $\mathbf{MX} = \mathbf{0}$. Therefore:

$$\mathbf{e} = \mathbf{M}\left(\mathbf{X}\beta + \boldsymbol{\varepsilon}\right) = \mathbf{M}\boldsymbol{\varepsilon} \tag{0.17}$$

Thus the LS residuals are linear combinations of the true disturbances. Therefore

$$E\left(\mathbf{e}\right) = \mathbf{M}E\left(\boldsymbol{\varepsilon}\right) = \mathbf{0} \tag{0.18}$$

$$E\left(\mathbf{e}\mathbf{e}^{'}\right) = \mathbf{M} \cdot E\left(\mathbf{e}\mathbf{e}^{'}\right) \cdot \mathbf{M}^{'} = \sigma^2\mathbf{M} \tag{0.19}$$

The last equation follows from **M** being symmetric and idempotent, i.e. $\mathbf{MM} = \mathbf{M}$.

Since $\mathbf{M} \neq \mathbf{I}$, the LS residuals are neither homoscedastic nor serially uncorrelated. Hence their properties do no mirror those of the true disturbances.

For a small sample it may be difficult to detect misspecification of the model (e.g. structural breaks) with the LS-residuals.

When T is large, M is closely approximated by the identity matrix and thus $\mathbf{e}$ is close to $\boldsymbol{\varepsilon}$.

4. Recursive Residuals

The $\left(T-K\right)\times 1$ vector of prediction errors is

$$\mathbf{v}_T = \begin{bmatrix} v_{K+1} \\ \vdots \\ v_t \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} y_{K+1} - \mathbf{x}_{K+1}^{'}\mathbf{b}_K \\ \vdots \\ y_t - \mathbf{x}_t^{'}\mathbf{b}_{t-1} \\ \vdots \\ y_T - \mathbf{x}_T^{'}\mathbf{b}_{T-1} \end{bmatrix} \tag{0.20}$$

The standardized prediction errors are called *recursive residuals*.

Properties:

1) The recursive residuals are LUS residuals, i.e. linear, unbiased with scalar covariance matrix.

Let us prove the result for a general case. T_1 observations are used to estimate $\mathbf{b}_{T_1}$. Forecasts are produced for the T_2 remaining observations. The available number of observations is $T_1+T_2=T$.

$$\mathbf{v}_{T_2} = \begin{bmatrix} v_{T_1+1} \\ v_{T_1+2} \\ \vdots \\ v_{T_1+T_2} \end{bmatrix} = \begin{bmatrix} y_{T_1+1} - \mathbf{x}_{T_1+1}^{'}\mathbf{b}_{T_1} \\ y_{T_1+2} - \mathbf{x}_{T_1+2}^{'}\mathbf{b}_{T_1} \\ \cdots \\ y_{T_1+T_2} - \mathbf{x}_{T_1+T_2}^{'}\mathbf{b}_{T_1} \end{bmatrix} = \begin{bmatrix} y_{T_1+1} \\ y_{T_1+2} \\ \vdots \\ y_{T_1+T_2} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_{T_1+1}^{'} \\ \mathbf{x}_{T_1+2}^{'} \\ \vdots \\ \mathbf{x}_{T_1+T_2}^{'} \end{bmatrix} \mathbf{b}_{T_1} = \mathbf{y}_{T_2} - \mathbf{X}_{T_2}\mathbf{b}_{T_1} \tag{0.21}$$

Take care to note that the notation is unusual: The expressions $\mathbf{v}_{T_2}$ and $\mathbf{X}_{T_2}$ include the observations from T_1+1 to T_2 (and not 1 to T_2).

Inserting the last T_2 equations of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ in (0.21) gives:

$$\mathbf{v}_{T_2} = \mathbf{y}_{T_2} - \mathbf{X}_{T_2}\mathbf{b}_{T_1} = \boldsymbol{\varepsilon}_{T_2} - \mathbf{X}_{T_2}\left(\mathbf{b}_{T_1} - \boldsymbol{\beta}\right) \tag{0.22}$$

Inserting (0.5) gives

$$\mathbf{v}_{T_2} = \mathbf{y}_{T_2} - \mathbf{X}_{T_2}\mathbf{b}_{T_1} = \boldsymbol{\varepsilon}_{T_2} - \mathbf{X}_{T_2}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{-1}\mathbf{X}_{T_1}\boldsymbol{\varepsilon}_{T_1} \tag{0.23}$$

If the disturbances are normal, the prediction errors are also normal. Furthermore, the prediction errors are unbiased:

$$E\left(\mathbf{v}_{T_2}\right)=E\left(\boldsymbol{\varepsilon}_{T_2}\right)-\mathbf{X}_{T_2}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{\mathbf{-1}}\mathbf{X}_{T_1}E\left(\boldsymbol{\varepsilon}_{T_1}\right)=\mathbf{0} \qquad (0.24)$$

The variance-covariance matrix of the forecast errors is:

$$\begin{aligned}\operatorname{var}\left(\mathbf{v}_{T_2}\right)&=E\left(\mathbf{v}_{T_2}\mathbf{v}_{T_2}^{'}\right)\\&=\sigma^2\mathbf{I}_{T_2}+\mathbf{X}_{T_2}\cdot\operatorname{var}\left(\mathbf{b}_{T_1}\right)\cdot\mathbf{X}_{T_2}^{'}\\&=\sigma^2\left[\mathbf{I}_{T_2}+\mathbf{X}_{T_2}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{-1}\mathbf{X}_{T_2}^{'}\right]\end{aligned} \qquad (0.25)$$

The predictions errors of the T$_2$-step-ahead forecasts are correlated. In the special case of a one-step-ahead prediction error, i.e. T$_1$=T-1 and T$_2$=1, we get:

$$Var\left(v_T\right)=\sigma^2\left[1+\mathbf{x}_T^{'}\left(\mathbf{X}_{T-1}^{'}\mathbf{X}_{T-1}\right)^{-1}\mathbf{x}_T\right]=\sigma^2 f_T=F_T \qquad (0.26)$$

Thus F_T is the variance of the one-step-ahead prediction error. In the updating equation for $\mathbf{b}_T$ the term f_T weights the forecast error.

The proof is as follows:

$$\begin{aligned}\operatorname{var}\left(\mathbf{v}_{T_2}\right)&=E\left(\mathbf{v}_{T_2}\mathbf{v}_{T_2}^{'}\right)\\&=E\left[\left(\boldsymbol{\varepsilon}_{T_2}-\mathbf{X}_{T_2}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{\mathbf{-1}}\mathbf{X}_{T_1}^{'}\boldsymbol{\varepsilon}_{T_1}\right)\left(\boldsymbol{\varepsilon}_{T_2}-\mathbf{X}_{T_2}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{\mathbf{-1}}\mathbf{X}_{T_1}\boldsymbol{\varepsilon}_{T_1}\right)^{'}\right]\\&=E\left(\boldsymbol{\varepsilon}_{T_2}\boldsymbol{\varepsilon}_{T_2}^{'}\right)+\mathbf{X}_{T_2}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{\mathbf{-1}}\mathbf{X}_{T_1}^{'}\cdot E\left(\boldsymbol{\varepsilon}_{T_1}\boldsymbol{\varepsilon}_{T_1}^{'}\right)\cdot\mathbf{X}_{T_1}\left(\mathbf{X}_{T_1}^{'}\mathbf{X}_{T_1}\right)^{\mathbf{-1}}\mathbf{X}_{T_2}^{'}\end{aligned} \qquad (0.27)$$

The cross-product terms vanish since $E\left(\mathbf{v}_{T_1}\mathbf{v}_{T_2}^{'}\right)=\mathbf{0}$ by assumption. Substituting $E\left(\boldsymbol{\varepsilon}_{T_2}\boldsymbol{\varepsilon}_{T_2}^{'}\right)=\sigma^2\mathbf{I}_{T_2}$ gives the desired result (0.25).

The one-step ahead prediction errors are uncorrelated. For all t and s such that $t < T$ and $K < s < t$ the correlation is

$$\begin{aligned}
E\left(\nu_t \nu_s\right) &= E\left[\left(\varepsilon_t - \mathbf{x}_t^{'}\left(\mathbf{X}_{t-1}^{'}\mathbf{X}_{t-1}\right)^{\mathbf{-1}}\mathbf{X}_{t-1}^{'}\boldsymbol{\varepsilon}_{t-1}\right)\left(\varepsilon_s - \mathbf{x}_s^{'}\left(\mathbf{X}_{s-1}^{'}\mathbf{X}_{s-1}\right)^{\mathbf{-1}}\mathbf{X}_{s-1}\boldsymbol{\varepsilon}_{s-1}\right)^{'}\right] \\
&= E\left(\varepsilon_t\varepsilon_s\right) \\
&\quad -E\left(\varepsilon_t\boldsymbol{\varepsilon}_{s-1}^{'}\right)\cdot\mathbf{X}_{s-1}^{'}\left(\mathbf{X}_{s-1}^{'}\mathbf{X}_{s-1}\right)^{\mathbf{-1}}\mathbf{x}_s \\
&\quad -\mathbf{x}_t^{'}\left(\mathbf{X}_{t-1}^{'}\mathbf{X}_{t-1}\right)^{\mathbf{-1}}\mathbf{X}_{t-1}^{'}\cdot E\left(\boldsymbol{\varepsilon}_{t-1}\varepsilon_s\right) \\
&\quad +\mathbf{x}_t^{'}\left(\mathbf{X}_{t-1}^{'}\mathbf{X}_{t-1}\right)^{\mathbf{-1}}\mathbf{X}_{t-1}^{'}\cdot E\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{s-1}^{'}\right)\cdot\mathbf{X}_{s-1}^{'}\left(\mathbf{X}_{s-1}^{'}\mathbf{X}_{s-1}\right)^{\mathbf{-1}}\mathbf{x}_s \\
&= -\sigma^2\mathbf{x}_t^{'}\left(\mathbf{X}_{t-1}^{'}\mathbf{X}_{t-1}\right)^{\mathbf{-1}}\mathbf{x}_s + \sigma^2\mathbf{x}_t^{'}\left(\mathbf{X}_{t-1}^{'}\mathbf{X}_{t-1}\right)^{\mathbf{-1}}\mathbf{X}_{s-1}^{'}\mathbf{X}_{s-1}^{'}\left(\mathbf{X}_{s-1}^{'}\mathbf{X}_{s-1}\right)^{\mathbf{-1}}\mathbf{x}_s = 0
\end{aligned} \tag{0.28}$$

The first two terms vanish. The vector $E\left(\boldsymbol{\varepsilon}_{t-1}\varepsilon_s\right)$ has zeros everywhere with the exception of row s, it follows that $\mathbf{X}_{t-1}^{'}\cdot E\left(\boldsymbol{\varepsilon}_{t-1}\varepsilon_s\right) = \sigma^2\mathbf{x}_s$. The diagonal matrix $E\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{s-1}^{'}\right)$ has zeros on rows s to t-1, therefore $\mathbf{X}_{t-1}^{'}\cdot E\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{s-1}^{'}\right) = \sigma^2\mathbf{X}_{s-1}^{'}$.

This result provides an intuitive rationale for the updating equation of $\mathbf{b}_t$. According to the updating equation, the previous prediction errors ν_s for $K+1 \le s \le t$, are incorporated in the current estimator $\mathbf{b}_t$. Since the prediction error ν_{t+1} is uncorrelated with the previous prediction errors ν_s, it seems reasonable that it represents the new information useful to update $\mathbf{b}_t$ to get $\mathbf{b}_{t+1}$.

The one-step-ahead prediction errors, have zero mean, a variance $\sigma^2 f_t$, and are uncorrelated.

The standardized prediction errors

$$\tilde{\nu}_t = \frac{\nu_t}{\sqrt{f_t}} = \frac{y_t - \mathbf{x}_t^{'}\mathbf{b}_{t-1}}{\sqrt{f_t}} \quad t = K+1,\ldots,T \tag{0.29}$$

are called *recursive residuals*. They are used to check the specification and stability of regression models.

The behavior of the recursive residuals is very similar to the one of the disturbances ε_t. If the disturbances ε_t are normally distributed, the recursive residuals are also normal, see equation (0.23). That is $\varepsilon_t \sim NID\left(0,\sigma^2\right)$ implies $\tilde{\nu}_t \sim NID\left(0,\sigma^2\right)$.

Sum of squares of the recursive residuals

The estimator of the variance can be updated:

$$SSE_t = SSE_{t-1} + \tilde{v}_t^2 \quad t = K+1,\cdots,T \tag{0.30}$$

where

$$SSE_t = \left(y_t - X_t b_t\right)' \left(y_t - X_t b_t\right) \quad t = k,\ldots,T\ . \tag{0.31}$$

A proof for (0.30) can be obtained by rewriting

$$\begin{gathered} SSE_t = SSE_{t-1} + \frac{v_t^2}{f_t} \\ \sum_{i=1}^{t} e_i^2 = \sum_{i=1}^{t-1} e_i^2 + \frac{v_t^2}{f_t} \end{gathered} \tag{0.32}$$

For (0.32) to be true we need that $e_t^2 = \frac{v_t^2}{f_t}$.

$$\begin{aligned} & e_t^2 = \left(y_t - x_t' b_t\right)' \left(y_t - x_t' b_t\right) = y_t^2 - 2x_t' b_t y_t + b_t' x_t x_t' b_t = y_t^2 - x_t' b_t y_t \\ & \frac{v_t^2}{f_t} = \frac{\left(y_t - x_t' b_{t-1}\right)' \left(y_t - x_t' b_{t-1}\right)}{f_t} = \frac{y_t^2 - x_t' b_{t-1} y_t}{f_t} \end{aligned} \tag{0.33}$$

$$\begin{aligned}
e_t^2 &= y_t^2 - x_t' b_t y_t \\
&= y_t^2 - x_t' \left[b_{t-1} + \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t \left(y_t - x_t' b_{t-1} \right) / \ f_t \right] y_t \\
&= \frac{1}{f_t} \left[y_t^2 \left(1 + x_t' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t \right) - x_t' b_{t-1} y_t \left(1 + x_t' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t \right) \right] \\
&\quad - \frac{1}{f_t} \left[x' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t y_t^2 - x' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t x_t' b_{t-1} y_t \right] \\
&= \frac{1}{f_t} \left[y_t^2 - x_t' b_{t-1} y_t \right] \qquad\qquad (0.34) \\
&\quad + \frac{1}{f_t} \left[x_t' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t y_t^2 - x_t' b_{t-1} y_t x_t' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t \right] \\
&\quad - \frac{1}{f_t} \left[x' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t y_t^2 - x_t' b_{t-1} y_t x' \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} x_t \right] \\
&= \frac{1}{f_t} \left[y_t^2 - x_t' b_{t-1} y_t \right] = \frac{v_t^2}{f_t}
\end{aligned}$$

The above result implies that

$$SSE_T = \sum_{t=1}^{T} e_t^2 = \sum_{t=K+1}^{T} \tilde{v}_t^2 \qquad (0.35)$$

Because of the properties of the recursive residuals it follows that

$$E\left(SSE\right) = \left(T - k\right)\sigma^2 \qquad (0.36)$$

and so

$$s^2 = \frac{SSE}{\left(T - K\right)} \qquad (0.37)$$

is an unbiased estimator of σ^2.

The use of (0.37) in conjunction with the updating equations (0.13), (0.14) and (0.30) allows the covariance matrix of $\mathbf{b}_t$ to be estimated as the recursion proceeds.

Is it clever to use the updating formula with $\mathbf{P}$?

The estimate for $\mathbf{b}_{t+1}$ and $\left(\mathbf{X}'_{t+1}\mathbf{X}_{t+1}\right)^{-1}$ will not change independently of sigma square. New forecast errors give rise to a new estimate for sigma square and all $\mathbf{P}$ have to be computed anew. Because the forecast errors are not affected, the new $\mathbf{P}$ cannot be used to improve the estimate of sigma in a new iteration. Because sigma square does not appear in them it is more sensible to use the updating equations (0.8) and (0.9) in the case where the model is not more complex than the regression (0.1), which consists of an observation equation only.

5. Test of parameter constancy

Section 1 to 4 provide the basis for test of parameter constancy (Chow Forecast Test, CUSUM and CUSUMQ).

See: Johnston DiNardo, page 112 and 119.

Exercises

Exercise 1:

Show the validity of the updating equation for $\left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1}$.

$$\left(\mathbf{X}_{t+1}^{'}\mathbf{X}_{t+1}\right)^{-1} = \left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1} - \frac{\left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1}\mathbf{x}_{t+1}\mathbf{x}_{t+1}^{'}\left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1}}{1+\mathbf{x}_{t+1}^{'}\left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1}\mathbf{x}_{t+1}}$$

Hint: Use thereby the Matrix Inversion Lemma (see exercice 3).

Exercise 2:

Show the validity of the updating equation for b.

$$\mathbf{b}_{t+1} = \mathbf{b}_t + \mathbf{P}_t\mathbf{x}_{t+1}\frac{\left(y_{t+1} - \mathbf{x}_{t+1}^{'}\mathbf{b}_t\right)}{F_t}$$

Hint: Insert the updating equation for $\left(\mathbf{X}_t^{'}\mathbf{X}_t\right)^{-1}$ into the LS estimator for $\mathbf{b}_{t+1}$ and use the decomposition $\mathbf{X}_{t+1}^{'}\mathbf{y}_{t+1} = \mathbf{X}_t^{'}\mathbf{y}_t + \mathbf{x}_{t+1}y_{t+1}$.

Exercise 3:

The Matrix Inversion Lemma states that the following equality holds:

$$\left(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{B}^{'}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{B}^{'}\mathbf{A}^{-1}\mathbf{B}\right)\mathbf{B}^{'}\mathbf{A}^{-1}$$

where A is $n \times n$, C is $m \times m$ and B is $n \times m$ is to be conformable. Convince yourself of the validity of the Lemma.