Least squares learning

George McCandless
BCRA and UCEMA

July 10, 2008

1   Introduction

An increasingly standard way of putting bounded rationality into a model is through the use of least squares learning for the expectation process. Instead of the rational expectations that we have been using, least squares learning implies that the agents inside the model have to use data produced by the model to make least squares forecasts of expectational variables and that these forecasts are updated each period as a new data point comes available. For the reader who is uncomfortable with rational expectaions as a working paradyme, least squares forecasting provides an attractive alternative. For those who are comfortable with rational expectations, under a range of conditions, equilibria with least squares forecasting converge to rational expectations equilibria.

In each period, all agents in the model use the same least squares forecasts for predicting expectational variables in the economy (the model). These forecasts are done inside the linear model of the economy and result in current values for the economy that are consistent with these projections. The data that results from this period provides another set of data points and this data is then used to update the coefficients of the least squares forecasting equations. The new coefficients are put into the model to solve for the next period.

It may be the cse that the modeller believes that the economy has undergone structural changes or changes in policies so that the parameters of the least squares forecasting problem are time dependent. In order to allow these parameters to adjust, it is possible to add a "forgetting" factor to the updating equation. This forgetting factor gives older data less weight in the estimation process so that the values of the parameters are weighted more by what has happened recently. Adding a forgetting factor tends to make the coefficients of the least squares forecasting equation vary with time and to be more responsive to recent outcomes of the economy.

Recursive least squares updating is a restricted form of the Kalman filter. A simple demonstration of this can be found in Branch and Evans [1] or in Sargent [6], Chapter 8, Appendix A.

2 Recursive least squares updating of coefficients

The first step is to present the recursive least squares updating procedure. This recursive least squares updating procedure is used along with a linear model that includes a expectations equations whose parameters are re -estimated each period from the updating procedure. In each period t, the linear model is run and is solved for the equilibrium values for the variables of the model for period t. The time t values of the variables are then used with the updating procedure to get new estimates for the paramters for the forecasting equations in the linear model. These new parameters are then put into the linear model and are used for forecasting in period $t + 1$. This process is repeated.

There are two versions of the recursive ordingary least squares presented in this section. The first version presented is equivalent to ordinary least squares and the later is equivalent to least squares with forgetting. In the version with forgetting, older data in the linear regression is given an exponentially smaller weight for determining current coefficients of the forecasting equations.

Consider a simple linear model of the form

$$y_t = x_t \varphi_t + \varepsilon_t,$$

where y_t is a vector of endogenous variables, x_t is the history up to moment t of the exogenous variables (that could include past values of y_t) that are being used for the estimation, φ_t is the estimate of the coefficients of the model using data up to time $t - 1$ and ε_t is a vector of error terms. Define $X_t = [x_t, x_{t-1}, ..., x_0]'$ and $Y_t = [y_t, y_{t-1}, ..., y_0]'$ as the available data on x_t and y_t in period t and where x_0 and y_0 are the data for the available initial period. Ordinary least squares of the estimate of the coefficients, φ_t, using the data available in period t is

$$\varphi_t = \left(X_t'X_t\right)^{-1} X_t'Y_t.$$

A way to find the least squares coefficients in a recursive manner is to write the matrices X_t and Y_t as

$$X_t = \left[\begin{array}{c} X_{t-1} \\ x_t \end{array} \right]$$

and

$$Y_t = \left[\begin{array}{c} Y_{t-1} \\ y_t \end{array} \right].$$

Then

$$X_t'X_t = \left[\begin{array}{cc} X_{t-1}' & x_t' \end{array} \right] \left[\begin{array}{c} X_{t-1} \\ x_t \end{array} \right] = X_{t-1}'X_{t-1} + x_t'x_t$$

and

$$X_t'Y_t = \left[\begin{array}{cc} X_{t-1}' & x_t' \end{array} \right] \left[\begin{array}{c} Y_{t-1} \\ y_t \end{array} \right] = X_{t-1}'Y_{t-1} + x_t'y_t$$

where both $x_t'x_t$ and $x_t'y_t$ are matrices of rank one. For our purposes, it is most interesting that the square matrix $x_t'x_t$ is of rank one. The inverse of the sum of an indentity matrix and a rank one matrix has a special, very simple solution.

Let $I + a'b$ be an $n \times n$ identity matrix and the product of two n dimension vectors to give an $n \times n$ matrix of rank one.[1] The inverse of the matrix $I + a'b$ can be written as

$$[I + a'b]^{-1} = I + ca'b$$

where

$$c = -\frac{1}{1 + ba'}.$$

Notice that $b'a$ is simply a scalar. To see that this formula for the inverse is correct, simply take

$$
\begin{aligned}
[I + ca'b][I + a'b] &= I + ca'b + a'b + ca'ba'b \\
&= I - \frac{1}{1 + ba'}a'b + a'b - \frac{a'b}{1 + ba'}a'b \\
&= I + a'b - \frac{1 + ba'}{1 + ba'}a'b \\
&= I + a'b - a'b = I.
\end{aligned}
$$

For a more general matrix (since $X'_{t-1}X_{t-1} \neq I$, in general), multiply $[I + ab']^{-1}$ by a non-singluar matrix B^{-1} to get

$$B^{-1}[I + a'b]^{-1} = [[I + a'b]B]^{-1} = [B + a'bB]^{-1}.$$

Using the formula for the inverse from the earlier discussion, we get

$$B^{-1}[I + a'b]^{-1} = B^{-1}[I + ca'b] = B^{-1} + cB^{-1}a'b.$$

Combining the above two results, we have that

$$[B + a'bB]^{-1} = B^{-1} + cB^{-1}a'b.$$

Define the vector as $f = bB$, and substitute this into the above expression to get

$$[B + a'f]^{-1} = B^{-1} + cB^{-1}a'fB^{-1}$$

where

$$c = -\frac{1}{1 + fB^{-1}a'}.$$

Using this method of finding an inverse, we can write

$$\left(X'_{t-1}X_{t-1} + x'_t x_t\right)^{-1} = \left(X'_{t-1}X_{t-1}\right)^{-1} + \mathbf{c}\left(X'_{t-1}X_{t-1}\right)^{-1} x'_t x_t \left(X'_{t-1}X_{t-1}\right)^{-1}, \tag{1}$$

with

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

---

[1] Any rank one $n \times n$ matrix can be expressed as the product of two n dimensional vectors.

Since

$$\begin{aligned} \varphi_t &= \left(X_t'X_t\right)^{-1} X_t'Y_t \\ &= \left(X_{t-1}'X_{t-1} + x_t'x_t\right)^{-1} \left(X_{t-1}'Y_{t-1} + x_t'y_t\right), \end{aligned}$$

we can replace the expression $\left(X_{t-1}'X_{t-1} + x_t'x_t\right)^{-1}$ by what it equals and after some algebra get

$$\varphi_t = \varphi_{t-1} + \frac{\left(X_{t-1}'X_{t-1}\right)^{-1} x_t'}{1 + x_t \left(X_{t-1}'X_{t-1}\right)^{-1} x_t'} \left(y_t - x_t\varphi_{t-1}\right).$$

Define $P_t = \left(X_t'X_t\right)^{-1}$, then this can be written as

$$\varphi_t = \varphi_{t-1} + \frac{P_{t-1}x_t'}{1 + x_tP_{t-1}x_t'} \left(y_t - x_t\varphi_{t-1}\right)$$

with the updating rule for P_t, from equation 1, of

$$P_t = \left[I - \frac{P_{t-1}x_t'}{1 + x_tP_{t-1}x_t'} x_t\right] P_{t-1}.$$

Beginning with some initial guesses for P_0 and φ_0, one can recursively calculate P_t and φ_t. These converge to the least square estimate.

The formulation of least squares updating derived here is quite robust, given that one works directly with the matrix P_t and the parameter vector φ_t and there is no need to take inverses of matrices.

3   Recursive least squares with forgetting

It may be the case that one wishes to give less weight to much older data, either because one believes that structural changes occur in a model or because one believes that individuals are somewhat myopic, and treat more recent observations as somehow more relevant than older observations. Least squares with forgetting is a version of the Kalman filter with constant "gain."

According to Lindoff [3], adding "forgetting" to recursive least squares estimation is simple. Choose a λ where $0 < \lambda < 1$ and adjust the updating rule to be

$$P_{t+1}^{-1} = \lambda P_t^{-1} + x_{t+1}'x_{t+1}.$$

Asymptotically, the recursive least squares model with forgetting is equivalent to a weighted least squares estimation of the form

$$\widehat{\varphi}_t = \left(\sum_{k=1}^t \lambda^{t-k} x_k'x_k\right)^{-1} \left(\sum_{k=1}^t \lambda^{t-k} x_k'y_k\right).$$

Notice that this form of a weighted least squares estimator gives every smaller weights, at the rate λ^j to data j periods in the past.

4

3.1 Linear version of recursive least squares with forgetting

To get a linear version of the recursive least squares with forgetting we follow the same methodology as above but define the matrices $X_t'X_t$ and $X_t'Y_t$ as

$$X_t'X_t = \begin{bmatrix} \lambda^{.5} X_{t-1}' & x_t' \end{bmatrix} \begin{bmatrix} \lambda^{.5} X_{t-1} \\ x_t \end{bmatrix} = \lambda X_{t-1}'X_{t-1} + x_t'x_t$$

and

$$X_t'Y_t = \begin{bmatrix} \lambda^{.5} X_{t-1}' & x_t' \end{bmatrix} \begin{bmatrix} \lambda^{.5} Y_{t-1} \\ y_t \end{bmatrix} = \lambda X_{t-1}'Y_{t-1} + x_t'y_t.$$

Notice that the forgetting factor λ is included in the definition. Using these definitions of $X_t'X_t$ and $X_t'Y_t$, the least squares estimator is

$$\begin{aligned} \varphi_t &= \left(X_t'X_t\right)^{-1} X_t'Y_t \\ &= \left(\lambda X_{t-1}'X_{t-1} + x_t'x_t\right)^{-1} \left(\lambda X_{t-1}'Y_{t-1} + x_t'y_t\right). \end{aligned}$$

Using the linear representation of the inverse from above, we have

$$\left(\lambda X_{t-1}'X_{t-1} + x_t'x_t\right)^{-1} = \frac{1}{\lambda}\left(X_{t-1}'X_{t-1}\right)^{-1} + \frac{\mathbf{c}}{\lambda}\left(X_{t-1}'X_{t-1}\right)^{-1} x_t'x_t \left(X_{t-1}'X_{t-1}\right)^{-1},$$

with

$$\mathbf{c} = -\frac{1}{\lambda + x_t \left(X_{t-1}'X_{t-1}\right)^{-1} x_t'}.$$

From a derivation similar to the one given above, we have that

$$\varphi_t = \varphi_{t-1} + \frac{\left(X_{t-1}'X_{t-1}\right)^{-1}}{\lambda + x_t \left(X_{t-1}'X_{t-1}\right)^{-1} x_t'} x_t' \left[y_t - x_t\varphi_{t-1}\right]$$

and the updating equation is

$$\begin{aligned} \left(X_t'X_t\right)^{-1} &= \left(\lambda X_{t-1}'X_{t-1} + x_t'x_t\right)^{-1} \\ &= \frac{1}{\lambda}\left(X_{t-1}'X_{t-1}\right)^{-1} + \mathbf{c}\frac{1}{\lambda}\left(X_{t-1}'X_{t-1}\right)^{-1} x_t'x_t \frac{1}{\lambda}\left(X_{t-1}'X_{t-1}\right)^{-1}. \end{aligned}$$

Defining $P_t = \left(X_t'X_t\right)^{-1}$, we can write the model as

$$\varphi_t = \varphi_{t-1} + \frac{P_{t-1}x_t'}{\lambda + x_t P_{t-1}x_t'}\left[y_t - x_t\varphi_{t-1}\right]$$

and the updating equation as

$$P_t = \frac{1}{\lambda}\left[I + \frac{P_{t-1}x_t}{\lambda + x_t'P_{t-1}x_t}x_t'\right]P_{t-1}.$$

As Branch and Evans [1] and Sargent [6] point out, least squares learning is a restricted form of the Kalman filter where the constant Kalman "gain" is equal to $1 - \lambda$.

5

4 Applying least squares learning to the Hansen model

Consider the most simple version of the Hansen model (for a development of this model, see McCandless [5], Chapter 6). The log-linear form of that model is

$$
\begin{aligned}
0 &= \widetilde{C}_t - E_t\widetilde{C}_{t+1} + \beta\bar{r}E_t\widetilde{r}_{t+1} \\
0 &= \widetilde{Y}_t - \frac{\widetilde{H}_t}{1 - \overline{H}} - \widetilde{C}_t \\
0 &= \overline{Y}\widetilde{Y}_t - \overline{C}\widetilde{C}_t + \overline{K}\left[(1 - \delta)\widetilde{K}_t - \widetilde{K}_{t+1}\right] \\
0 &= \widetilde{\lambda}_t + \theta\widetilde{K}_t + (1 + \theta)\widetilde{H}_t - \widetilde{Y}_t \\
0 &= \widetilde{Y}_t - \widetilde{K}_t - \widetilde{r}_t
\end{aligned}
$$

with a stochastic process for technology of

$$
\widetilde{\lambda}_t = \gamma\widetilde{\lambda}_{t-1} + \widetilde{\varepsilon}_t.
$$

We assume that the expected variables $E_t\widetilde{C}_{t+1}$ and $E_t\widetilde{r}_{t+1}$ are found by projection from a regression model of the form

$$
\left[\begin{array}{c} E_t\widetilde{C}_{t+1} \\ E_t\widetilde{r}_{t+1} \end{array}\right] = \left[\begin{array}{cc} \varphi_{11}^1 & \varphi_{12}^1 \\ \varphi_{21}^1 & \varphi_{22}^1 \end{array}\right]\left[\begin{array}{c} \widetilde{K}_t \\ \widetilde{Y}_t \end{array}\right],
$$

where the coefficients of the regression are determined using data up to period $t - 1$. The reason we use a regress with $\widetilde{K}_t$ and $\widetilde{Y}_t$ instead of using a VAR with $\widetilde{C}_t$ as the second independent variable is because the first line of the model has both $\widetilde{C}_t$ and $E_t\widetilde{C}_{t+1}$ in it with coefficients equal to one, so the A matrix can become less than full rank, for example when

$$
\left[\begin{array}{cc} \varphi_{11}^1 & \varphi_{12}^1 \\ \varphi_{21}^1 & \varphi_{22}^1 \end{array}\right] = \left[\begin{array}{cc} 2 & q \\ 1/\beta\bar{r} & q/\beta\bar{r} \end{array}\right],
$$

for any q. However, a VAR forecasting model is frequently possible and would be preferred if multiple step-ahead forecasts are needed.

Using the recursive least squares method described above, the parameters of the model are updated using

$$
\varphi_t = \varphi_{t-1} + \frac{P_{t-1}x_t'}{\lambda + x_tP_{t-1}x_t'}\left(y_t - x_t\varphi_{t-1}\right)
$$

and the updating equation for P_t is

$$
P_t = \frac{1}{\lambda}\left[I + \frac{P_{t-1}x_t'}{\lambda + x_tP_{t-1}x_t'}x_t\right]P_{t-1}.
$$

In this last equation, λ is the foregetting factor. $\lambda = 1$ for a ordinary least squares forecast and can take values less than one (normally between .9 and .999) when the model includes foregetting.

We write out the Hansen model in a state space format. The variables to be determined at date t are

$$x_t = \begin{bmatrix} K_{t+1} \\ H_t \\ Y_t \\ C_t \\ r_t \\ E_t \widetilde{C}_{t+1} \\ E_t \widetilde{r}_{t+1} \\ \widetilde{\lambda}_t \end{bmatrix}.$$

The model is comprised of two parts. The state space version of the log-linear model, written as

$$A_t \left(\varphi_{t-1} \right) x_t = B_t \left(\varphi_{t-1} \right) x_{t-1} + C \varepsilon_t,$$

and the updating equation for the parameters of the least squares learning VAR, which can be given as

$$\begin{bmatrix} \varphi_t \\ P_t \end{bmatrix} = G \left(\begin{bmatrix} \varphi_{t-1} \\ P_{t-1} \end{bmatrix}, x_t \right).$$

The matrix A is written as $A_t \left(\varphi_{t-1} \right)$ to remind us that it has parameters in the least squares updating equation that were found using the data up to period $t - 1$. If the matrix $A_t \left(\varphi_{t-1} \right)$ is invertable, then the state space model can be solved directly in each period as

$$x_t = \left[A_t \left(\varphi_{t-1} \right) \right]^{-1} B_t \left(\varphi_{t-1} \right) x_{t-1} + \left[A_t \left(\varphi_{t-1} \right) \right]^{-1} C \varepsilon_t.$$

The matrices A_t, B_t, and C_t for the above version of the Hansen economy are

$$A_t = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & -1 & \beta \overline{r} & 0 \\ 0 & -\frac{1}{1-\overline{H}} & 1 & -1 & 0 & 0 & 0 & 0 \\ -\overline{K} & 0 & \overline{Y} & -\overline{C} & 0 & 0 & 0 & 0 \\ 0 & 1-\theta & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ -\varphi_{11}^1(t-1) & 0 & -\varphi_{12}^1(t-1) & 0 & 0 & 1 & 0 & 0 \\ -\varphi_{21}^1(t-1) & 0 & -\varphi_{22}^1(t-1) & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -(1-\delta)\overline{K} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\theta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma \end{bmatrix},$$

and

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}'.$$

Notice how the parameters of the regression for estimating $E_t\widetilde{C}_{t+1}$ and $E_t\widetilde{r}_{t+1}$, in the form,

$$\begin{matrix} -\varphi^1_{11}(t-1) & 0 & -\varphi^1_{12}(t-1) \\ -\varphi^1_{21}(t-1) & 0 & -\varphi^1_{22}(t-1) \end{matrix},$$

are included directly matrix A of the state space version of the model.

A time path for the economy is found by solving the economy, beginning with some initial guesses (priors) for φ_0 and P_0 and using the data that comes from the model to update φ_0 and P_0 and get new values for φ_1 and P_1. The parameter values φ_1 and then put into matrix A and the model is solved again. If $\lambda = 1$, the matrix P_t becames small quite quickly and the parameter values, φ_t, tend to converge slowly. Nor are the parameters is very responsive to a structural change (or policy change), since all data, new and old, are given the same weight and, if there is substantial data over the old policy regime, it may take a long time for the effects of the new regime to be observed in the parameters.. Running example economies, one can find that the convergence is so slow as to be almost useless. For example, using the above economy with the parameters from McCandless , Chapter 6, the rational expectations values for φ are

$$\varphi = \begin{bmatrix} .5130 & .2614 \\ -1.007 & .9662 \end{bmatrix}.$$

If one starts the recursive least squares process with both φ and P equal to 2×2 identity matrices, φ^1_{11} is still not very close after $200,000$ periods. Figure 1 shows the time path of the four paramters in this simulation. At the end of the $200,000$ periods φ is equal to

$$\varphi = \begin{bmatrix} 0.6803 & 0.2750 \\ -1.0058 & 0.9597 \end{bmatrix}$$

and the updating matrix P is equal to

$$P = \begin{bmatrix} .00009069 & -.00005938 \\ -.00005938 & .00005983 \end{bmatrix}$$

so each new data point has very little influence on the parameter values.

With forgetting, the coefficients of φ_t do not usually converge exactly, but are distributed over a range that contains the fixed point of the version without
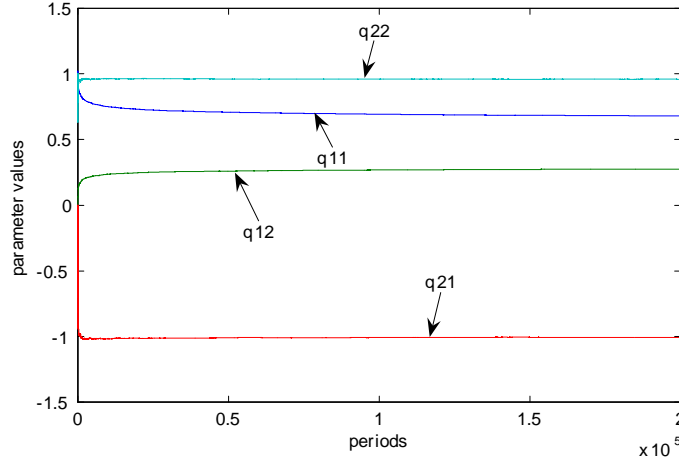
Figure 1: Parameter values over 200,000 periods: $\lambda = 1$

forgetting. If $\lambda < 1$, P_t stays relatively large and while φ_t can show substantial variation, it tends to converge faster into a range that includes the rational expectations values. Figure 2 shows the time path for the parameters for the same economy as above with the same starting points but where $\lambda = .999$. After about 20,000 periods, the parameters are close to their rational expectations values. Even at the end of the period, the updating matrix P is still relatively large with

$$P = \left[\begin{array}{cc} 0.0098 & -0.0084 \\ -0.0084 & 0.0102 \end{array} \right],$$

so that new data has much more impact on updating the parameters and in the case where $\lambda = 1$.

5 Forecasting inside a linear rational expectations model

If the economy that one is studying is a linear rational expectations model, such as the basic Hansen model, then one can use the linear policy functions found from solving the economy to find a set of linear forecasting functions that use the data generated in any period t to get the rational expectations forecast for the values of variables in period $t + 1$. A single agent or econometrician living in this economy and producing a least squares forecast on some set of variables will end up with this same linear equation as the result of their estimation process. We first show how to find a linear forecasting equation given the policy functions and then compare the one that can be calculated from the
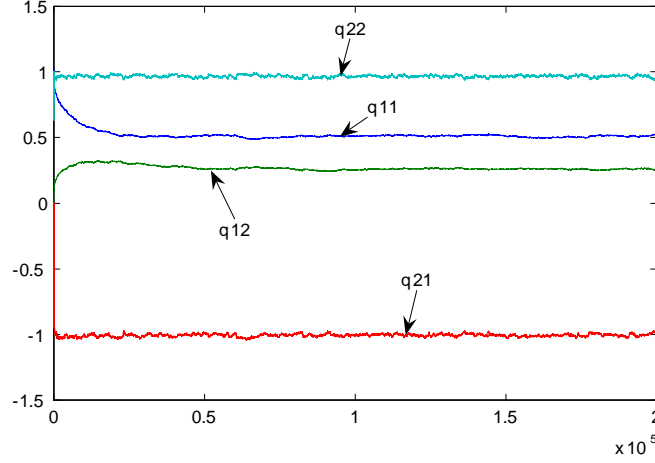
9

Figure 2: Parameter values over 200,000 periods: $\lambda = .999$

policy functions for the basic Hansen model to a linear regression on the same variables from simulations of the same Hansen model.

To see how this works, assume that we have the policy functions (plans) for the rational expectations model. From those, we can write the policy functions for the two variables we want to forecast in the form

$$
\left[\begin{array}{c} x_t^d \\ x_t^e \end{array} \right] = \left[\begin{array}{cc} d^k & d^\lambda \\ e^k & e^\lambda \end{array} \right] \left[\begin{array}{c} k_t \\ \lambda_t \end{array} \right]
$$

and the equation for updating the state variable as

$$
k_{t+1} = \left[\begin{array}{cc} k^k & k^\lambda \end{array} \right] \left[\begin{array}{c} k_t \\ \lambda_t \end{array} \right].
$$

Since there are only two state variables in the Hansen model we are considering here, k_t and λ_t, only two variables[2] from the model are needed to get the rational expectations forecast. Here we illustrate a case where the econometrician is using variables d and e to forecast the variables a and b. If one were doing a VAR, then variables $a = d$ and $b = e$. We want to find the regression model that an agent inside the rational expectations model would finding for forecasting

$$
\left[\begin{array}{c} E_t x_{t+1}^a \\ E_t x_{t+1}^b \end{array} \right] = \left[\begin{array}{cc} c^{a1} & c^{a2} \\ c^{b1} & c^{b2} \end{array} \right] \left[\begin{array}{c} x_t^d \\ x_t^e \end{array} \right].
$$

---

[2] This is true if the two variables are not perfectly correlated or that $\left[\begin{array}{cc} a^1 & b^1 \\ a^2 & b^2 \end{array} \right]$ is invertible, which is to say the same thing.

10

Since we know the policy functions, we know that the rational expectations forecast of these variables is

$$
\begin{bmatrix} E_t x_{t+1}^a \\ E_t x_{t+1}^b \end{bmatrix} = \begin{bmatrix} a^k & a^\lambda \\ b^k & b^\lambda \end{bmatrix} \begin{bmatrix} k_{t+1} \\ \gamma\lambda_t \end{bmatrix},
$$

where since $E_t\varepsilon_{t+1} = 0$, $E_t\lambda_{t+1} = \gamma\lambda_t$. The rational expectations forecasts for $E_t k_{t+1}$ and $E_t\lambda_{t+1}$ are

$$
\begin{bmatrix} k_{t+1} \\ \gamma\lambda_t \end{bmatrix} = \begin{bmatrix} k^k & k^\lambda \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} k_t \\ \lambda_t \end{bmatrix}.
$$

Therefore, the expectations for x_{t+1}^a and x_{t+1}^b can be written in terms of the state variables at date t as

$$
\begin{bmatrix} \begin{bmatrix} E_t x_{t+1}^a \\ E_t x_{t+1}^b \end{bmatrix} \end{bmatrix} = \begin{bmatrix} a^k & a^\lambda \\ b^k & b^\lambda \end{bmatrix} \begin{bmatrix} k^k & k^\lambda \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} k_t \\ \lambda_t \end{bmatrix}
$$

Since we have assumed that $\begin{bmatrix} d^k & d^\lambda \\ e^k & e^\lambda \end{bmatrix}$ is invertible, then one can write

$$
\begin{bmatrix} k_t \\ \lambda_t \end{bmatrix} = \begin{bmatrix} d^k & d^\lambda \\ e^k & e^\lambda \end{bmatrix}^{-1} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}.
$$

Using this, the regression equation can be written as

$$
\begin{bmatrix} E_t x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = \begin{bmatrix} a^k & a^\lambda \\ b^k & b^\lambda \end{bmatrix} \begin{bmatrix} k^k & k^\lambda \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} d^k & d^\lambda \\ e^k & e^\lambda \end{bmatrix}^{-1} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix},
$$

so that the regression equation under rational expectation should be

$$
\begin{bmatrix} c^{11} & c^{12} \\ c^{21} & c^{22} \end{bmatrix} = \begin{bmatrix} a^k & a^\lambda \\ b^k & b^\lambda \end{bmatrix} \begin{bmatrix} k^k & k^\lambda \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} d^k & d^\lambda \\ e^k & e^\lambda \end{bmatrix}^{-1}.
$$

From Chapter 6, the rational expectations version of Hansen's basic model gives the policy functions (or plans),

$$
\begin{aligned}
P &= 0.9537, \\
Q &= 0.1132, \\
R &= \begin{bmatrix} 0.2045 \\ 0.5691 \\ -0.2430 \\ -0.7955 \end{bmatrix}, \\
S &= \begin{bmatrix} 1.4523 \\ 0.3920 \\ 0.7067 \\ 1.4523 \end{bmatrix}.
\end{aligned}
$$

11

For the Hansen economy, the only variables that are expectational variables needed are the time $t + 1$ consumption and interest rate. Using these as the variables we wish to estimate and time t capital and output as the exogenous variable, we have

$$
\begin{bmatrix} c^{11} & c^{12} \\ c^{21} & c^{22} \end{bmatrix} = \begin{bmatrix} .5691 & .3929 \\ -.79955 & 1.4523 \end{bmatrix} \begin{bmatrix} .9537 & .1132 \\ 0 & .95 \end{bmatrix} \begin{bmatrix} .9537 & .1132 \\ .2045 & 1.4523 \end{bmatrix}^{-1},
$$

or that

$$
\begin{bmatrix} c^{11} & c^{12} \\ c^{21} & c^{22} \end{bmatrix} = \begin{bmatrix} 0.513\,0 & 0.261\,4 \\ -1.007 & 0.966\,2 \end{bmatrix}.
$$

We carried out simulations of 4000 periods using the basic Hansen model with the parameters in McCandless [5], Chapter 6, and an normal error term with a standard error of .1. After throwing away the first 2000 results of the simulation to get rid of initialization problems, running a least squares regression on the remaining 2000 observations gives the estimates of the above regression as

$$
\begin{bmatrix} c^{11} & c^{12} \\ c^{21} & c^{22} \end{bmatrix} = \begin{bmatrix} 0.5127 & 0.2605 \\ -1.0045 & 0.9651 \end{bmatrix}.
$$

The equation that we found above is the one a single individual would encounter if s/he were estimating a least squares forecasting model using the capital stock and output as the exogenous variables to estimate consumption and interest rates. S/he could just as easily have estimated a VAR using one lag of consumption and interest as the exogenous veriables in the estimation and then use current consumption and interest rate to forecast the economy's (rationally) expected values for time $t + 1$ consumption and interest rates.

Exercise 1 *Find the coefficients in the VAR model that a single agent would find for forecasting consumption and interest rates using the calculations described above. Then run simulations of the basic Hansen model for 4000 periods and use the last 2000 observations of that data in a least squares estimation for finding the estimated coefficients of the VAR.*

A agent living in a rational expectations economy could use the data generated each period to recursively update his or her forecasting equation. Since recursive least squares estimation techniques converge to the least squares estimate generated over the same data set, the recursive forecasting model would converge to the model found above. This is one sense in which least squares learning can converge to rational expectations.

A point needs to be mentioned with respect to the least squares or recursive least squares estimates generated by an agent living in a rational expections economy. Probably the most important special characteristic of the example given above is that it used sufficient exogenous variables in the estimation to exactly cover the number of state variables in the economy. Two variables, d and e, were used as exogenous variables to be able to exactly determine the values of the state variables k_t and λ_t that occurred in period t. Since the values

of the full set of state variables could be determined, putting these values into the appropriate policy functions could give the values of all other variables and, especially necessary for the forecasts, the expected values of the state variables in period $t + 1$. In other words, the set of exogenous variables that were used exactly *span* the space of the state variables.

It is of interest to know what happens when the exogenous variables do not span the space of state variables. This question is relevant because the number of state variables included in any model will normally be less (and possibly much less) than the number of state variables in the world that is being modelled. It is not unreasonable to assume that the econometrician is normally working with fewer exogenous variables than there are states in the world.

To be able to get the parameters that are those of the rational expectations model, one would need to be able to solve a problem such as the following. Suppose that we have data on three endogenous varables, v_t^i, $i = 1, 2, 3$, from an economy with four state variables, s_t^j, $j = 1$ to 4 and we want to calculate what the VAR for these variables should look like. We want to find the matrix C for the forecasting problem

$$
\begin{bmatrix} E_t v_{t+1}^1 \\ E_t v_{t+1}^2 \\ E_t v_{t+1}^3 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \end{bmatrix}.
$$

From the policy functions, we know that

$$
\begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} s_t^1 \\ s_t^2 \\ s_t^3 \\ s_t^4 \end{bmatrix},
$$

and that there exists a matrix for updating the state variables of the form

$$
\begin{bmatrix} s_t^1 \\ s_t^2 \\ s_t^3 \\ s_t^4 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \begin{bmatrix} s_{t-1}^1 \\ s_{t-1}^2 \\ s_{t-1}^3 \\ s_{t-1}^4 \end{bmatrix},
$$

so we can write

$$
\begin{bmatrix} E_t v_{t+1}^1 \\ E_t v_{t+1}^2 \\ E_t v_{t+1}^3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \begin{bmatrix} s_t^1 \\ s_t^2 \\ s_t^3 \\ s_t^4 \end{bmatrix}.
$$

Unfortunately, the final step would involve inverting

$$
\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}
$$

13

to be able to find an expression for the states, $\begin{bmatrix} s_t^1 & s_t^2 & s_t^3 & s_t^4 \end{bmatrix}'$, as a function of the varables, $\begin{bmatrix} v_t^1 & v_t^2 & v_t^3 \end{bmatrix}'$. This inverse does not exist.

Assume that we add an error term of the form

$$
\begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix} \begin{bmatrix} s_t^1 \\ s_t^2 \\ s_t^3 \\ s_t^4 \end{bmatrix}
$$

so that the augmented matrix is invertible. Then we can find

$$
\begin{bmatrix} E_t v_{t+1}^1 \\ E_t v_{t+1}^2 \\ E_t v_{t+1}^3 \end{bmatrix}
$$

$$
= \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}^{-1} \begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \\ \varepsilon_t \end{bmatrix} .
$$

Define

$$
\begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \end{bmatrix}
$$

$$
= \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}^{-1}
$$

and one gets

$$
\begin{bmatrix} E_t v_{t+1}^1 \\ E_t v_{t+1}^2 \\ E_t v_{t+1}^3 \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \end{bmatrix} + \begin{bmatrix} f_{14} \\ f_{24} \\ f_{34} \end{bmatrix} \begin{bmatrix} \varepsilon_t \end{bmatrix}
$$

where the d_i's are chosen so as to minimize the sum of squared errors,

$$
\begin{bmatrix} f_{14} & f_{24} & f_{34} \end{bmatrix} \begin{bmatrix} f_{14} \\ f_{24} \\ f_{34} \end{bmatrix} var\,(\varepsilon_t) .
$$

The variance of the error term, $var\,(\varepsilon_t)$, and the vector $\begin{bmatrix} f_{14} & f_{24} & f_{34} \end{bmatrix}'$ will normally be non-zero and will depend not only on the policy matrices but also on the variances of the state variables. In an economy of this type, the least squares forecases will assume that the error term, ε_t, is zero and the least squaresd forecasts that come from this equation will differ from the rational expectations forecasts by the size of that error term times $\begin{bmatrix} f_{14} & f_{24} & f_{34} \end{bmatrix}'$.

14

References

[1] Branch, W., and G. Evans (2005) "A simple recursive forecasting model," *Economic Letters*, 91, p. 158-66.

[2] Evans, G., and S. Honkapohja (2008) "Expectations, Learning and Monetary Policy: An Overview of Recent Research," Working paper, Cambridge University.

[3] Lindoff, B. (1997) "On the Optimal Choice of the Forgetting Factor in the Recursive Least Squares Estimator," Working paper, Lund Institute of Technology.

[4] Marcet, A. and T. J. Sargent (1989) "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models," *Journal of Economic Theory*, 48 (2), August, p. 337-368.

[5] McCandless, G. (2008) *The ABCs of RBCs: An Introduction to Dynamic Macroeconomic Models*, Harvard University Press, Cambridge.

[6] Sargent, T. J., (1993) *The Conquest of American Inflation*, Princeton University Press, Princeton.