Time Series Models
Supervision 4

Samuel Lee

Question 1

We have a VAR(1) in two variables:

$$\begin{pmatrix} a_{11}^0 & a_{12}^0 \\ a_{21}^0 & a_{22}^0 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} a_{11}^1 & a_{12}^1 \\ a_{21}^1 & a_{22}^1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

where $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N(0, \Sigma)$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$. There are 8 parameters in the coefficients, 2 in the intercept, and 3 in the error covariance matrix, which means there are 13 parameters to estimate. This is 4 more compared to the parameters from the reduced form, which characterise the likelihood function completely for a given dataset. Therefore we need 4 restrictions to identify the model.

(i.) If we impose $a_{21}^0 = a_{12}^0 = 0$ and $a_{11}^0 = a_{22}^0 = 1$, the reduced form is the same as the structural form (the coefficient matrix on the left-hand side is the identity matrix). We can just estimate the parameters by OLS.

(ii.) If we impose $a_{11}^0 = a_{22}^0 = 1$, $a_{21}^0 = 0$, and $\sigma_{12} = 0$, the coefficient matrix on the left-hand side is an upper triangular matrix with 1 on the diagonals, and the innovation terms are orthogonal to one another. We can always re-arrange the equations to make the coefficient matrix on the left-hand side a lower triangular matrix instead, and we assume that this has been done if only to adhere to the standard setup.

The VAR can be denoted as

$$B_0 \mathbf{x}_t = \boldsymbol{\gamma} + B_1 \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

and assuming B_0 is invertible, we can manipulate the above to get

$$\mathbf{x}_t = B_0^{-1} \boldsymbol{\gamma} + B_0^{-1} B_1 \mathbf{x}_{t-1} + B_0^{-1} \boldsymbol{\varepsilon}_t = A_0 + A_1 \mathbf{x}_{t-1} + \mathbf{u}_t$$

When we estimate the reduced form of the VAR, we also get an estimate of the variance-covariance matrix Ω of $\mathbf{u}_t$. Ω would be positive definite and symmetric, which means there is a unique triangular factorisation $\Omega = ADA'$ where A is a lower triangular matrix with 1 along the diagonal and D is a diagonal matrix. But with $\mathbf{u}_t = B_0^{-1} \varepsilon_t$, we have

$$\Omega = \mathrm{E}[\mathbf{u}_t \mathbf{u}_t'] = B_0^{-1} \mathrm{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] (B_0^{-1})' = B_0^{-1} D (B_0^{-1})'$$

and by uniqueness we have that $A = B_0^{-1}$. The model would be just-identified and we can infer the values of B_0, B_1, and γ by estimating Ω, applying the triangular factorisation (otherwise known as the Cholesky decomposition), and premultiplying the estimated coefficients by B_0.

The dynamic relationship between the variables in the first case is that there is only a lagged effect of one variable on another. One example could be chickens and eggs: the number of chickens today affects how many eggs are laid in the future, while the number of eggs laid today affects how many chickens there are in the future.

In the second case, x_2 is not affected by present values of x_1, but x_1 is affected by present values of x_2. An example could be monetary policy and inflation: inflation might respond to monetary policy after a lag, but monetary policy is set with reference to the latest information about inflation.

Question 2

We have a VAR in two variables:

$$A_0 x_t = A_1 x_{t-1} + \varepsilon_t$$

Assuming A_0 is invertible,

$$x_t = A_0^{-1} A_1 x_{t-1} + A_0^{-1} \varepsilon_t = B x_{t-1} + u_t$$

where $B = A_0^{-1} A_1$ and $u_t = A_0^{-1} \varepsilon_t$, which is what is estimated when estimating the reduced form. The problem in trying to estimate the structural form is that estimating the reduced form gives us estimates of B (4 parameters) and $\mathrm{E}[u_t u_t']$ (3 parameters), giving us 7 parameter estimates. But there are 11 parameters in the structural form, and we are left with 4 free parameters. Normalising the diagonals in A_0 to 1 gives leaves us with 2 extra parameters, and assuming the diagonals of Σ are 0 leaves us with one extra parameter. Therefore we need one more restriction to identify the structural form.

A common restriction is to set A_0 to be lower triangular, in addition to the previously imposed value of 1 on the diagonals. We can then recover an estimate of A_0^{-1} from the estimate of $\mathrm{E}[u_t u_t']$, which allows us to recover the estimate of A_0 and A_1 by inverting the estimate of A_0^{-1} and multiplying it by the estimate of B respectively. To see this, we note that if A_0 is lower triangular with 1 on the diagonals, then so is A_0^{-1} (this can be seen by writing any valid A_0 and solving $A_0 P = I$). Then u_t takes the following structure:

$$\begin{aligned} u_{1t} &= \varepsilon_{1t} \\ u_{2t} &= a\varepsilon_{1t} + \varepsilon_{2t} \end{aligned}$$

We then have

$$\begin{aligned} \mathrm{Var}[u_{1t}] &= \sigma_1^2 \\ \mathrm{Var}[u_{2t}] &= a^2 \sigma_1^2 + \sigma_2^2 \\ \mathrm{Cov}[u_{1t}, u_{2t}] &= a\sigma_1^2 \end{aligned}$$

which gives us 3 unknowns in 3 equations. We can then solve recursively for A_0^{-1} by noting $\sigma_1^2 = \text{Var}[u_{1t}]$ and $a = \frac{\text{Cov}[u_{1t},u_{2t}]}{\sigma_1^2}$, and our estimate of A_0^{-1} comes from replacing the population moments with their sample counterparts.

The impulse response functions in this case are not invariant to the ordering of variables for the same reason that they are not invariant to our choice between making A_0 a lower triangular or upper triangular matrix. Imposing a lower triangular matrix on A_0 implies that the realisation of $\varepsilon_{1t} = \delta$ has a contemporaneous effect on x_{2t}, whereas the realisation of $\varepsilon_{2t} = \delta$ only has a lagged effect on x_{1t}. Imposing an upper triangular matrix on A_0 (or equivalently, reordering x_t in a 'flipped' manner) implies the opposite. This affects the impulse response function; a trivial example is that the impulse response function will chart a different contemporaneous response to a given realisation of ε_t.

Mathematically, this can be shown through the following

$$\begin{aligned}
x_t &= A_0^{-1}A_1x_{t-1} + A_0^{-1}\varepsilon_t \\
(I - A_0^{-1}A_1L)x_t &= A_0^{-1}\varepsilon_t \\
x_t &= (I - A_0^{-1}A_1L)^{-1}A_0^{-1}\varepsilon_t \\
&= \left(I + A_0^{-1}A_1L + (A_0^{-1}A_1L)^2 + \ldots\right) A_0^{-1}\varepsilon_t \\
&= \sum_{i=0}^{\infty}(A_0^{-1}A_1)^i A_0^{-1}L^i\varepsilon_t
\end{aligned}$$

which means that the impulse response function $I(h, \delta, \Omega_{t-1})$ is equal to

$$\begin{aligned}
I(h, \delta, \Omega_{t-1}) &= \text{E}[x_{t+h}|\varepsilon_t = \delta, \varepsilon_{t+1} = 0, \ldots, \varepsilon_{t+h} = 0, \Omega_{t-1}] - \text{E}[x_{t+h}|\varepsilon_t = 0, \ldots, \varepsilon_{t+h} = 0, \Omega_{t-1}] \\
&= \sum_{i=h+1}^{\infty}(A_0^{-1}A_1)^i A_0^{-1}L^i\varepsilon_{t+h} + (A_0^{-1}A_1)^h A_0^{-1}\delta - \sum_{i=h+1}^{\infty}(A_0^{-1}A_1)^i A_0^{-1}L^i\varepsilon_{t+h} \\
&= (A_0^{-1}A_1)^h A_0^{-1}\delta
\end{aligned}$$

and we can see that this can lead to different outputs depending on whether A_0 is a upper or lower triangular matrix. For example, if A_0 is lower triangular, then a shock of $\delta = \begin{pmatrix} 0 \\ c \end{pmatrix}$ yields $I(0, \delta, \Omega_{t-1}) = \begin{pmatrix} 0 \\ c' \end{pmatrix}$ since $A_0^{-1}\delta = 0$. This does not hold if A_0 is upper triangular.

Question 3

(a)

With $A = \begin{pmatrix} 0.5 & -1 \\ -0.25 & 0.5 \end{pmatrix}$ and $L = \begin{pmatrix} 1 & -2 \\ 0.5 & 1 \end{pmatrix}$, we have $L^{-1} = \begin{pmatrix} 0.5 & 1 \\ -0.25 & 0.5 \end{pmatrix}$, and

$$LAL^{-1} = \begin{pmatrix} 1 & -2 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & -1 \\ -0.25 & 0.5 \end{pmatrix} \begin{pmatrix} 0.5 & 1 \\ -0.25 & 0.5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \Lambda$$

(b)

If $y_t = Lx_t$, we have

$$\begin{aligned} x_t &= Ax_{t-1} + \varepsilon_t \\ Lx_t &= LAx_{t-1} + L\varepsilon_t \\ y_t &= LAL^{-1}Lx_{t-1} + L\varepsilon_t \\ &= \Lambda y_{t-1} + L\varepsilon_t \end{aligned}$$

which is what is needed. From the above, we have that $y_{1t} = x_{1t} - 2x_{2t}$ is an $I(1)$ variable, while $y_{2t} = 0.5x_{1t} + x_{2t}$ is stationary. A linear combination of an $I(1)$ variable and a stationary variable is still $I(1)$, so we can observe that $\frac{y_{1t}+2y_{2t}}{2} = x_{1t}$ and $-\frac{y_{1t}-2y_{2t}}{4} = x_{2t}$ are both $I(1)$.

(c)

We have

$$\Pi = A - I = \begin{pmatrix} -0.5 & -1 \\ -0.25 & -0.5 \end{pmatrix} \implies \det(\Pi) = (-0.5)(-0.5) - (-1)(-0.25) = 0$$

and also that

$$\alpha\beta' = \begin{pmatrix} -0.5 \\ -0.25 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} = \begin{pmatrix} -0.5 & -1 \\ -0.25 & -0.5 \end{pmatrix} = \Pi$$

(d)

We know from before that $y_t = \Lambda y_{t-1} + L\varepsilon_t$. Since $y_{2t} = 0.5x_{1t} + x_{2t}$ and $L\varepsilon_t = 0.5\varepsilon_{1t} + \varepsilon_{2t}$, and $\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, this means that $\begin{pmatrix} 0.5 & 1 \end{pmatrix} x_t = \begin{pmatrix} 0.5 & 1 \end{pmatrix} \varepsilon_t$, or equivalently, $\beta' x_t = \beta' \varepsilon_t$.

(e)

The above implies x_{1t} and x_{2t} are co-integrated with the co-integration vector β.

Question 4

(a)

With only one lag, the Granger Causality tests basically amount to single tests of whether the coefficient on lagged values of ret_t are significant in the equation for div_t, and vice versa. The standard t-tests suffice for this: the t-statistic of $\hat{\alpha}_2$ is $\frac{0.0274}{0.013} \approx 2.108$, and that of $\hat{\alpha}_6$ is $\frac{0.6224}{0.599} \approx 1.039$. The former is statistically significant at the 5% level, while the latter is not statistically significant at the 10% level. The hypothesis that div_t does not Granger-cause ret_t is rejected at the 5% level, and the hypothesis that ret_t does not Granger-cause div_t is not rejected at the 10% level.

(b)

We have that

$$(1-\beta_4 L)\,div_t = \beta_3 + v_t^2$$

From the equation for ret_t, we have

$$\begin{aligned}
ret_t &= \beta_1 + \beta_2 div_{t-1} + v_t^1 \\
(1-\beta_4 L) ret_t &= \beta_1(1-\beta_4) + \beta_2(1-\beta_4 L) div_{t-1} + (1-\beta_4 L) v_t^1 \\
&= \beta_1(1-\beta_4) + \beta_2(\beta_3 + v_t^2) + (1-\beta_4 L) v_t^1 \\
&= \beta_1(1-\beta_4) + \beta_2\beta_3 + \beta_2 v_t^2 + (1-\beta_4 L) v_t^1
\end{aligned}$$

On the right-hand side, we have an MA(1) process and a white noise process, the sum of which is also an MA(1) process. This can be confirmed by checking that the autocovariance function of the right-hand side cuts off to 0 after the first lag. Therefore, we have

$$ret_t = \beta_1(1-\beta_4) + \beta_2\beta_3 + \underbrace{\beta_4 ret_{t-1}}_{\text{AR(1)}} + \underbrace{\beta_2 v_t^2 + v_t^1 - \beta_4 v_{t-1}^1}_{\text{MA(1)}}$$

which shows that ret_t is ARMA(1,1).

(c)

By Wilks's theorem, $2(\ell_{ur} - \ell_r) \xrightarrow{D} \chi^2(2)$ where ℓ_{ur} is the log-likelihood from the first regression and ℓ_r is the log-likelihood from the second (restricted) regression. Therefore the test statistic is $2(24.3 - 21.6) = 5.4$, and the hypothesis is rejected at the 10% level but not the 5% level (the 5% critical value for a $\chi^2(2)$ distribution is 5.991). Taking the 5% level as our standard, this implies returns are not better predicted by the ARMA(1,1) model relative to the model of white noise with non-zero mean.

(d)

From the formula given, we can see that the Schwarz Bayesian Criterion includes a (positive) penalty for the number of estimated parameters but is more negative the higher the likelihood of the estimated parameters is. This means the criterion incorporates a trade-off between model parsimony and model fit. Using this as a criterion, the lower SBC value should be preferred: in this case, the simpler ARMA(0,0) model is preferred for its parsimony and the same implications follow from above.

(e)

From (b), we found the following expression

$$ret_t = \beta_1(1-\beta_4) + \beta_2\beta_3 + \underbrace{\beta_4 ret_{t-1}}_{\text{AR(1)}} + \underbrace{\beta_2 v_t^2 + v_t^1 - \beta_4 v_{t-1}^1}_{\text{MA(1)}}$$

If this is the true representation of the data-generating process, $\beta_2 = 0$ would imply that the AR and MA coefficients are of the same magnitude but opposite signs. Finding some hint of this in the data might imply that β_2 is indeed equal to 0, in which case ret_t is just white noise with a possible intercept. This is consistent with ret_t not Granger-causing div_t, but it might be harder to explain why div_t seems to Granger-cause ret_t.

Question 5

(a)

Letting $y_t = \begin{pmatrix} 1 & LHRP_t & LREARN_t \end{pmatrix}'$, if house prices and earnings are cointegrated, there must be a cointegrating vector β such that $\beta' y_t$ is stationary. Furthermore, the cointegrating vector must be unique. If there is a true cointegrating relationship between the two variables, the OLS regression of $LRHP_t$ on $LREARN_t$ will be super-consistent and the residuals should be stationary. In the second set of Dickey–Fuller regressions, the AIC and SBC both favour the non-augmented Dickey–Fuller test, which yields a test statistic of -1.6657. The Engle–Yoo 5% critical value for a Dickey–Fuller test with 2 parameters in the cointegrating relationship is -3.37, which means the null hypothesis that $\hat{u}_t$ has a unit root is not rejected at the 5% level. However, the null hypothesis is rejected at the 5% level in the ADF(3) and ADF(4) tests.

As mentioned, the cointegrating vector should be unique, so if the estimated relationship in the regression above contains the true cointegrating vector, then $LRHP_t - LREARN_t$ should still be I(1). For the tests done in the first set of Dickey–Fuller regressions, the 1%, 5%, and 10% critical values for a test with an intercept are -3.43, -2.86, and -2.57. The AIC and SIC favour the non-augmented Dickey–Fuller test, which does not reject the null hypothesis that $LRHP_t - LREARN_t$ has a unit root at the 10% level. However, our conclusion does depend on which Dickey–Fuller test we choose to believe; the ADF(2) and ADF(3) test statistics are significant at the 5% level, while the ADF(4) test statistic is significant at the 1% level. It seems that on the whole the evidence favours the hypothesis that $LHRP_t - LREARN_t$ is stationary, if only relative to the likelihood that $\hat{u}_t$ is stationary. But there is probably still some reason to doubt that $LHRP_t$ and $LREARN_t$ are cointegrated.

(b)

If it is true that $LRHP_t - LREARN_t$ is stationary, which must mean $\hat{u}_{t-1}$ is I(1), then the first regression equation includes a non-stationary covariate. If true, the regression equation must be misspecified since the left-hand side is stationary; it cannot possibly be explained by an I(1) variable. Furthermore, the distribution of the estimated coefficient is non-standard, even asymptotically. Since the data seems to reject the null hypothesis of nonstationarity more convincingly in the case of $LRHP_t - LREARN_t$, there is a case to prefer the second specification.

(c)

In the first sub-sample ($1970Q1$–$2000Q4$), there are 124 observations, making the mean squared error $0.046713/124 \approx 0.00038$. For the second sub-sample ($2001Q1$–$2006Q4$), the mean squared error is $0.004717/24 \approx 0.00020$. Squaring the mean prediction error gives us $0.079504^2 \approx 0.00632$. Due to Jensen's inequality, this would be a lower bound on the mean squared prediction error, but this is already one order of magnitude greater than the mean squared errors from either of the sample periods. We might conclude that the out of sample performance of this specification is poor. There might be some reason to suspect there is some structural break in the long-run relationships within the data.

(d)

Estimated over the full sample, the Dickey–Fuller statistics are much less negative and the null hypothesis of no unit roots is not rejected at the 10% level. In the second regression, the coefficient on the interaction term is (a) highly statistically significant and (b) close in magnitude to the coefficient on $LRHP_{t-1} - LREARN_{t-1} - 3.7205$ but of the opposite sign. This would seem to indicate that the long-run cointegrating relationship between $LRHP_t$ and $LREARN_t$ disappears in the second subsample, and our earlier suspicion of a structural break seems to be supported. This is possibly why the forecast performance before was relatively poor.

(e)

Whatever the conclusions about the model's ability to 'explain' house prices (which have more to do with how well it predicts them), the more interesting takeaway might be that there was possibly an uncoupling of house prices and average earnings some time around 2000.

Question 6

Looking at equation 1, we can use the regressions involving $\Delta\hat{u}_t$ to test if the estimated relationship in $Eq1a$ includes the true cointegrating vector. Ignoring the presence of $DUM79_t$, the relevant critical values are the Engle–Yoo critical values for 2 parameters. The 5% critical value is -3.37 and the 10% critical value is -3.02. The Dickey–Fuller statistic is $-0.1371/0.045 \approx -3.0467$, and the Augmented Dickey–Fuller statistic with 1 lagged dependent variable is $-0.1178/0.045 \approx -2.6178$. The non-augmented Dickey–Fuller test would marginally reject the null hypothesis that $\hat{u}_t$ is I(1) at the 10% level, but not at the 5% level, and it is not the test favoured by the SBC. However, it is worth noting that the coefficient on $LSILVER_t$ is close (in economic terms) to 1, although it is significantly different from 1 (with t-statistic $(1.0198 - 1)/0.041 \approx 0.4829$). Also, the estimated constant is close to $\ln 50 \approx 3.912$, which would be the true value if there is a long-run tendency for gold to be 50 times the price of silver. The difference from $\ln 50$ is not significant at the 10% level (with t-statistic $(3.8221 - 3.912)/0.051 \approx -1.7627$). One caveat is that the Durbin–Watson statistic for $Eq1a$ is much closer to 0 than 2, which indicates some autocorrelation in the residuals and could be a symptom of some dynamics not captured by the model.

The second equation can be used to test if $RATIO_t$ is stationary, that is, if $LGOLD_t$ and $LSILVER_t$ are cointegrated with cointegrating vector $\begin{pmatrix}1 & -1\end{pmatrix}'$. The relevant critical values are the Dickey–Fuller critical values for a model with intercept and trend. The 5% critical value is -3.45, and the 10% critical value is -3.24. The test statistics are $-0.1843/0.048 \approx -3.8396$ for the non-augmented Dickey–Fuller test and $-0.1605/0.049 \approx -3.2755$ for the Augmented Dickey–Fuller test with first lags. In both cases the null hypothesis that $RATIO_t$ is I(1) is rejected at the 10% level, and it is rejected at the 5% level in the non-augmented test, which is favoured by the SBC. There is a stronger case to argue that $RATIO_t$ is stationary from these results. Furthermore, the trend terms are not significantly different from 0, which implies that the long-run relationship is somewhat stable (though the intercept is not close to $\ln 50$; this may be due to the likely nonstationary regressor).

Equation 3 is a vector error-correction model where the true cointegrating vector is assumed to be $\begin{pmatrix} 1 & -1 \end{pmatrix}'$. The estimated coefficients on $\begin{pmatrix} LGOLD_{t-1} \\ LSILVER_{t-1} \end{pmatrix}$ imply that when $LGOLD_{t-1} - LSILVER_{t-1}$ is higher than usual, the price of gold will tend to decrease (since the coefficient in $Eq3a$ is negative) and the price of silver will tend to increase (since the coefficient in $Eq3b$ is positive). This provides some support for the investor's investment strategy.