Time Series Models
Supervision 2

Samuel Lee

Question 1

i.

If we want to say that shocks leading to severe recessions are more persistent than other shocks, the estimated coefficients on $DUM1 \times y_{t-1}$ and $DUM2 \times y_{t-1}$ are of the correct sign. The dummies are coded as 1 when the preceding growth rates are relatively high, so a negative coefficient implies that less of the previous year's growth rate "carries through" to the current year if growth has been relatively high.

For the coefficient on $DUM1 \times y_{t-1}$, the standard error is even larger than the point estimate, so it is definitely not statistically significant. In the case of the model with *DUM2*, the t-statistic is $-\frac{0.6032}{0.2561} \approx -2.3553$, and the estimated coefficient is significantly negative at the 1% level (as well as being of a much greater magnitude than the previous estimated coefficient).

The second model thus provides some statistically significant and quantitatively important evidence that shocks leading to severe recessions are more persistent. The evidence from the first model is a bit noisier and hints at a much smaller difference in persistence, but is nevertheless suggestive of some increased persistence for negative shocks.

ii.

If the recession dummy was included as the result of a lengthy specification search, the p-value is not an accurate representation of the certainty we have about our point estimate, since we are ignoring model uncertainty. If we think the lengthy search was a form of p-hacking, our posterior belief on the probability that recessionary shocks are more persistent should be updated downwards; the negative estimated coefficient has given us less information on the true coefficient than we initially thought it did. Also, $\Pr(\beta < 0 \mid \text{search was lengthy})$ is probably small.

Question 2

i.

A simple expansion gives

$$\begin{aligned} 1 - \rho_1 L - \rho_2 L^2 &= (1 - \lambda_1 L)(1 - \lambda_2 L) \\ &= 1 - (\lambda_1 + \lambda_2)L + \lambda_1 \lambda_2 L^2 \end{aligned}$$

which means

$$\begin{aligned} \rho_1 &= \lambda_1 + \lambda_2 \\ \rho_2 &= -\lambda_1 \lambda_2 = -\lambda_1(\rho_1 - \lambda_1) \\ \lambda_1^2 - \lambda_1 \rho_1 - \rho_2 &= 0 \\ \lambda_1 &= \frac{\rho_1 \pm \sqrt{\rho_1^2 + 4\rho_2}}{2} \\ \lambda_2 &= \frac{\rho_1 \mp \sqrt{\rho_1^2 + 4\rho_2}}{2} \end{aligned}$$

ii.

If y_t is I(1) then its lag polynomial has a factor of $(1 - L)$. If $\rho_1 + \rho_2 = 1$, we have

$$\begin{aligned} \lambda_1 &= \frac{1 - \rho_2 \pm \sqrt{(1 - \rho_2)^2 + 4\rho_2}}{2} \\ &= \frac{1 - \rho_2 \pm \sqrt{1 + 2\rho_2 + \rho_2^2}}{2} \\ &= \frac{1 - \rho_2 \pm (\rho_2 + 1)}{2} \\ &= 1 \text{ or } -\rho_2 \\ \lambda_2 &= -\rho_2 \text{ or } 1 \end{aligned}$$

which means $(1 - L)(1 + \rho_2)y_t = \varepsilon_t$. Thus $\rho_1 + \rho_2 = 1$ implies y_t is I(1) if $|\rho_2| < 1$. Actually, if we just evaluate $P(L) = 1 - \rho_1 L - \rho_2 L$ at $L = 1$ we get $P(1) = 0 \iff \rho_1 + \rho_2 = 1$, where $P(1) = 0$ implies $P(L)$ has a factor of $(1 - L)$ by the factor theorem.

iii.

We found that y_t has a unit root if and only if $\rho_1 + \rho_2 = 1$, so testing for $\rho_1 + \rho_2 = 1$ is the same as testing for a unit root. So we could perform one of the usual unit root tests, say, the Augmented Dickey-Fuller test. The test statistic is the unadjusted t-statistic of the coefficient on y_{t-1} in a regression of Δy_t on y_{t-1} and $\Delta y_{t-\tau}$, $\tau \in \{1, \ldots, q\}$, and possibly a trend and intercept (unless we already know *a priori* that $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t$). We could select q using some information criterion. Whatever happens, we can compare the Dickey-Fuller statistic against the critical values of the Dickey-Fuller distribution, where the specific distribution used depends on whether we modelled an intercept and/or trend.

Question 3

i.

With some substitutions, we get

$$y_t = (\lambda - 1)c_{t-1} + \varepsilon_t^c + \varepsilon_t^\tau$$

If y_t is an ARMA(1,1) process, $(1 - \varphi L)y_t$ is an MA(1) process for some $\varphi \neq 0$. We have

$$\begin{aligned}(1 - \varphi L)y_t &= (\lambda - 1)c_{t-1} - \varphi(\lambda - 1)c_{t-2} + \varepsilon_t^c - \varphi\varepsilon_{t-1}^c + \varepsilon_t^\tau - \varphi\varepsilon_{t-1}^\tau \\ &= (\lambda - \varphi)(\lambda - 1)c_{t-2} + \varepsilon_t^c + (\lambda - 1 - \varphi)\varepsilon_{t-1}^c + \varepsilon_t^\tau - \varphi\varepsilon_{t-1}^\tau\end{aligned}$$

For the right-hand side to be an MA(1) process, we need at least for the c_{t-2} term to disappear, that is, $\varphi = \lambda$ (we ignore the possibility that $\lambda = 1$). If so, then

$$(1 - \lambda L)y_t = \varepsilon_t^c - \varepsilon_{t-1}^c + \varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau$$

which is the sum of two uncorrelated MA(1) processes. We can show that the autocovariance structure of such processes is identical to that of an MA(1) process by taking simple covariances; because the innovations are orthogonal, all the terms in the autocovariance function beyond lag 1 for the above process are washed out by the independence and zero means of the innovation terms.

The above might only(?) be a necessary condition for y_t to be ARMA(1,1), though it would be a sufficient condition if the innovations are Gaussian.

ii.

The left hand side is just the autocorrelation at lag 1 of $\eta_t - \theta\eta_{t-1}$. This must be equal to

$$\begin{aligned}\rho(1) &= \frac{\operatorname{Cov}\left[\varepsilon_t^c - \varepsilon_{t-1}^c + \varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau,\ \varepsilon_{t-1}^c - \varepsilon_{t-2}^c + \varepsilon_{t-1}^\tau - \lambda\varepsilon_{t-2}^\tau\right]}{\operatorname{Var}\left[\varepsilon_t^c - \varepsilon_{t-1}^c + \varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau\right]} \\ &= \frac{-\sigma_c^2 - \lambda\sigma_\tau^2}{2\sigma_c^2 + (1 + \lambda^2)\sigma_\tau^2} \\ &= \frac{-\frac{\sigma_c^2}{\sigma_\tau^2 + \sigma_c^2} - \lambda\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_c^2}}{2\frac{\sigma_c^2}{\sigma_\tau^2 + \sigma_c^2} + (1 + \lambda^2)\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_c^2}} \\ &= \frac{-q - \lambda(1 - q)}{2q + (1 + \lambda^2)(1 - q)} \\ &= \frac{-\lambda - (1 - \lambda)q}{1 + \lambda^2 + (1 - \lambda^2)q}\end{aligned}$$

As $q \to 0$, the variance of ε_t^τ overpowers everything else in η_t, the MA(1) term of y_t. The autocorelation function then reduces to that of $\varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau$. As $q \to 1$, ε_t^c dominates η_t, and the autocorrelation function reduces to that of $\varepsilon_t^c - \varepsilon_{t-1}^c$.

iii.

If $\lambda = 0$ then $-\frac{\theta}{1+\theta^2} = -\frac{q}{1+q}$ and $\theta = (1+\theta^2)\frac{q}{1+q} > 0$ since $q > 0$.

iv.

If we plot $y = \frac{\theta}{1+\theta^2}$ and $y = \frac{\lambda+(1-\lambda)q}{1+\lambda^2+(1-\lambda^2)q}$ on the same axes, we can see that any set of (θ, λ) which satisfies a given value of y will involve $\theta > \lambda$ as long as $|\lambda|, |\theta| < 1$. That is, we restrict ourselves to the subset of the parameter space which yields a stationary AR component and an invertible MA component. Furthermore, $q \in (0,1)$ since q is a variance-based weight.

We have that $\frac{\theta}{1+\theta^2}$ is monotonically increasing over $(-1,1)$:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{\theta}{1+\theta^2}\right] = \frac{1}{1+\theta^2} - \frac{2\theta^2}{(1+\theta^2)^2} = \frac{1-\theta^2}{(1+\theta^2)^2} > 0 \qquad \text{for } |\theta| < 1$$

The same goes for $\frac{\lambda+(1-\lambda)q}{1+\lambda^2+(1-\lambda^2)q}$ with a slightly messier derivation:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{\lambda+(1-\lambda)q}{1+\lambda^2+(1-\lambda^2)q}\right] &= \frac{1-q}{1+\lambda^2+(1-\lambda^2)q} - \frac{2\lambda(1-q)\left[\lambda+(1-\lambda)q\right]}{\left[1+\lambda^2+(1-\lambda^2)q\right]^2} \\ &= \frac{\left[1+\lambda^2+(1-\lambda^2)q\right](1-q) - 2\lambda(1-q)\left[\lambda+(1-\lambda)q\right]}{\left[1+\lambda^2+(1-\lambda^2)q\right]^2}\end{aligned}$$

The numerator simplifies to

$$\begin{aligned}&(1-q)\left\{1+\lambda^2+(1-\lambda^2)q - 2\lambda\left[\lambda+(1-\lambda)q\right]\right\} \\ =&(1-q)\left[1+\lambda^2+(1+\lambda)(1-\lambda)q - 2\lambda^2 - 2\lambda(1-\lambda)q\right] \\ =&(1-q)\left[1-\lambda^2+(1-\lambda)^2 q\right] > 0 \qquad \text{for } |\lambda| < 1,\ q \in (0,1)\end{aligned}$$

Furthermore, when plotted on the same axes, the graph of $y = \frac{\lambda+(1-\lambda)q}{1+\lambda^2+(1-\lambda^2)q}$ is always above the graph of $y = \frac{\theta}{1+\theta^2}$ over $\theta, \lambda \in (0,1)$:

$$\frac{x+(1-x)q}{1+x^2+(1-x^2)q} - \frac{x}{1+x^2} = \frac{(1+x^2)\left[x+(1-x)q\right] - x\left[1+x^2+(1-x^2)q\right]}{(1+x^2)\left[1+x^2+(1-x^2)q\right]}$$

Again, the numerator simplifies to

$$\begin{aligned}&x(1+x^2) + (1+x^2)(1-x)q - x(1+x^2) - x(1+x)(1-x)q \\ =&\left[1+x^2-x(1+x)\right](1-x)q = (1-x)^2 q > 0 \qquad \text{for } q \in (0,1)\end{aligned}$$

Given that both functions are monotonically increasing over $(0,1)$, this means that the graph of $y = \frac{\lambda+(1-\lambda)q}{1+\lambda^2+(1-\lambda^2)q}$ is always to the left of the graph of $y = \frac{\theta}{1+\theta^2}$, and that the value of λ which attains a given value of y is always smaller than the value of θ which attains that same value of y.

We can apply the Beveridge-Nelson decomposition to the reduced form ARMA(1,1) representation:

$$
\begin{aligned}
y_t &= \lambda y_{t-1} + \eta_t - \theta\eta_{t-1} \\
y_t &= \frac{1-\theta L}{1-\lambda L}\eta_t \\
&= \frac{1-\theta}{1-\lambda}\eta_t + \left(\frac{1-\theta L}{1-\lambda L} - \frac{1-\theta}{1-\lambda}\right)\eta_t \\
&= \frac{1-\theta}{1-\lambda}\eta_t + \frac{(1-\theta L)(1-\lambda) - (1-\lambda L)(1-\theta)}{(1-\lambda L)(1-\lambda)}\eta_t \\
&= \frac{1-\theta}{1-\lambda}\eta_t + \frac{1-\lambda-\theta L+\theta\lambda L-1+\theta+\lambda L-\theta\lambda L}{(1-\lambda L)(1-\lambda)}\eta_t \\
(1-L)Y_t &= \frac{1-\theta}{1-\lambda}\eta_t - \frac{\lambda-\theta}{(1-\lambda L)(1-\lambda)}(1-L)\eta_t \\
Y_t &= \underbrace{\frac{1-\theta}{1-\lambda}\frac{\eta_t}{1-L}}_{\text{Stochastic trend}} - \underbrace{\frac{\lambda-\theta}{1-\lambda}\frac{\eta_t}{1-\lambda L}}_{\text{Stationary AR(1) cycle}}
\end{aligned}
$$

With $\theta > \lambda$, the coefficient on the stochastic trend is less than one, which means that the long run impact of a shock to the trend is less than unity.

v.

Rehashing the steps in 3i., we have

$$
\begin{aligned}
y_t &= (\lambda-1)c_{t-1} + \varepsilon_t^c + g_t + \varepsilon_t^\tau \\
(1-\varphi L)y_t &= (\lambda-1)c_{t-1} - \varphi(\lambda-1)c_{t-2} + \varepsilon_t^c - \varphi\varepsilon_{t-1}^c + g_t - \varphi g_{t-1} + \varepsilon_t^\tau - \varphi\varepsilon_{t-1}^\tau \\
(1-\varphi L)y_t &= (\lambda-\varphi)(\lambda-1)c_{t-2} + \varepsilon_t^c + (\lambda-1-\varphi)\varepsilon_{t-1}^c + g_t - \varphi g_{t-1} + \varepsilon_t^\tau - \varphi\varepsilon_{t-1}^\tau \\
(1-\lambda L)y_t &= \varepsilon_t^c - \varepsilon_{t-1}^c + g_t - \lambda g_{t-1} + \varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau
\end{aligned}
$$

(a) If $g_t = g \ \forall\ t$, then $(1-\lambda L)y_t = g(1-\lambda) + \varepsilon_t^c - \varepsilon_{t-1}^c\varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau$, which means y_t is still an ARMA(1,1) process albeit now with a non-zero mean g. Y_t now has a unit root with drift.

(b) If $g_t = g_{t-1} + \varepsilon_t^g$, then

$$
\begin{aligned}
(1-\lambda L)y_t &= \varepsilon_t^c - \varepsilon_{t-1}^c + g_t - \lambda g_{t-1} + \varepsilon_t^\tau - \lambda\varepsilon_{t-1}^\tau \\
(1-L)(1-\lambda L)y_t &= \varepsilon_t^g - \lambda\varepsilon_{t-1}^g + \varepsilon_t^c - 2\varepsilon_{t-1}^c + \varepsilon_{t-2}^c + \varepsilon_t^\tau - (1+\lambda)\varepsilon_{t-1}^\tau + \lambda\varepsilon_{t-2}^\tau
\end{aligned}
$$

which means Y_t is now an ARIMA(1,2,2) process, and y_t is ARIMA(1,1,2).

Question 4

i.

With the information we have we can carry out the Augmented Dickey-Fuller test. The relevant critical values are those corresponding to a null hypothesis of a unit root with intercept and trend. The Dickey-Fuller statistics are $-\frac{0.0363}{0.015} = -2.42$ and $-\frac{0.0379}{0.014} \approx -2.707$ for Δy_t and Δc_t respectively. Neither of the null hypotheses are rejected at the 10% level; the evidence is not strong enough to reject that the log of income and consumption are integrated.

ii.

Depending on whether the data is seasonally adjusted, I might first de-seasonalise the growth rates by keeping the residuals from the regression of income growth on a set of seasonal dummies. I would then fit several ARMA models of varying lag orders to the data, and also with and without a trend or intercept. I would then select a preferred model using an information criterion, and inspect the autocorrelograms and residual plots to check for anything weird.

iii.

This is just the innovation to the stochastic trend in the Beveridge-Nelson decomposition derived before.

iv.

Very broadly, the permanent income hypothesis would predict a marginal propensity to consume close to 1 out of permanent shocks to income (the exact value depends on the discount factor and interest rate). Because the innovations to the stochastic trend never die out, they are analogous to permanent shocks to income (they move the trend of income permanently onto a higher path).

(a) The t-statistic is $\frac{0.2611}{0.042} \approx 6.2167$ which is statistically significant for any reasonable significance level. This suggests that the marginal propensity to consume is positive, though the point estimate is quite low and difficult to reconcile with the permanent income hypothesis.

(b) The point estimate is even further from 1 than it is from 0, so it is surely statistically significant for whatever significance level we used before. This is evidence against the permanent income hypothesis for the same reason as before.

v.

I would include a dummy variable for the sign of the innovation (0 if positive and 1 otherwise), and also interact it with the innovation term. Then we have $\frac{\partial \Delta c_t}{\partial innovation_t} = \beta + \gamma \times \mathbb{1}(innovation_t \leq 0)$, where β is the original coefficient and γ is the coefficient on the interaction term. Testing for the statistical significance of the estimated γ is one way to test between the rival explanations, though to do this properly we probably should get a subsample of liquidity constrained consumers (perhaps based on household balance sheets) and run the test on those as well.