Section A (200 Marks)

Please attempt TWO questions from the three questions in this section.

Question 1 (100 Marks) – Identification & Frequency Related Filtering.

Part (a): (70 Marks)

i. By the law of total probability (LTP)

$$P[Y(a)=0]=P[Y(a)=0|Z=a]P(Z=a)+P[Y(a)=0|Z=b]P(Z=b)$$

We know that $P(Z=a)=P(Z=b)=0.5$. We can observe the distribution $P[Y(a)=0|Z=a]=P[Y=0|Z=a]$ and calculate it:

$$P(Y=0|Z=a)=P(Y=0,Z=a)/P(Z=a)=.1/.5=0.2$$

We do not observe $P[Y(a)=0|Z=b]$ but since it is a probability, we know that $P[Y(a)=0|Z=b]\in[0,1]$. So, our identification region is

$$H[P[Y(a)=0]]=[0.2\times 0.5, 0.2\times 0.5+0.5]=[0.1,0.6]$$

ii. By the law of iterated expectations (LIE)

$$\begin{aligned}E[Y(b)] &\overset{\text{LIE}}{=} E[Y(b)|Z=a]P(Z=a)+E[Y(b)|Z=b]P(Z=b)\\ &= E[Y(b)|Z=a](.50)+E[Y|Z=b](.50)\end{aligned}$$

The only thing we know about the counterfactual quantity $E[Y(b)|Z=a]$ is that

$$E[Y(b)|Z=a]\in[0,2]$$

For $E[Y(b)|Z=b]=E[Y|Z=b]$, we have

$$
\begin{aligned}
E[Y|Z=b] &= \\
&= 0\cdot P(Y=0|Z=b)+1\cdot P(Y=1|Z=b)+2\cdot P(Y=2|Z=b) \\
&= \frac{.3}{.5}+2\cdot\frac{.08}{.50} \\
&= 0.6+0.32 \\
&= 0.92
\end{aligned}
$$

Hence we have that

$$
\begin{aligned}
E[Y(b)] &\in 0.92(0.50)+[0,2](0.50) \\
&= [0.46, 1.46]
\end{aligned}
$$

We can do an analogous calculation for $E[Y(a)]$. We have

$$
\begin{aligned}
E[Y(a)] &= E[Y(a)|Z=a]P(Z=a)+E[Y(a)|Z=b]P(Z=b) \\
&= E[Y|Z=a](.50)+E[Y(a)|Z=b](.50) \\
&= E[Y|Z=a](.50)+[0,2](.50)
\end{aligned}
$$

and

$$
\begin{aligned}
E[Y|Z=a] &= 0\cdot P(Y=0|Z=a)+1\cdot P(Y=1|Z=a)+2\cdot P(Y=2|Z=a) \\
&= 0+\frac{.25}{.50}+2\cdot\frac{.15}{.50} \\
&= 0.5+0.6 \\
&= 1.1
\end{aligned}
$$

Hence we have that

$$\begin{aligned} E[Y(a)] &\in 1.1(.50) + [0,2](.50) \\ &= [0.55, 1.55] \end{aligned}$$

Putting these two regions together, we have the following bounds on the average treatment effect:

$$\begin{aligned} E[Y(b)] - E[Y(a)] &\in [.46 - 1.55, 1.46 - .55] \\ &= [-1.09, 0.91] \end{aligned}$$

iii. For here and the rest of this problem, let α and β denote the true values of $E[Y(a)]$ and $E[Y(b)]$, respectively. Let $[\alpha_L, \alpha_U]$ and $[\beta_L, \beta_U]$ denote their respective identified sets. The maximin rule is

$$\delta_{\text{MM}} = \begin{cases} 0 & \text{if } \alpha_L > \beta_L \\ 1 & \text{if } \beta_L > \alpha_l \\ [0,1] & \text{if } \alpha_L = \beta_L \end{cases}$$

Since $\alpha_L = 0.55 > 0.46 = \beta_L$, choose $\delta_{\text{MM}} = 0$. Assign everyone treatment a.
As in Manski (2007)

$$\delta_{\text{MMR}} = \frac{\beta_U - \alpha_L}{(\alpha_U - \beta_L) + (\beta_U - \alpha_L)}$$

Using the numbers from part i,

$$\begin{aligned}\delta_{\text{MMR}} &= \frac{1.46 - 0.55}{(1.55 - 0.46) + (1.46 - 0.55)} \\ &= \frac{0.91}{1.09 + 0.91} \\ &= 0.455\end{aligned}$$

The doctor will assign 45.5% of the patients to the new treatment b and the others to the status quo a.

iv. No, this statement does not accurately describe the empirical finding. We can only say that the patients who received the status quo drug and are members of Jazz bands on average lived longer than those receiving the status quo drug who do not have such a background. We cannot say that the membership of a Jazz band increased the probability of patient who received the status quo drug living for 2 years after treatment.
Asking what would happen to this $E(Y|X)$ when we vary X is akin to a hypothetical change in X, where we have no data and so the researcher has confused correlation with causation and has used a counterfactual (expressing what has not happened but what might or would happen if circumstances, i.e. data, were different). The researcher is in effect extrapolating using the assumption of external validity.
However, if the patients who received the status quo drug were randomly assigned as being members of Jazz bands (an impossibility in this case, but possible in more general cases where X could include randomly distributing some other aspect to patients), then the researcher would be correct in saying that an increase in that covariate increases the probability that a patient will live longer on average than a patient who does not have that aspect. But since the distri-

bution of being a member of a Jazz band is non-random and we are dealing with what actually happened (descriptive) we cannot say that being in a Jazz band increases the probability that a patient who receives the status quo drug lives for 2 years.

Part (b): (30 Marks)

i. Period $p = \frac{2\pi}{\omega}$ where ω is the frequency.

$$p = \frac{2\pi}{\omega} = \frac{2\pi}{0.3} \approx 20.9 \text{ years}$$

ii. The gain function is

$$g(\omega) = |c(e^{-i\omega})|$$

where $|c(e^{-i\omega})|$ is the modulus of $c(e^{-i\omega})$, the frequency response function, i.e.

$$|c(e^{-i\omega})| = \sqrt{c(e^{-i\omega})c(e^{i\omega})}$$

As the Kutznets filter is

$$c(L) = b(L)a(L)$$

we can compute the gain

$$g(\omega) = |c(e^{-i\omega})| = |b(e^{-i\omega})||a(e^{-i\omega})|$$

Let us focus on each of the polynomials separately. First consider $b(L) = L^{-5} - L^5$

$$\begin{aligned} |b(e^{-i\omega})| &= \sqrt{b(e^{-i\omega})b(e^{i\omega})} = \sqrt{(e^{i\omega 5} - e^{-i\omega 5})(e^{-i\omega 5} - e^{i\omega 5})} \\ &= \sqrt{2 - 2\cos(10\omega)} = \sqrt{2}\sqrt{1 - \cos(10\omega)} \end{aligned} \quad (1)$$

Next consider $a(L) = \frac{1}{5}(L^{-2} + L^{-1} + L^0 + L^1 + L^2)$

$$\begin{aligned}
|a(e^{-i\omega})| &= \sqrt{\frac{1}{5}(e^{i\omega 2} + e^{i\omega} + e^0 + e^{-i\omega} + e^{-i\omega 2})\frac{1}{5}(e^{-i\omega 2} + e^{-i\omega} + e^0 + e^{i\omega} + e^{i\omega 2})} \\
&= \frac{1}{5}\sqrt{5e^0 + 4(e^{i\omega} + e^{-i\omega}) + 3(e^{i\omega 2} + e^{-i\omega 2}) + 2(e^{i\omega 3} + e^{-i\omega 3}) + e^{i\omega 4} + e^{-i\omega 4}} \\
&= \frac{1}{5}\sqrt{5 + 8\cos(\omega) + 6\cos(2\omega) + 4\cos(3\omega) + 2\cos(4\omega)} \qquad (2)
\end{aligned}$$

Putting (1) & (2) together, we get the result.

Question 2 (100 Marks) – Univariate Time Series & Forecasting.

Part (a): (70 Marks)

i. The MA(2) process is defined by

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

where ϵ_t is a sequence of independent random variables from a distribution with zero mean and constant variance and where θ_1 and θ_2 are parameters. So

$$\mu = E(\epsilon_t) + \theta_1 E(\epsilon_{t-1}) + \theta_2 E(\epsilon_{t-2}) = 0$$

and we have that the autocovariance function at lag 0 (i.e. variance) is

$$\begin{aligned}\gamma(0) &= E[(\epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2})(\epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2})] \\ &= E(\epsilon_t^2) + \theta_1^2 E(\epsilon_{t-1}^2) + \theta_2^2 E(\epsilon_{t-2}^2) \\ &= (1 + \theta_1^2 + \theta_2^2)\sigma^2\end{aligned}$$

where σ^2 is the variance. The autocovariance function at lag 1 is

$$\begin{aligned}\gamma(1) &= E[(\epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2})(\epsilon_{t-1} + \theta_1 \epsilon_{t-2} + \theta_2 \epsilon_{t-3})] \\ &= \theta_1 E(\epsilon_{t-1}^2) + \theta_2 \theta_1 E(\epsilon_{t-2}^2) \\ &= \theta_1(1 + \theta_2)\sigma^2\end{aligned}$$

The autocovariance function at lag 2 is

$$\begin{aligned}\gamma(2) &= E[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_{t-2} + \theta_1\epsilon_{t-3} + \theta_2\epsilon_{t-4})] \\ &= \theta_2 E(\epsilon_{t-2}^2) \\ &= \theta_2\sigma^2\end{aligned}$$

The autocovariance function for any lag $\tau > 2$ is $\gamma(\tau) = 0$. Putting all this together:

$$\gamma(\tau) = \begin{cases} (1+\theta_1^2+\theta_2^2)\sigma^2 & \tau = 0 \\ \theta_1(1+\theta_2)\sigma^2 & \tau = 1 \\ \theta_2\sigma^2 & \tau = 2 \\ 0 & \tau > 2 \end{cases}$$

We see that the process is stationary since the mean, variance and autocovariances are independent of t. Now note that the autocorrelation function at lag τ is $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$. So the autocorrelation function is given by

$$\rho(\tau) = \begin{cases} 1 & \tau = 0 \\ \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} & \tau = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} & \tau = 2 \\ 0 & \tau > 2 \end{cases}$$

ii. From visual inspection, the data are clearly nonstationary as seen from the mean of US real GDP, which increases with time. The mean US real GDP between 1950 and 1970 is somewhere around $\$3$ trillion, while the mean US real GDP between 1990 and 2010 is somewhere around $\$10$ trillion. The correlogram, which plots the sample autocorrelation function against lags of stationary series drops off as the number of lags becomes large, but this does not usually happen for a nonstationary series. So, we might expect to see the correlogram not dropping off even as the number of lags becomes large.

iii.

$$y_t = \alpha + \beta t + u_t \tag{3}$$

There are different ways to induce stationarity in this model. Students get full credit for following any approach correctly.

(a) Trend removal (detrending) to induce stationarity:

$$y_t - \beta t = \alpha + \epsilon_t$$

which is stationary; hence description as ‘trend stationary’. In practice, we would have to estimate β, of course.

(b) We can also induce stationarity by taking first differences, albeit the model is more naturally made stationary through detrending.

$$\Delta y_t = y_t - y_{t-1} = \beta + \epsilon_t - \epsilon_{t-1} = \beta + \Delta\epsilon_t$$

which is a stationary process; hence description as ‘difference stationary’.

(c) Students may also discuss applying other filters to the series to remove nonstationarity.

Note that this model has a deterministic trend and a purely random (white noise) disturbance. Random shocks are only transitory; there is reversion to trend. The mean of y_t is

$$E(y_t) = \alpha + \beta t$$

which is time dependent; hence y_t is nonstationary. The variance is $V(y_t) = V(\epsilon_t) = \sigma^2$, which is independent of time. The autocovariance is also independent of time:

$$C(y_t, y_{t-j}) = E(\epsilon_t \epsilon_{t-j}) = 0 \quad \forall j \neq 0$$

iv. The process in figure 2 seems to be an MA(2) model since the sample autocorrelation function cuts off at lag length 2, which corresponds to the theoretical counterpart for an MA(2) model. The process in figure 3 seems to be an AR(2) model since the sample autocorrelation function does not cut off but the sample partial autocorrelation function cuts off at lag length 2, which corresponds to the theoretical counterparts for an AR(2) model.

Regarding estimation of an MA(2) model figure 2 (zero mean, $\mu = 0$)

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

this cannot be estimated by regression since nothing on the right-hand side is known. We do not know the mean (μ) except assuming it is zero and we do not know the random shocks. We would use maximum likelihood estimation. Also note that from above

$$\rho(\tau) = \begin{cases} \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} & \tau = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} & \tau = 2 \end{cases}$$

Consider

$$r_1 = \frac{\hat{\theta}_1 + \hat{\theta}_1 \hat{\theta}_2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2}$$

and

$$r_2 = \frac{\hat{\theta}_2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2}$$

We have two simultaneous equations involving two variables ($\hat{\theta}_1$ and

$\hat{\theta}_2$) with quadratic terms ($\hat{\theta}_1^2$ and $\hat{\theta}_2^2$) and a cross product $\hat{\theta}_1\hat{\theta}_2$. We can either solve this through tedious algebra or use a computer program to help us.

Regarding estimation of an AR(2) model figure 3, the AR(2) is a legitimate regression equation

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

so OLS may be used – asymptotically valid. The complication is that there are lags of dependent variables on the right hand side so the independent variables are not fixed. OLS estimates of the ϕ_1 parameters are consistent, even thought the regressors are lagged dependent variables, as long as the disturbances, ϵ_t are non-autocorrelated, which can be assessed by plotting and inspecting correlogram of residuals ($\hat{\rho}(\hat{u})$ should be one at lag zero and zero for any lags greater than one). Why? Because autocorrelation at lag 0 is the variance divided by the variance, i.e. one and for lags other than one it is the autocovariance between residuals at different lags divided by the variance and there should be no correlation between residuals at different time periods, hence the autocorrelation should be zero at lags other than 0. Randomness of residuals can be checked as in OLS regressions by plotting and inspecting residuals over time or by using a test statistic (see next part); if the residuals are not random, then the model is inadequate; asymptotically, $\hat{\rho} = r \overset{a}{\sim} N(0, \frac{1}{n})$.

v. A zero mean $\mu = 0$ MA(1) model is written

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

We need to generate observations from what is unknown. We can convert MA to AR if we have a stability assumption that $|\theta_1| < 1$:

$$\begin{aligned}\epsilon_N &= y_N - \theta_1 \epsilon_{N-1} \\ &= y_N - \theta_1 (y_{N-1} - \theta_1 \epsilon_{N-2}) \\ &= y_N - \theta_1 y_{N-1} + \theta_1^2 \epsilon_{N-2}\end{aligned}$$

We will follow this technique by considering the one-period-ahead forecast for an MA(1):

$$\hat{y}_{N+1} = \hat{\epsilon}_{N+1} + \hat{\theta}_1 \hat{\epsilon}_N$$

where

$$\begin{aligned}\hat{\epsilon}_N &= y_N - \hat{\theta}_1 y_{N-1} + \hat{\theta}_1^2 y_{N-2} - \cdots + (-\hat{\theta}_1)^k y_{N-k} \qquad (4) \\ \hat{\epsilon}_{N+1} &= E(\epsilon_{N+1}) = 0 \\ \therefore \hat{y}_{N+1} &= \hat{\theta}_1 \hat{\epsilon}_N\end{aligned}$$

Similarly, the two-period-ahead forecast will be

$$\begin{aligned}\hat{y}_{N+2} &= \hat{\epsilon}_{N+2} + \hat{\theta}_1 \hat{\epsilon}_{N+1} \\ &= E(\epsilon_{N+2}) + \hat{\theta}_1 E(\epsilon_{N+1}) \\ &= 0\end{aligned}$$

So, the two-period-ahead forecast for an MA(1) will be zero, which reflects the short memory of MA processes, which display rapid convergence to the mean (*mean reversion*), i.e. there is rapid reversion of forecasts to the mean value 0 (μ more generally).

The mean square error of the predictor is defined as $MSE(y_{T+j|T}) =$

$E_T[(y_{T+j|T} - y_{T+j})^2]$, so for the MA(1) process this is equal to

$$\begin{aligned} MSE(y_{T+1|T}) &= E_T(\epsilon_{T+1}^2) = \sigma^2 \\ MSE(y_{T+j|T}) &= E_T(\epsilon_{T+j}^2) + 2\theta_1 E_T(\epsilon_{T+j}\epsilon_{T+j-1}) + \theta_1^2 E_T(\epsilon_{T+j-1}^2) \\ &= \sigma^2(1+\theta_1^2) \text{ for } j \geq 1 \end{aligned}$$

Question 3 (100 Marks) – Volatility & Kalman Filtering.

Part (a): (20 Marks)

i. First note that given the independence of η_t and ϵ_t, the mean of y_t is zero:

$$E(y_t) = E(\sigma_t \epsilon_t) = E(\sigma_t)E(\epsilon_t) = 0 \ \because E(\epsilon_t) = 0$$

So the kurtosis of y_t is

$$\frac{E[(y_t - E(y_t))^4]}{(E[(y_t - E(y_t))^2])^2} = \frac{E[y_t^4]}{(V(y_t))^2}$$

Now using the lemma, we get that the kurtosis of y_t is given by

$$3\exp\{2\gamma_h + 2\sigma_h^2 - 2\gamma_h - \sigma_h^2\} = 3\exp\{\sigma_h^2\}$$

When $\sigma_h^2 > 0$, equivalently $\sigma_\eta^2 > 0$, which will be the case for the stochastic volatility model

$$3\exp\{\sigma_h^2\} > 3$$

and since the kurtosis of the standard normal distribution is 3, we can see that the stochastic volatility model displays excess kurtosis relative to the standard normal distribution, i.e. the tails of the distribution will be fatter in the stochastic volatility model than in the standard Normal distribution.

Part (b): (80 Marks)

i. When the state follows a stationary process as it does here, the initial conditions for the Kalman filter are given by its unconditional mean and variance:

$$\mathbf{a}_0 = \mathbf{a}_{1|0} = \mathbf{0}$$
$$\mathbf{P}_0 = \mathbf{P}_0 = \frac{1}{\sigma^2} E(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t') = \begin{bmatrix} 1+\theta^2 & \theta \\ \theta & \theta^2 \end{bmatrix}$$

where the second line follows since $\boldsymbol{\alpha}_t = (y_t, \theta\epsilon_t)'$.
Alternative answer: When the model is stationary, we get the initial conditions for the state from

$$\mathbf{a}_0 = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{c}$$
$$\text{vec}(\mathbf{P}_0) = [\mathbf{I} - \mathbf{T} \otimes \mathbf{T}]^{-1}\text{vec}(\mathbf{RQR}')$$

ii.

$$\mathbf{a}_0 \mathbf{a}_{1|0} = \mathbf{0}$$
$$\mathbf{P}_0 = \mathbf{P}_0 = \frac{1}{\sigma^2} E(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t') = \begin{bmatrix} 1+\theta^2 & \theta \\ \theta & \theta^2 \end{bmatrix}$$

The first prediction error is $v_1 = y_1$ and $f_1 = 1 + \theta^2$.

iii. Updating, we get

$$\mathbf{a}_1 = \begin{pmatrix} y_1 \\ \frac{\theta y_1}{1+\theta^2} \end{pmatrix} \quad \text{and} \quad \mathbf{P}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\theta^4}{1+\theta^2} \end{pmatrix}$$

iv. We get the following prediction equations for α_2:

$$
\begin{aligned}
\mathbf{a}_{2|1} &= \begin{pmatrix} \frac{y_1\theta}{1+\theta^2} \\ 0 \end{pmatrix} \\
\text{and} \quad \mathbf{P}_{2|1} &= \begin{pmatrix} \frac{\theta^4}{1+\theta^2} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & \theta \\ \theta & \theta^2 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1+\theta^2+\theta^4}{1+\theta^2} & \theta \\ \theta & \theta^2 \end{pmatrix} \\
\therefore v_2 &= y_2 - \frac{\theta y_1}{1+\theta^2} \\
\text{and} \quad f_2 &= \frac{1+\theta^2+\theta^4}{1+\theta^2}
\end{aligned}
$$