Problem Set #1

MACS 30000, Dr. Evans
Jingyuan Zhou

Problem 1

Part (b). The articles that I've found is: Soroka, Stuart N., Dominik A. Stecula, and Christopher Wlezien. "It's (Change In) the (Future) Economy, Stupid: Economic Indicators, the Media, and Public Opinion." American Journal of Political Science 59.2 (2014): 457-74. Web. 7 Jan. 2017.

Part (c). The model they used is time-series error correction model(ECM).
$\Delta Y_t = \alpha_0 + \beta_0 \Delta X_t + \Phi(Y_{t-1} - \gamma X_{t-1}) + \varepsilon_t$
ΔY_t is change in US economy at time t, ΔX_t is change in media tone at time t, Y_{t-1} is US economy at time t-1, X_{t-1} is the media tone at time t-1. The error correction rate is captured by Φ and the long-run multiplier is captured by γ.

Part (d). Exogenous variables of this model are $\Delta X_t, Y_{t-1}, X_{t-1}$, and the only endogenous variable of of this model is ΔY_t.

Part (e). This model is dynamic, linear and stochastic. It's dynamic because values of t make a different to X_t; it's linear because the equation shows a linear relationship; the error term makes the model stochastic.

Part (f). Another variable that might be taken into consideration is the global economic indicator that shows the global economic environment because it will also have an effect on the change in US economy.

Problem 2

Part (a). $Y = \alpha_1 D + \alpha_2 A + \alpha_3 M + Average + \varepsilon$
Y is predicted lifespan, D is drug consumption, A is alcohol consumption, M is mental health status, Average is the average age of the population and ε is a normally distributed error.

Part (d). I think the key factor out of these are drug consumption and alcohol consumption because high drug and alcohol consumption could potentially lead to higher chances of accidents, suicide and homicide, which are all ways of death according to analysis from other people.

Part (e). I choose these factors because higher consumptions of drug and alcohol and different mental health status are reasons that lead to lower lifespan of musicians than general population.

Part (f). At first, I need to select a sample of 1000 musicians with information on their real lifespan, drug consumption, alcohol consumption and mental health to be my training dataset. I would also need another 500 musicians's lifespan data to be my test set.

Naively, without any factor, predicted lifespans would all be the same as average lifespan of general population for the decade of their death.

With only the drug consumption factor, I would use information of 1000 musicians to fit my model of $Y = \alpha_1 D + Average$. Then, with information of drug consumption of each of the 500 musicians, I could predict the lifespan of each. Then, we could compare if these predictions are statistically significantly closer to real values compared to what we get from naive model. If the difference is statistically significantly smaller, then it shows that the drug consumption variable is statistically significant.

I'll follow the same approach to test the other factors.

Related online article https://theconversation.com/stairway-to-hell-life-and-death-in-the-pop-music-industry-32735
https://www.researchgate.net/publication/300969552$_Life_expectancy_and_cause_of_death_in_popular_musicia$