Midterm 1 Study Guide

ECON 4753 — University of Arkansas

Prof. Kyle Butts

Overview of Topics:

1. Bivariate Regression
   - Given a regression coefficient, how do we interpret the marginal effect
     - A one unit change in X is associated with a $\hat{\beta}_1$ change in y
     - DO NOT use causal langauge here
   - Understand regressing y on an indicator variable
     - How do you interpret the intercept and the coefficient on the indicator variable (difference-in-means)
   - Regression y on a set of indicator variables for each value of a discrete variable
     - Know what the 'omitted' category means and know the coefficients are difference-in-means
2. Time-series
   - Autocorrelation coefficient and how to calculate it (e.g. calculating ρ_1)
3. Smoothing Methods
   - Two-sided moving average:
     - The formula to calculate: $\hat{y}_t = \sum_{k=-K}^{K} \frac{1}{2K+1} y_{t+k}$
     - Why are there no forecasts on the ends of the time-series
     - How does increasing the number of observations on each side change the forecasting ("smoothing")
     - What sort of time-series features does a high-level of smoothing miss out on?
   - Decomposition of time-series into a Trend term, Seasonality term, and remanining noise: $y_t = T_t + S_t + \varepsilon_t$
     - How to estimate each part:
       1. $2 \times L$ moving average to estimate T_t
       2. regressing $y_t - \hat{T}_t$ on seasonal dummies to estimate S_t

iii. $\hat{\varepsilon}_t = y_t - \hat{T}_t - \hat{S}_t$

- How to interpret each estimate

- One-sided moving average:
  - The formula to calculate: $\hat{y}_t = \sum_{k=1}^{K} \frac{1}{K+1} y_{t-k}$
  - How does increasing the number of observations included change the forecasting ("smoothing")
  - What sort of time-series features does a high-level of smoothing miss out on?
- Simple exponential smoothing
  - The formula to calculate: $\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$
  - The notion of α and 'updating' the forecast
  - How the forecasting smoothness changes with the value of $\alpha \in [0, 1]$

4. Time-series Regression
   - Know the trade-off between local smoothing methods and time-series *regression* models (e.g. better understanding sudden shocks vs. understanding long-term trends and seasonality)
   - Estimating seasonality
     - How to interpret regression table and omitted category
     - Significance tests of coefficients and how to interpret them
     - Comparing two month's via a coefficient test and how to set up the regression properly to do this
     - Year-by-month indicators vs. Monthly indicators
   - Linear time-trend model
     - Interpret coefficient estimate
     - How to forecast with a linear time-trend
     - Why should you not use higher-order polynomial terms for time-trends
   - Piecewise linear trends
     - What is the advantage of piecewise linear trends over a single linear trend
     - What are "breakpoints"?
     - Intuition on how you might select breakpoints (using the MSPE)
   - Indicator for "weird shocks"
     - Understand why you might include an indicator for a weird period of the time-series

Study Questions

Time-series

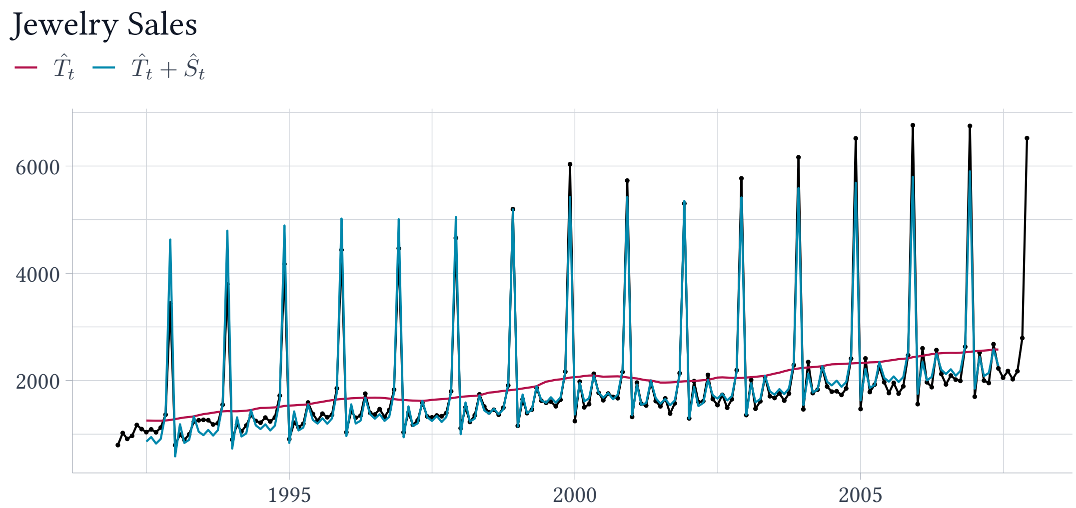
1. Calculate the first lag autocorrelation coefficient, ρ_1, for the following time-series data:

   $1.34, 1.78, 3.20, 4.32, 5.10, 6.24, 6.80$

Smoothing Methods

1. You observe the following time-series for $t = 1, \dots, 7$. By hand calculate the two-sided moving average for $\hat{y}_4$ with $K = 2$ periods on each side

   $1.34, 1.78, 3.20, 4.32, 5.10, 6.24, 6.80$
2. Why might you want to use a larger K for a rolling average when the data is measured noisily (like in aggregate survey statistics)?
3. Say you have a time-series on data where there are a lot of sudden jumps up or down for short periods. For this time-series, would you prefer a larger or smaller K and why?
4. Say you have a time-series observed every month. In words, describe (1) what seasonality is and (2) describe why a smoothing method might not capture seasonality well. You can use an example if that helps
5. Say you have weekly data on the amount of movie tickets sold by AMC Theatres over the last 10 years. Think about decomposing the time-series. Describe the difference between the time-trend, T_t, and seasonality, S_t.
6. Below we present the classical decomposition of monthly jewlery sales in the US.



- Why are there no estimates on the ends of the time-series?
- Would a 3-month moving average do a good job at forecasting into the future? Why or why not?

Time-series Regression

1. We present time-series regression estimates using monthly jewlery sales in the US.

```
OLS estimation, Dep. Var.: jewelry_sales
                  Estimate Std. Error  t value   Pr(>|t|)
(Intercept)      1460.2500    66.2759 22.03289  < 2.2e-16 ***
quarter(date)::2  201.0208    87.7863  2.28989 2.3138e-02 *
quarter(date)::3   86.1458    78.3546  1.09944 2.7298e-01
quarter(date)::4 1550.4375   270.5341  5.73102 3.9133e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Which quarter has the highest sales?
- Does the 4th quarter have a significantly different sales than the 1st quarter?
- Describe how you could modify this regression to test if quarter 2 has significantly different sales than quarter 3

2. Consider extending the time-series past the 2008 recession and into the 2010s

- Why might we be concerned with our time-series regression estimates when including a recession? What could a possible solution be?

3. Given that this data is trending, we add a linear time-trend to our jewelry sales regression

```
OLS estimation, Dep. Var.: jewelry_sales
Standard-errors: Heteroskedasticity-robust
                     Estimate Std. Error  t value   Pr(>|t|)
(Intercept)      -164797.9300 27392.9818 -6.01606 9.2227e-09 ***
year(date)            83.1499    13.7020  6.06844 7.0290e-09 ***
quarter(date)::2     201.0208    48.8162  4.11791 5.7330e-05 ***
quarter(date)::3      86.1458    44.3752  1.94131 5.3724e-02 .
quarter(date)::4    1550.4375   254.4599  6.09305 6.1837e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Are sales trending? In which direction?
- What is the omitted category in this regression?