

CHICAGO BOOTH BUS 41100
SOLUTIONS TO FINAL EXAM SAMPLE #3

INSTRUCTOR: MAX H. FARRELL

These solutions are a guide only! Your answers should show more work/detail/reasoning.

1 Short Answer & Multiple Choice

(a) Choice (ii).

(b) (5) $Y_t = 0.7Y_{t-1} + \varepsilon_t$

(4) $Y_t = 0.9Y_{t-1} + \varepsilon_t$

(1) $Y_t = -0.6Y_{t-2} + \varepsilon_t$

(6) $Y_t = \varepsilon_t$

(2) $Y_t = \sin(2\pi t/12) + \cos(2\pi t/12) + \varepsilon_t$

(3) $Y_t = \sin(2\pi t/12) - t/50 + \varepsilon_t$

(c) FALSE. A time series can exhibit higher order autocorrelations without a first order autocorrelation. For example, suppose $Y_t = \beta_0 + \beta_1 Y_{t-2} + \varepsilon_t$.

(d) (i) $\Delta\varepsilon_t \sim \mathcal{N}(0, 2)$

(ii) No, independence is violated. From above $\Delta\varepsilon_t$ has mean zero and constant variance. To see that the errors are correlated over time, consider $\text{Cov}(\Delta\varepsilon_t, \Delta\varepsilon_{t-1}) = \text{Cov}(\varepsilon_t - \varepsilon_{t-1}, \varepsilon_{t-1} - \varepsilon_{t-2}) = -\mathbb{V}[\varepsilon_{t-1}] = -1 \neq 0$.

(e) (i) **forward** means starting with a base model (empty or not) and building up the model by testing all one-variable **additions** and adding the best, then repeating this, each time adding one variable at a time. **backward** starts with the full model and **deletes** on variable at a time, each time seeing which single variable is the best to delete.

(ii) AIC and BIC penalize the dimension differently, with the AIC penalty not depending on the sample size, while for BIC it does. As long as $\log(n) > 2$, BIC penalizes more, and so will prefer smaller models.

(f) (i) Capital per worker is $k_t := K_t/L_t$. Then using $Y_t = AK_t^{\beta_1} L_t^{\beta_2} e^{\varepsilon_t} = Ak_t^{\beta_1} L_t^{\beta_2 + \beta_1} e^{\varepsilon_t}$, we get $\log(Y_t) = \beta_0 + \beta_1 \log(k_t) + (\beta_2 + \beta_1) \log(L_t) + \varepsilon_t$, so that the elasticity of capital per worker, controlling for the number of workers, is simply β_1 , which can be estimated using

$$\log(Y_t) = \beta_0 + \beta_1 \log(k_t) + (\beta_2 + \beta_1) \log(L_t) + \varepsilon_t \quad \text{or} \quad \log(Y_t) = \beta_0 + \beta_1 \log(K) + \beta_2 \log(L) + \varepsilon,$$

where either way $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and $\beta_0 = \log(A)$.

(ii) Yes: it is linear in parameters, independent over time, and the errors have mean zero, constant variance, and are Normal.

2 Simple Linear Regression

- (a) We already have linearity, we just need Normality and independence: $R_A = \alpha + \beta R_M + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and independent over time and of R_M .
- (b) (i) The phrase “better explain observed stock returns” means all we care about is R^2 , which is higher for GE.
- (ii) For both stocks $\alpha = 0$ is not rejected but $\beta = 1$ is rejected; the former comes directly from the **t value** or the p-value in the table, whereas for the latter forming a confidence interval using the standard errors gives the result, i.e. 1 is more than two standard errors away from β_1 in both cases.
- (c) (i) Since **MarketReturn** is zero, $\hat{R}_{\text{Ford}} = b_0 = 0.12$. The prediction interval is $\hat{R}_{\text{Ford}} \pm 2 \times s_{\text{pred}}$, where

$$s_{\text{pred}} = s \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{(n-1)s_x^2}} = 10 \sqrt{1 + \frac{1}{433} + \frac{(0 - 1.25)^2}{(432)(4.1)^2}} \approx 10.01$$

- (ii) The residual standard error for GE is lower, and that’s the only piece of the above formula that change, so the interval will be narrower.

3 Multiple Linear Regression and Model Building

- (a) We already have linearity, we just need Normality and independence: $R_A = \alpha + \beta R_M + S + V + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and independent over time and of R_M, S, V .
- (b) The phrase “explains the observed variation” means all we care about is this sample, and the biggest R^2 will be the model with the most variables in it, so the Fama-French three-factor model is best for this. We don’t need any output at all for this actually.
- (c) Model (iii) has the smallest BIC value, which is what defines “best” here. This is different because R^2 always increases with more variables, but BIC does not.
- (d) (i) The four choices are (1) where to start (the CAPM model), (2) the universe to search (all main effects), (3) the direction to search (forward), and (4) how to evaluate the steps and stop (BIC and at most two steps).
- (ii) The two shares wind up with different models. This is not surprising since they are very different companies. Even without knowing anything about GE and Ford returns, since they are different Y variables, there’s no reason to expect the same model to be chosen.
- (iii) It is not possible. The CAPM and three-factor model are supposed to be generic: they are for modeling any asset. We have developed two alternatives but we did so in a specific way: each model is specific to a single asset. There’s no reason to believe that either of our new models do any better in general, across different assets, than either the CAPM or the three-factor model. But there’s also no evidence they do any worse either.

4 Logistic Regression

- (a) (i) Give someone a card that should not get one, and thus lose the profit since they will default on their debt.

(ii) Deny someone a card that should get one, and lose the profit from fees, interest, etc.

You can make a sensible argument for either error being worse. Defaulting is usually bad, because the bank gets nothing, but on the other hand, credit card laws are incredibly anti-consumer (in the US), so it is generally worthwhile for a bank to give anyone a card and let them pile up debt.

(b) (i) $81/(92 + 81)$. (ii) $68/(68+571)$.

(c) We first learn that $\log(\text{ratio.spending.income})$ is the best single variable to add first, in terms of BIC improvement. We then learn that once we have it in the model, only two other variables are worth controlling for.

Finally, to interpret the coefficient, for a 1% increase in the spending ratio the odds ratio is multiplied by $\exp\{b_1 \times 0.01\}$, or alternatively $b_1 = 2.75$ is the **percent** change in the odds ratio. See the handout on the course website.

(d) $e^{-1.97}$ times less likely.

(e) (i) As κ goes up, fewer people are given cards in general, and so you make less of the first type of mistake but more of the second. Increasing κ effectively makes the criteria more stringent, so these patterns make sense in that only very good candidates will get cards.

(ii) It depends on how exactly you weight the errors, if indeed you flagged one as worst than the other. If the first type of error is just horrible for you, then you will want the highest κ , but if, for example, the first error is roughly twice as bad, then maybe you prefer the middle choice of κ .

5 Transformations

(a) Holding fixed the own price and the competitors price this would indicate a 290% **fall** in sales from moving from a bad display position to a good one. At face value, this makes no sense, but this interpretation ignores the interaction effect of $\text{Display.QualityGood}:\log(\text{Price})$. When we hold prices fixed and examine the change in sales for a change in display, we have to take these into account as well. For example, if the own **Price** was \$100, then the actual change in moving from bad to good display would be $-2.9 + 0.7 \times \log(100) \approx -2.9 + 0.7 \times 4.6 \approx 0.32$, meaning a 32% **increase** in sales.

(b) Anything outside of the interval $[1.8 \pm 2 \times 0.2]$ is rejected at the 5% level.

(c) Bad: -2.0. Good: $-2.0 + 0.7$.

(d) The elasticity is always negative, which makes sense: your price goes up, your sales fall. The pattern is that for a 1% increase in price, you get a 2% decrease in sales for a carseat in a bad display position, but a smaller fall in the better shelf position. Interpretation: consumers respond to price changes by decrease purchases (or shifting to competitors' products) but that putting your product in a better shelf position (better marketing) can help mitigate this effect: better marketing lowers the price elasticity.

(e) (i) (3) is the best, because: (1) 20% decrease in price, which means a 40% increase in sales; (2) 20% increase in competitor's price, which means a 36% increase in sales; (3) change of $-2.9 + 0.71 \times \log(150) \approx 0.66$, so a 66% increase in sales; (4) 10% decrease in price and a 10% increase in competitor price, meaning a 38% increase in sales.

- (ii) The decision maker would learn a range of likely values for the change in **Sales**. They might want to know this, for example, to assess if the change was really worth making. If the prediction interval included zero, the data does not indicate that the change is worth it for sure.
- (iii) You would need to know the sample variance-covariance matrix of the X variables and their mean.