Ben Allen

Dr. Losak

SAL 366

February 02, 2022

Assignment 2

1.) Approximate equation from *Brown and Jewell (2006)*…

$$realrev_i = \alpha + \beta_1 market_i + \beta_2 footballEntertainment_i + \beta_3 playersall_i + \beta_4 avgrankpointsthisseason_i + \beta_5 oppschedulerank_i + \varepsilon_i$$

Modeling for real revenue using OLS & quantile regression (at three levels) yields the following output…

	Dependent variable:			
	realrev			
	OLS		*quantile regression*	
	OLS	0.25	0.50	0.75
	(1)	(2)	(3)	(4)
market	-0.720	-0.470	-0.700	-0.450
	(0.480)	(0.440)	(0.530)	(0.980)
footballEntertainment	-711,585.000	-380,129.000	-723,073.000	-1,463,480.000***
	(941,220.000)	(376,250.000)	(675,357.000)	(549,684.000)
playersall	1,823,873.000***	1,463,184.000***	1,669,465.000***	2,377,535.000***
	(280,224.000)	(286,789.000)	(288,528.000)	(478,177.000)
avgrankpointsthisseason	83,819.000	335,871.000*	516,944.000	213,676.000
	(355,312.000)	(173,621.000)	(361,850.000)	(755,860.000)
oppschedulerank	7,318,765.000***	5,855,139.000***	6,830,275.000***	5,762,595.000***
	(1,036,264.000)	(743,699.000)	(935,815.000)	(1,470,722.000)
Constant	4,026,826.000	-34,264.000	3,028,343.000	10,532,778.000***
	(3,110,622.000)	(1,225,917.000)	(2,460,510.000)	(2,989,542.000)
Observations	117	117	117	117
R^2	0.720			
Adjusted R^2	0.700			
Residual Std. Error	15,185,950.000 (df = 111)			
F Statistic	56.000*** (df = 5; 111)			
Note:				*p<0.1; **p<0.05; ***p<0.01

Interpretation of ‘playersall’

Throughout all four modeling techniques, the resulting coefficients for the draft pick variables are positive and significant, yet the nature of OLS vs quantile regression leads to slightly different conclusions when testing at different levels. Specifically, the conditional mean approach that comes with simple OLS predicts a future drafted player to bring an additional \$1.823 million to their team. Building off this baseline value, we can more accurately gage the relative confidence of our ‘real revenue’ prediction by running the same model at the 25^{th}, 50^{th}, & 75th percentile. As such, we can interpret that for each model we are x% confident that each beta estimate will be below the predicted (\$1.4 million for .25 quantile, \$2.38 million for .75) Naturally, lower quantiles will take on more conservative estimates of real revenue. One last factor that should be considered when looking at these values is the relatively large standard error/errors for each of these coefficients. As Losak notes in his rebuttal to Brown and Jewell, these values lesson the precision of our potential conclusions.

2.) Estimating with 2SLS

$$\begin{aligned} playersall_i = \alpha_0 + \alpha_1 pool_i + \alpha_2 multiyearrate_i + \alpha_3 avgrankpointsthisseason_i \\ + \alpha_4 oppschedulerank_i + \alpha_5 market_i + \alpha_6 footballEntertainment_i + \varepsilon_1 \quad (1) \end{aligned}$$

$$\begin{aligned} realrev_i = \beta_0 + \beta_1 playersall_i + \beta_2 avgrankpointsthisseason_i + \beta_3 oppschedulerank_i \\ + \beta_4 market_i + \beta_5 footballEntertainment_i + \varepsilon_2 \quad (2) \end{aligned}$$

a.) Upon running the 2SLS model shown above, we can estimate that the true value of an elite college football player is equal to the playersall beta coefficient shown below…

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            -2.54e+06   7.00e+06   -0.36    0.718
playersall              6.11e+06   2.88e+06    2.12    0.036 *
```

Using the 2SLS modeling technique, our estimate of an elite player **<u>(\$6.11 million)</u>** is significantly higher than any of the OLS or quantile regression techniques.

b.) Hausman test output with additional residual variable

```
Call:
lm(formula = realrev ~ market + footballEntertainment + playersall +
    avgrankpointsthisseason + oppschedulerank + residuals, data = data2015)

Residuals:
      Min        1Q    Median        3Q       Max
-41325876  -7067497   -946117   4860543  76172912

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             -2.54e+06   3.86e+06   -0.66   0.5117
market                  -1.71e+00   5.91e-01   -2.89   0.0047 **
footballEntertainment    3.97e+05   1.00e+06    0.40   0.6923
playersall               6.11e+06   1.59e+06    3.85   0.0002 ***
avgrankpointsthisseason -3.24e+06   1.26e+06   -2.57   0.0114 *
oppschedulerank          1.06e+06   2.49e+06    0.42   0.6722
residuals               -4.42e+06   1.61e+06   -2.74   0.0071 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.5e+07 on 110 degrees of freedom
Multiple R-squared:  0.734,     Adjusted R-squared:  0.719
F-statistic: 50.6 on 6 and 110 DF,  p-value: <2e-16
```

In the above output, the presence of statistical significance for the “residuals” variable means that the ‘playersall’ variable is indeed an endogenous regressor in predicting the true value of a player.

c.) By investigating and subsequently accounting for the endogeneity problem within the OLS model, we get a more accurate view of player value that is uncorrelated with selection bias.

d.) Based on the incredibly large beta estimate differences we can see in the output above, this model could very easily be overvaluing a player in an attempt to reduce endegeneity. In addition, a more accurate descriptor of player value (not just playersall, but draft position for example) seems to be more important in this 2SLS model, compared to the OLS model from question 1.

3.)

a. By clustering types of variables together (school, year, etc.), we would hope to decrease heteroskedasticity in connection with the value of an elite college football player. Because of this subsequent inclusion of team fixed effects, our estimates of player value is much smaller ($304,451) – shown below.

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -3463092    2007388   -1.73  0.08482 .
playersall                  304451      60455    5.04  5.7e-07 ***
```

b. In order to derive the value of the 1st and 218th player in each draft class, we would multiply the value coefficient (shown below) with the corresponding pick value.

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -3379069    2017145   -1.68  0.09423 .
value                          546        136    4.00  6.8e-05 ***
```

As such, we get the resulting values…

1st pick: 546 * 1000 = $546,000

218th pick: 546 * 128 = $69,888

c. The accuracy of the following rational entirely depends on the usability of the draft value coefficients. If these values (1000 for 1st, 128 for 218th, etc.) are somewhat accurate, it would seem that the logical approach to estimating true player value would have to further evaluate a player's future value to a more concrete extent than simply whether or not the player was drafted. In the long run, the future value of a player drafted in the first round is much higher than a player drafted in the 7th.

d. Results from both CSE models are shown below…

Clustered SE Models

	Dependent variable:	
	realrev	
	(1)	(2)
playersall	304,451.000	
value		545.000
as.factor(season)2007	4,553,458.000	4,598,219.000
as.factor(season)2008	4,867,165.000	4,939,830.000
as.factor(season)2009	6,471,085.000	6,537,586.000
as.factor(season)2010	7,384,410.000	7,424,755.000
as.factor(season)2011	8,725,264.000	8,776,802.000
as.factor(season)2012	10,048,122.000	10,092,461.000
as.factor(season)2013	11,481,778.000	11,539,710.000
as.factor(season)2014	13,880,524.000	13,967,654.000
as.factor(season)2015	16,111,345.000	16,208,265.000
schoolAlabama	70,297,216.000	71,113,458.000
schoolAlabama-Birmingham	2,374,901.000	2,232,727.000
schoolArizona	19,440,905.000	20,584,075.000
schoolArizona St.	29,897,337.000	31,347,246.000
schoolArkansas	49,111,124.000	51,334,773.000
schoolArkansas State	-620,244.000	-766,726.000
schoolAuburn	65,578,340.000	66,924,047.000
schoolBall State	424,619.000	279,133.000
schoolBaylor	13,857,041.000	15,118,090.000
schoolBoise State	9,855,944.000	9,709,462.000
Constant	-3,463,092.000**	-3,379,069.000***
	(1,634,225.000)	(81,538.000)
Observations	1,073	1,073
R^2	0.950	0.950
Adjusted R^2	0.940	0.940
Residual Std. Error (df = 945)	5,497,061.000	5,523,755.000
F Statistic (df = 127; 945)	143.000***	142.000***
Note:	*p<0.1; **p<0.05; ***p<0.01	

<-break here for rest of schools

4.)

a.) Right off the bat, I would imagine that accounting for a variable like “recruiting value” over some period of time (& not just for a single year) would prove to be more accurate, as it takes time for recruits to develop their skills and ultimately provide value to the program. In the vast majority of instances, players do not enter a team and immediately contribute their future value. In turn, including the lagged variable based on this logic would ultimately help solve the endogeneity problem because it simultaneously reduces the model’s dependency on single year trends.

b.)

Lagged Recruiting Model

	Dependent variable:
	realrev
market	-0.900
footballEntertainment	381,635.000
playersall	964,972.000
avgrankpointsthisseason	427,563.000
oppschedulerank	3,527,031.000
pool	0.015
multiyearrate	59,660.000
laggedRecruitingScore	133,612.000
as.factor(season)2007	-12,779,683.000
as.factor(season)2008	-12,341,782.000
as.factor(season)2009	-10,858,433.000
as.factor(season)2010	-10,539,469.000
as.factor(season)2011	-10,266,619.000
as.factor(season)2012	-10,592,244.000
as.factor(season)2013	-10,573,613.000
as.factor(season)2014	-10,164,802.000
as.factor(season)2015	-9,099,273.000
Constant	-55,759,808.000***
	(19,082,214.000)
Observations	1,071
R^2	0.730
Adjusted R^2	0.720
Residual Std. Error	12,220,645.000 (df = 1053)
F Statistic	166.000*** (df = 17; 1053)
Note:	*p<0.1; **p<0.05; ***p<0.01

The value of a premium college football player is $965,000 based on this methodology. Because there is no accounting for school fixed effects, this prediction is larger (but similar) to the panel approach model, yet well below the 2SLS prediction of over $6 million (as we might expect). Overall, the inclusion of the lagged recruiting variable corresponds to a slight, yet comparable increase in expected player value, which I would surmise is more accurate than school fixed effects in the long run.