Comments on "Essays in Sports Economics" by Radek Janhuba

This dissertation is composed of three separate essays, all of which are related to sports economics. I will provide comments on each of the three essays in turn but I want to make it clear from the start that my comments are intended to be constructive. None of my comments or suggestions should be considered binding. Radek has written a dissertation that, without a doubt, satisfies any reasonable standard for a Ph.D. thesis in economics and, therefore, I am recommending it for a Ph.D. defense.

Comments on "Do Victories and Losses Matter? Effects of Football on Life Satisfaction"

This essay examines the effect of emotional shocks as measured by the outcomes of college football games on life satisfaction. Using data from the Behavioral Risk Factor Surveillance System (BRFSS) for the period 2005-2010, Radek finds that unexpected wins at home lead to increases in the probability that BRFFS respondents report that they are "very satisfied" with their life. By contrast, the outcomes of away games to not appear to have an appreciable impact on life satisfaction. This pattern of results is interpreted as evidence that "emotional shocks caused by unexpected football wins are larger when the experience is shared with others." Consistent with this interpretation, Radek finds a positive association between stadium size and the magnitude of the home-game effect.

Major comments, questions and suggestions

1. In all but one table, estimated coefficients from an ordered logit are reported. These coefficients make gauging the magnitude of the effects nearly impossible. Marginal effects are reported in one table (Table 1.5.2), but I think they are interpreted incorrectly. For example, a marginal effect of .103 should be interpreted as 10 percentage points, which corresponds to a 23% increase relative to the sample mean (.103/.452 = .228). On page 23, Radek refers to this estimate as showing that unexpected wins are associated with an increase of "over 10%" in the probability of being very satisfied.

2. One solution to the problem described above would be to continue to report estimates from an ordered logit model and just include more tables (such as Table 1.52.) with marginal effects. Another solution, which I find much more appealing, is to start with OLS and then, as a robustness check, report marginal effects from an ordered logit. A variety of dependent variables could be used: for instance, an indicator for whether the respondent is very satisfied, an indicator for whether the respondent is satisfied or very satisfied, etc. The advantage of this approach is its transparency and simplicity. Transparency and simplicity are hallmarks of good applied work these days.

3. Speaking of simplicity and transparency, I would have liked to have seen estimates of the effect of any game and the effects of wins and losses before the estimates of unexpected wins and losses.

4. The three-day cut off seems arbitrary. Why not look at the day immediately after the game is played? Why not look at the entire week after the game is played? There's nothing wrong with the three-day window, but readers are going to want to see how sensitive the estimates are to other perfectly plausible windows.

5. In the same vein as the comment above, moving to OLS is going to make constructing an event-study figure easier. An event-study figure will allow the "data to speak". What is the correct cutoff to use? An event-study analysis is the best way to answer this question.

6. I like using Facebook "likes" to identify which team is dominant in a particular county, but I would not start the analysis by restricting the sample in this way. Why not begin the analysis by looking at all BRFSS respondents in counties with college football teams?   Then one might restrict the analysis in several ways. For instance, one might restrict the analysis to counties with only one FBS football team. Again, the idea is to start simple and then get fancy.   Immediately turning to Facebook data is too fancy too fast.

7. Finally, I have a number of concerns regarding interpretation. In general, Radek is claiming too much based on his results.   In particular, the fact that home games are associated with increased life satisfaction but away games are not is, I agree, consistent with the hypothesis that "emotional shocks caused by unexpected football wins are larger when the experience is shared with others", but an alternative interpretation is that the outcomes of home games are simply more salient. Perhaps fans are more likely to watch home games on TV or are more likely to talk about the results of home games. The positive association between stadium size and the size of the estimated effect doesn't get you out of this dilemma. Communities that care more about college football are likely to be served by larger stadiums. If away games don't register with the typical fan (i.e., if their outcome is not salient), then we would expect to see the pattern in Figure 1.5.2. If the home games do register, then the pattern in Figure 1.5.1 could reflect shared experience or it could reflect how much the community cares about college football games regardless of whether they are viewer in person or on the TV.

8. How many games were observed? How many unexpected losses were there in the data? How many unexpected wins? Are the answers to these questions somewhere in the essay and I just missed them? These are important questions for any reader who is interested in gauging how much identifying variation there is in the data. If the answers are tucked away in the appendix, they should be given much more prominence.

Minor Comments:

1. I wouldn't describe the BRFSS as "cross-sectional". It's composed of repeated cross-sections. This is important because you get to see the same county again and again, year after year, which means that you can control for county fixed effects. It occurs to me, because there are multiple games within the year, it is possible to use county-by-year fixed effects.   Has Radek tried putting in county-by-year fixed effects?
2. Try to avoid the phrase "whether or not". Instead, "whether" along works nicely.
3. Try to avoid parentheses within parentheses as in "(see Smith et al. (2008))."
4. There is a great article on football and birth weight that recently came out in the JHR by Duncan, Mansour and a third author whose name I can't remember. It might be worth citing.
5. The essays are written by one person, Radek, but he refers to himself as "we" and he refers to his results as "our results". I found this VERY distracting.
6. I think "home game days" should be "home-game days".

Comments on "Criminals on the Field: A Study of College Football?"

This essay examines the effect of one-the-field officials on penalties incurred by college football players. Using two years of data at the game-play level, Radek and a coauthor find a positive association between the number of game officials and the number of penalties called. Specifically, they find that more offensive holding and more roughing the passer calls are made when an extra official is on the field. They interpret these results as evidence of a monitoring effect.

Comments, questions and suggestions

1. This study exploits the gradual implementation of a policy requiring that games be officiated by 8 as opposed to 7 referees. In the 2013 season, only the Big 12 conference adopted the policy. In the 2014 season, three more conferences adopted the policy. In the 2015 season, the policy was adopted by every team. Because the policy is at the conference-year level, the standard errors should be clustered at the conference-year level (or even at the conference level). Instead, the authors are clustering their standard errors at the referee level.

2. In the same vein as the pervious comment, what advantage is gained by using the play-by-play data? Because the policy is at the conference-year level, the regression could (should?) be run at that level too. Were there any games played there any games played after the adoption of this policy that only used 7 officials?

3. The 8-offical policy was adopted by every team in 2015 so the estimate of λ_2 is really being identified off of the 2015 year dummy. This is made clear in the text on page 61, but it illustrate the dangers of trying to read too much into this natural experiment. Could something else have happened between 2014 and 2015 that affected the propensity to call penalties? Perhaps a change in the interpretation of the rules or a change in the culture of the game?

4. The data being used are rich and it occurs to me that there are other questions that could be addressed using them. Have the authors considered doing something closer to Price and Wolfers (2010) or Parsons et al. (2011)?

References

Price, Joseph and Justin Wolfers. 2010. "Racial Discrimination among NBA Referees." The Quarterly Journal of Economics, Volume 125(4): 1859–1887.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation."American Economic Review, 101 (4): 1410-35.

Comments on "High Bets for the Win? The Role of Stake Size in Sports Betting"

This essay examines the effect of stake size (defined as the amount of money at risk) is associated with the probability of winning. Using data an enormous data set (over 29 million observations) on sports betting, Radek and a coauthor find evidence of an association, but this association is not strong enough to increase net revenue. My plan is to provide a few comments, but the reader should be warned from the outset that I know absolutely nothing about betting or this literature. My comments should be taken with a grain of salt (or even ignored completely).

Comments, questions and suggestions

1. The bottom line seems to be "the effects of stake size on the net revenue of the bettor are statistically insignificant." This conclusion struck me as unsurprising. Isn't this the way markets are supposed to work? If there was a positive association between stake size and revenue, wouldn't this market unravel?

2. Could the positive association between stake size and the probability of winning be driven by how much time and effort is being expended by the bettor? If a lot of money is at stake, then the bettor better do a good job of researching the teams, the players and the conditions.   If not much is at stake, then why expend much effort? In other words, could the positive association simply be a reflection of the effort? If so, then isn't the net revenue always going to be zero taking into account the cost of the effort? I suppose some bettor could be better at researching games than others, in which case some bettors could still see a positive return to their effort.

3. Given that the authors can identify individual bettors, would it be possible to observe how the probability of winning evolves over a "career"? Could the positive association between stakes and winning be driven by this sort of learning? As bettors gain experience, their probability of winning goes up as does their willingness to bet large amounts of money.

4. Could these data be used to explore what types of games and situations encourage the worst bets? When do emotions get in the way of making a profit? Are there games and situations that even seasoned bettors systematically misjudge?

Comments on "Criminals on the Field: A Study of College Football?"

This essay examines the effect of one-the-field officials on penalties incurred by college football players. Using two years of data at the game-play level, Radek and a coauthor find a positive association between the number of game officials and the number of penalties called. Specifically, they find that more offensive holding and more roughing the passer calls are made when an extra official is on the field. They interpret these results as evidence of a monitoring effect.

Comments, questions and suggestions

1. This study exploits the gradual implementation of a policy requiring that games be officiated by 8 as opposed to 7 referees. In the 2013 season, only the Big 12 conference adopted the policy. In the 2014 season, three more conferences adopted the policy. In the 2015 season, the policy was adopted by every team. Because the policy is at the conference-year level, the standard errors should be clustered at the conference-year level (or even at the conference level). Instead, the authors are clustering their standard errors at the referee level.

2. In the same vein as the pervious comment, what advantage is gained by using the play-by-play data? Because the policy is at the conference-year level, the regression could (should?) be run at that level too. Were there any games played after the adoption of this policy that only used 7 officials? One potential advantage would be an increase in precision from being able to included additional Xs, but this argument is not made. (An analogous argument can be made when, for instance, a researcher is using individual-level data to explore the effect of a state-level policy.)

3. The 8-offical policy was adopted by every team in 2015 so the estimate of λ_2 is really being identified off of the 2015 year dummy. This is made clear in the text on page 61, but it illustrates the dangers of trying to read too much into this natural experiment. Could something else have happened between 2014 and 2015 that affected the propensity to call penalties? Perhaps a change in the interpretation of the rules or a change in the culture of the game?

4. The data being used are rich and it occurs to me that there are other questions that could be addressed using them. Have the authors considered doing something closer to Price and Wolfers (2010) or Parsons et al. (2011)?

References
Price, Joseph and Justin Wolfers. 2010. "Racial Discrimination among NBA Referees." *Quarterly Journal of Economics*, 125(4): 1859–1887.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation."*American Economic Review*, 101 (4): 1410-