Journal of Sports Economics

http://jse.sagepub.com

Comments on ``Measuring Parity''

Rodney Fort

Journal of Sports Economics 2007; 8; 642 originally published online May 8, 2007;
DOI: 10.1177/1527002506298122

Comments on "Measuring Parity"

RODNEY FORT
Washington State University

From historical odds Cain and Haddock assign the probability of a tie at 25% and offer functional forms for the standard deviation of a league with equal chances of a win for each team under different point assignment schemes. They then track the ratio of actual to these idealized standard deviations and argue against using an approximation that assigns half a win to both teams in a tie game or match. Their article contains some confusing use of nomenclature and there are some important lapses in scholarship. But perhaps more important, they ignore the fact (after acknowledging it) that if one uses percentages rather than absolute points, their measures of competitive balance are invariant to the point system actually used. Although their point is well-taken for competitive balance questions where absolute points might prove informative, for other competitive balance questions their point is not.

Keywords: *sports economics; competitive balance; measurement; ratio of standard deviations*

From historical odds Cain and Haddock (2006, henceforth, C-H) assign the probability of a tie at 25% and offer functional forms for the standard deviation of a league with equal chances of a win for each team under different point assignment schemes. They then track the ratio of actual to these idealized standard deviations (ISD) for English football and Major League Baseball (MLB) and argue against using an approximation that assigns half a win to both teams in a tie game or match. The aims of this comment are to correct some errors and suggest that C-H miss the actual usefulness of their argument.

C-H breaks with the usual nomenclature in the area, distracting from comparisons with past work. Furthermore, there are scholarship issues. C-H incorrectly identifies the originators of the approaches they detail, does not cite the extant literature on trinomial outcomes in soccer, and ignores a substantial body of literature on the treatment of competitive balance time series. Finally, C-H ignores the fact (after acknowledging it) that if the competitive balance measure

JOURNAL OF SPORTS ECONOMICS, Vol. 8 No. 6, December 2007 642–651
DOI: 10.1177/1527002506298122

they analyze is calculated in percentage terms, it is identical across all of the point assignment schemes they cover. That measure is only different across point assignment schemes if calculated in terms of absolute points. From this perspective, what matters is not which measure is ideal, but how different measures might be more effective at answering the main questions about competitive balance. Each of these issues—nomenclature confusion, scholarship, and the usefulness of balance measurements—is examined in the sections that follow.

For those unfamiliar with the issues concerning the measurement of competitive balance, in all literature prior to C-H, ISD is the standard deviation of game or match outcomes at the end of a season for a league where the probability that one team beats another is equal. If ASD is the actual standard deviation of end-of-season outcomes, then the measure commonly used is the ratio $RSD = \frac{ASD}{ISD}$. RSD is a useful measure of competitive balance because it controls for both season length and the number of teams, facilitating a comparison of competitive balance over time and between leagues. Because RSD = 1 for an actual, perfectly balanced league (because ASD = ISD), competitive balance worsens as RSD increases over unity. RSD has been employed extensively to analyze the behavior of competitive balance over time and in some estimates of the impact of uncertainty of outcome on fan demand (see the review in Fort, 2006). Of course, RSD is only one among many possible measures of competitive balance as explored by Humphreys (2002), Eckard (2003), Lee and Fort (2005), and Fort and Lee (in press).

In the remaining sections, the following convention is adopted. The approximation granting half a win for a tie is denoted (1,0.5,0). It ends up this point assignment scheme actually was used in college football prior to the adoption of game tie-breakers in the middle 1990s. The point assignment scheme that assigns two points for a win, one point for a tie, and zero for a loss is denoted (2,1,0). This scheme was used in English football prior to 1981 and in the National Hockey League prior to the 1998-1999 season (after that, an additional point was granted for an overtime loss). Finally, the scheme that assigns three points for a win, one for a tie, and zero for a loss is denoted (3,1,0) and is the current scheme in many world football leagues.

NOMENCLATURE CONFUSION

C-H breaks with previously developed descriptive nomenclature and it proves confusing. In all literature to date, ISD is used to describe the standard deviation of performance results from a hypothetical completely balanced league. RSD is, instead, the ratio of the actual standard deviation to ISD (as in the introduction, $RSD = \frac{ASD}{ISD}$). But C-H consistently confuses the two, using ISD when they actually are discussing or portraying tabled results for RSD. Relative to the use of that term in all previous literature, C-H never show a single ISD calculation in

TABLE 1: RSD Comparisons, English Football, 1888

					(1,0.5,0)		*(2,1,0)*	
Team	*Wins*	*Ties*	*Losses*	*Games*	*Points*	*Wins %*	*Points*	*%*
1	18	4	0	22	20.000	0.909	40.000	0.909
2	12	5	5	22	14.500	0.659	29.000	0.659
3	12	4	6	22	14.000	0.636	28.000	0.636
4	10	6	6	22	13.000	0.591	26.000	0.591
5	10	2	10	22	11.000	0.500	22.000	0.500
6	10	2	10	22	11.000	0.500	22.000	0.500
7	6	8	8	22	10.000	0.455	20.000	0.455
8	9	2	11	22	10.000	0.455	20.000	0.455
9	7	3	12	22	8.500	0.386	17.000	0.386
10	7	2	13	22	8.000	0.364	16.000	0.364
11	5	2	15	22	6.000	0.273	12.000	0.273
12	4	4	14	22	6.000	0.273	12.000	0.273
				ASD	3.971	0.181	7.943	0.181
				ISD	2.345	0.107	4.062	0.107
				RSD	1.693	1.693	1.955	1.693

NOTE: RSD = ratio standard deviation; ASD = actual standard deviation; ISD = standard deviation of game or match outcomes at the end of a season for a league where the probability that one team beats another is equal.

their entire article. Instead they perform the calculation just described; ASD and ISD are used to calculate RSD, and all of their tabled values are RSD.

This deviation from previous nomenclature contributes to confusion in the C-H Table 1. First, C-H states (2006) that the last row of their Table 1 is "the ISD based on half-win/half-loss assumption for ties" (p. 332). But, as just stated, what really appear in the table are RSD values, not ISD values. Unfortunately, adding further to the confusion, the last row of their Table 1 does not calculate RSD values for the (1,0.5,0) approximation but rather for the (2,1,0) and (3,1,0) point allocation schemes.

Furthermore and especially interesting because C-H argues against use of the (1,0.5,0) approximation, there actually never is any comparison to RSD computed using that approximation in the entire article! The comparison it would seem C-H would want to provide in their Table 1, between RSD for the (1,0.5,0) approximation and RSD for the (2,1,0) point scheme for English football in 1888, is offered in Table 1 with all ASD, ISD, and RSD included for each of the alternatives.

SCHOLARSHIP

C-H incorrectly attributes the original RSD approach to Quirk and Fort (1992) and Fort and Quirk (1995). Although it would be flattering to go down in sports economics history as co-inventor of such a useful measure, it actually was Roger Noll (1988) and Gerald Scully (1989) who developed the RSD approach. Quirk and I made this clear with citations of these sports economics pioneers in Quirk and Fort (1992, p. 244).

C-H also fails to cite an important body of work that already has elegantly developed the trinomial case in soccer. Ruud Koning (2000) pays homage to the originators of the trinomial analysis of soccer outcomes and extends that work significantly (Koning also correctly attributes the RSD approach to Noll [1988] and Scully [1989]). Koning's (2000) analysis of the trinomial case is more sophisticated than what appears in C-H. He examines parity in Dutch soccer using ordered-probit for the trinomial possibilities, adjusted for home field advantage. He also extends our understanding of the underlying probability issues with extensive simulations. Koning's treatment of the trinomial case is now textbook material; Dobson and Goddard (2001, p. 165) adopt it wholesale.

Finally, a growing literature suggests that the C-H analysis of the behavior of parity over time in English football and MLB is not very insightful. It ends up that one cannot simply eyeball a chart and say much of anything about the time series behavior of competitive balance. The lessons are clearly presented in Schmidt (2001), Schmidt and Berri (2001, 2003), Lee and Fort (2005) and Fort and Lee (in press). This literature on the analysis of competitive balance time series includes tests for stationary time series, first-differences approaches, and tests for break points in the time series. Without first performing these tests and transformations, C-H cannot be sure there even is the "u-shaped" behavior of competitive balance over time that puzzles them (their Figure 2).

The C-H measurement choice for MLB deserves special note. C-H states (p. 333), "We believe the similarity of the magnitudes and changes in the ISDs (actually, RSDs) for the two leagues separately provides a justification for looking at a single ISD for each season of MLB" (parenthetical added). This should be a much more careful statement. In MLB, the data actually used to calculate RSD stem from play primarily between teams in separate leagues. Until very recently, with the introduction of a very few interleague games, play is between teams in the same league. It is one thing to *test* that RSD in the American League is equal to RSD in the National League over some time period. It is quite another to say that the researcher *believes* the standard deviation *should be calculated* across all teams in both leagues! RSD calculation across an entire division is valid in the National Basketball Association and National Hockey League (NHL), or across an entire conference in the National Football League, but not across the two leagues in MLB.

THE USEFULNESS OF BALANCE MEASURES

The main C-H contention is that the (1,0.5,0) approximation fails to capture the variation actually occurring under (2,1,0) and (3,1,0) point systems. C-H actually does "tie into the ISD" to make this point, providing ISD formulas (although they do not refer to them by that name) for the following three cases (p. 332 and footnote 4):

$$\text{ISD }(1, 0.5, 0) = \sqrt{0.25 \times N}, \tag{1}$$
$$\text{ISD }(2, 1, 0) = \sqrt{0.75 \times N}, \tag{2}$$
$$\text{ISD }(3, 1, 0) = \sqrt{1.734375 \times N}. \tag{3}$$

If one used percentages instead of absolute point levels, all three point-scoring schemes have the same ISD, namely, $\frac{0.5}{\sqrt{N}}$ (see their footnote 4).

C-H then demonstrates the differences in RSD under the last two alternatives using two full-page appendices and a table. But all of their results actually follow from the fact that they are scalar transformations of the (1,0.5,0) case. First, (because $\sqrt{C \times N} = \sqrt{C} \times \sqrt{N}$), note that for a given N, say, a given league in a given year:

$$\text{ISD }(2, 1, 0) \cong 1.712 \times \text{ISD }(1, 0.5, 0), \tag{4}$$
$$\text{ISD }(3, 1, 0) \cong 3.424 \times \text{ISD }(1, 0.5, 0). \tag{5}$$

Second, because all of the actual variances are based on the same underlying raw data on wins, ties, and losses:

$$\text{ASD }(2, 1, 0) = 2.000 \times \text{ASD }(1, 0.5, 0), \tag{6}$$
$$\text{ASD }(3, 1, 0) \cong 3.938 \times \text{ASD }(1, 0.5, 0). \tag{7}$$

These relationships, in turn, imply (using $\text{RSD} = \frac{\text{ASD}}{\text{ISD}}$):

$$\text{RSD }(2, 1, 0) \cong 1.168 \times \text{RSD }(1, 0.5, 0), \tag{8}$$
$$\text{RSD }(3, 1, 0) \cong 1.150 \times \text{RSD }(1, 0.5, 0). \tag{9}$$

So, the main C-H admonition actually is quantifiable. RSD(1,0.5,0) understates the actual level of imbalance in a (2,1,0) league by about 17% and by about 15% in a (3,1,0) league. Furthermore, because they only differ by scalars, the correlation between any of these measures will be very close to unity.

Now, in discussing the different outcomes they find for RSD, C-H states (p. 332), "This raises the question of which ISD (actually, RSD) is ideal?" The parenthetical is added for clarity. They answer in favor of RSD (2,1,0) or RSD (3,1,0) over RSD (1,0.5,0) because "(a) it is based on a point system used by leagues with draws around the world, (b) the probability of a draw is considered *a priori*, and (c) the denominator is comparatively simple" (p. 332). But the question is more complex.

TABLE 2: RSD Comparisons, English Premier League, 1995-1996 to 2004-2005

	RSD (1,0.5,0)		*RSD (3,1,0)*	
Season	*Points*	*Wins %*	*Points*	*%*
1995-1996	1.641	1.641	1.868	1.640
1996-1997	1.261	1.261	1.466	1.288
1997-1998	1.347	1.329	1.563	1.357
1998-1999	1.561	1.561	1.724	1.514
1999-2000	1.731	1.731	1.979	1.737
2000-2001	1.469	1.469	1.721	1.511
2001-2002	1.768	1.768	2.056	1.805
2002-2003	1.664	1.664	1.887	1.657
2003-2004	1.615	1.615	1.870	1.641
2004-2005	1.771	1.771	2.104	1.848
Average	1.583	1.581	1.824	1.600

NOTE: RSD = ratio standard deviation.

To see this, it is extremely important to return to an earlier finding—there is no difference at all in the RSD measures for the three different point schemes if RSD measures are compared in percentage terms. C-H notes this fact (literally, refer to their footnote 4) but fail to recognize its importance for the question of which RSD is "ideal." Choosing one over another is not just a matter of the underlying point scheme, but also a matter of what recommends the comparison of RSD based on absolute points rather than percentages? I'll return to one answer directly.

Because C-H never actually shows any comparisons of RSD (1,0.5,0) to other RSD measures, a few are offered here. Returning to Table 1, note that the relationships just detailed between ISD, ASD, and RSD all hold. Table 2 compares the (1,0.5,0) approximation and the (3,1,0) point system in use in the English Premier League after 1981, for that league over the last 10 seasons. Again, note that the scalars separating the RSDs are as stated earlier. Furthermore, correlations between any two of the RSD variables exceed 0.98. Figure 1 provides a nice summary of these observations.

For additional comparison purposes, the NHL was a (2,1,0) league until the end of the 1998-1999 season. Since then, an additional point has been awarded for losing an overtime match. Table 3 shows the past 10 years of the (2,1,0) regime for the NHL because I have no knowledge of the distribution of overtime losses. In Table 3, as in the previous tables, all of the scalar differences between RSD hold. Furthermore, the correlations between any pair of measures all are literally unity. Figure 2 is another way of showing the same observations.

Returning to the question of which RSD is ideal, the preceding shows that C-H offer only a partial answer because (a) using percentage measurements gives identical RSD values for all point assignment schemes and (b) the correlation between any of these measures is for all intents and purposes unity. Rather

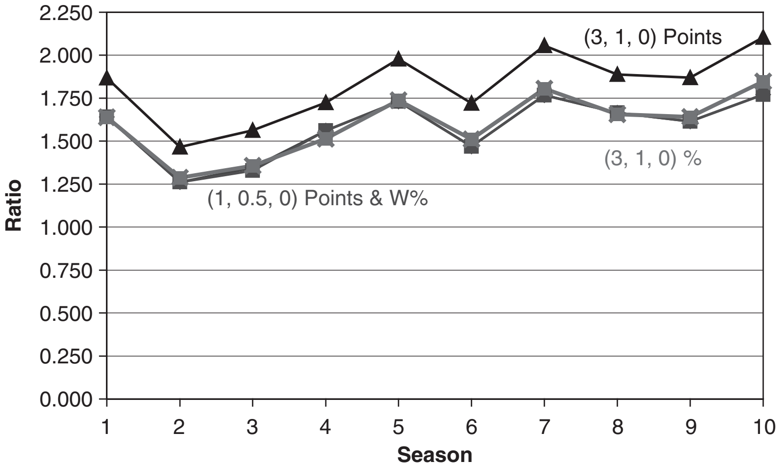

Figure 1: Comparing RSDs from Table 2
NOTE: RSD = ratio standard deviation.

than wonder which is "ideal," let us pose the more pragmatic question, "Which ISD would provide the most precise measure for some question being asked about competitive balance?"

Along these lines, Fort (2006), building on the observations in Fort and Maxcy (2003), suggests that two important questions pervade the competitive balance literature (many other balance issues are raised in Sanderson [2002] and Sanderson and Siegfried [2003]). There is the question of the behavior of competitive balance over time and there is the question the impact of outcome uncertainty on fan demand. The second follows from Rottenberg's (1956) original uncertainty of outcome hypothesis. Since those writings, a third distinction has arisen, namely, estimating the incentive effects of changing from one point allocation scheme to another (2,1,0) to (3,1,0) in football leagues around the world or adding an additional point for an overtime loss in the previously (2,1,0) NHL.

If one aims to only track the behavior of competitive balance over time, there is no reason to prefer either the absolute point version or the winning percentage version over the other. Their correlation is nearly perfect so either measure produces the same track (return to Figures 1 and 2). Surely regression analysis of that behavior would be identical using any of the RSD measures in C-H.

But answers to the other questions may well be more precise using the absolute point version over the winning percentage version. First, RSD has been used to measure one type of outcome uncertainty impact on fan demand (most recently, see Dawson and Downward, 2005). For this question, the "right" measure is the one that most precisely captures the impact of outcome uncertainty

TABLE 3: RSD Comparisons, English Premier League, 1995-1996 to 2004-2005

	RSD(1,0.5,0)		*RSD(2,1,0)*	
Season	*Points*	*Wins %*	*Points*	*%*
1989-1990	1.686	1.686	1.947	1.686
1990-1991	1.868	1.868	2.157	1.868
1991-1992	1.716	1.716	1.982	1.716
1992-1993	2.660	2.660	3.072	2.660
1993-1994	1.882	1.882	2.173	1.882
1994-1995	1.541	1.541	1.779	1.541
1995-1996	2.092	2.092	2.416	2.092
1996-1997	1.411	1.411	1.630	1.411
1997-1998	1.742	1.742	2.011	1.742
1998-1999	1.731	1.731	1.998	1.731
Average	1.833	1.833	2.116	1.833

Note: RSD = ratio standard deviation.

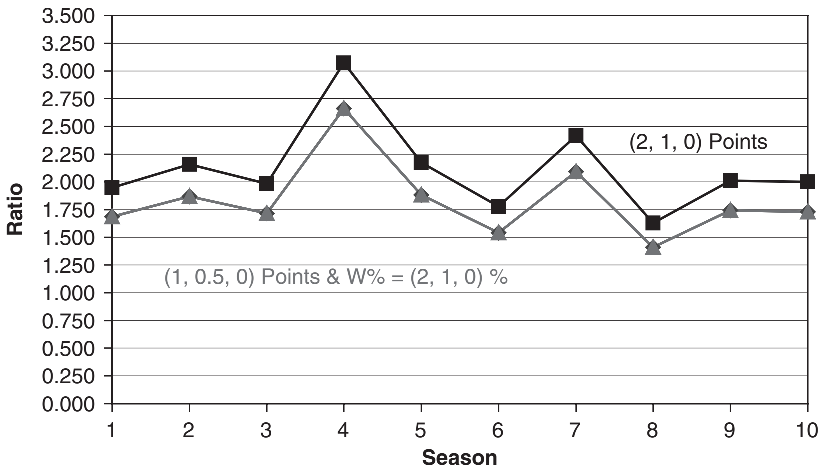

Figure 2: Comparing RSDs from Table 3
NOTE: RSD = ratio standard deviation.

on fan demand; both the level of that uncertainty and its behavior over time are important to that end (Fort & Maxcy, 2003). For fan demand estimation, then, there is a justifiable suspicion that the (1,0.5,0) approximation only imprecisely captures the impact of outcome uncertainty on fan enjoyment and, hence, willingness to pay. But there does remain the nagging result that RSD for each of the point allocation schemes are nearly perfectly correlated whether calculated from absolute points or percentages.

A similar suspicion seems justified when addressing the incentives inherent in, say, moving to (3,1,0) as a replacement for (2,1,0). The literature on this topic is small to date (Abreveya, 2004; Brocas & Carillo, 2004; Easton & Rockerbie, 2005; Newson, 1984) and thus far focused on incentive effects on effort and team outcomes. But once this literature moves on to the impacts of different point allocation schemes on fan demand, absolute point versions of RSD may more precisely capture fan enjoyment. Again, both the level of competitive balance as well as its behavior that effect fan appeal matter.

Given all of this, the C-H "wrap up" is both more modest and more important than they claim. Because it understates imbalance, parity in the NHL was actually about 17% worse than calculated under the RSD (1,0.5,0) approximation prior to the additional one point for an overtime loss point scheme (for example, in Quirk and Fort, 1992, and Szymanski and Smith, 2002). Parity in European football was actually 15% worse than calculated under the RSD (1,0.5,0) approximation (for example, in Szymanski and Smith, 2002). But that is about all that can be said given the C-H article. Over time, absolute point measures and percentage measures of RSD all yield identical changes. So there is nothing in the C-H article to suggest that previous findings about the behavior of competitive balance over time using any measure of RSD need to be revisited. But it may well be that measures based on absolute point schemes will be more effective at capturing the impact of RSD on fan excitement in either demand studies or studies of the effects of different point schemes.

CONCLUSIONS

C-H offer thoughtful and correctly calculated insights about RSD in the case of ties. But their break from nomenclature is needlessly confusing, they incorrectly ascribe the origins of the RSD approach, important previous contributions on the trinomial soccer outcome are missed, and developments in the literature on competitive balance time series are completely absent. Furthermore, the gains in precision they find with RSD measures based on actual point assignment schemes may enhance the analysis of fan demand or the analysis of incentive effects of alternative final season scoring schemes. But no similar gains occur for efforts at tracking competitive balance.

REFERENCES

Abrevaya, J. (2004). Fit to be tied: The incentive effects of overtime rules in professional hockey. *Journal of Sports Economics*, *5*, 292-306.

Brocas, I., & Carillo, J. D. (2004). Do the "three-point victory" and "golden goal" rules make soccer more exciting? *Journal of Sports Economics*, *5*, 169-185.

Cain, L. P., & Haddock, D. D. (2006). Measuring parity: Tying into the idealized standard deviation. *Journal of Sports Economics*, *7*, 330-338.

Dawson, A., & Downward, P. (2005). Measuring short-run uncertainty of outcome in sports leagues: A comment. *Journal of Sports Economics*, *6*, 303-313.

Dobson, S., & Goddard, J. (2001). *The economics of football*. Cambridge, UK: Cambridge University Press.

Easton, S. T., & Rockerbie, D. W. (2005). Overtime! Rules and incentives in the National Hockey League. *Journal of Sports Economics*, *6*, 178-202.

Eckard, E. W. (2003). The ANOVA-based competitive balance measure: A defense. *Journal of Sports Economics*, *4*, 74-80.

Fort, R. (2006). Competitive balance in North American professional sports. In J. Fizel (Ed.), *Handbook of sports economics research* (pp. 190-208). Armonk, NY: M.E. Sharpe.

Fort, R., & Lee, Y. H. (In press). Structural change, competitive balance, and the rest of the major leagues. *Economic Inquiry*.

Fort, R., & Maxcy, J. (2003). Comment: "Competitive balance in sports leagues: An introduction." *Journal of Sports Economics*, *4*, 154-160.

Fort, R., & Quirk, J. (1995). Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature*, *XXXIII*, 1265-1299.

Humphreys, B. R. (2002). Alternative measures of competitive balance in sports leagues. *Journal of Sports Economics*, 3, 133-148.

Koning, R. H. (2000). Balance in competition in Dutch soccer. *The Statistician*, *49*(Part 3), 419-431.

Lee, Y. H., & Fort, R. (2005). Structural change in baseball's competitive balance: The great depression, team location, and racial integration. *Economic Inquiry*, 43, 158-169.

Newson, G. (1984). Three points for a win: Has it made any difference? *Mathematical Gazette*, *68*, 87-91.

Noll, R. G. (1988). *Professional basketball* (Studies in Industrial Economics Paper No. 144). Stanford University, Stanford, California.

Quirk, J., & Fort, R. D. (1992). *Pay dirt: The business of professional team sports*. Princeton, NJ: Princeton University Press.

Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy*, *64*, 242-258.

Sanderson, A. R. (2002). The many dimensions of competitive balance. *Journal of Sports Economics*, *3*, 204-228.

Sanderson, A. R., & Siegfried, J. J. (2003). Thinking about competitive balance. *Journal of Sports Economics*, *4*, 255-279.

Schmidt, M. B. (2001). Competition in major league baseball: The impact of expansion. *Applied Economic Letters*, *8*, 21-26.

Schmidt, M. B., & Berri, D. J. (2001, May). Competitive balance and attendance: The case of major league baseball. *Journal of Sports Economics*, *2*(No. 2), 145-167.

Schmidt, M. B., & Berri, D. J. (2003). On the evolution of competitive balance: The impact of an increasing global search. *Economic Inquiry*, *41*, 692-704.

Scully, G. W. (1989). *The business of Major League Baseball*. Chicago: University of Chicago Press.

Szymanski, S., & Smith, R. (2002). Equality of opportunity and equality of outcome: Static and dynamic competitive balance in European and American sports leagues. In C. P. Barros, M. Ibrahimio, & S. Szymanski (Eds.), *Transatlantic sport: The comparative economics of North American and European sport* (pp. 109-226). Cheltenham, UK: Edward Elgar.

Rodney Fort, PhD, is professor in the School of Economic Sciences at Washington State University. His research program in sports examines all areas of professional and college sports.