Benjamin Baker
05/05/2022

Model Testing of *Subperfect Game: Profitable Biases of NBA Referees*

By: Ben Baker

University of North Carolina at Charlotte

Benjamin Baker
05/05/2022

Introduction

An important aspect of any econometric analysis is to specify the model that explains the relationship between variables correctly. The specification of a model depends on the data being used and the questions being asked. If we incorrectly specify a model and make conclusions based on incorrect estimates can lead to dire consequences, especially in situations that involve healthcare, taxes, or other important aspects of our daily lives. When trying to narrow down the model that best fits the data, one methodology is to compare the AIC and BIC values of models and find the one with the lowest values. This can provide insight into which models outperform the others. Another methodology is to use a model assumption test that determines whether the given model is appropriate in the given context. Both methods can provide important information that can help find the correct model.

This paper aims to look at multiple count variables as dependent variables and determine what type of model would be most appropriate to estimate the relationship between the given variables. This analysis is an extension of the paper *Subperfect Game: Profitable Biases of NBA Referees*, written by Joseph Price, Marc Remer, and Daniel F. Stone. In this paper, the authors estimate whether NBA referees have biases that favor home teams, teams losing during a game, and teams losing in a playoff series. The authors use the Poisson estimation technique since the dependent variables of interest are count variables. To determine whether the Poisson model is appropriate, we first compare AIC and BIC values between multiple models and find that the Poisson model best fits the data for nondiscretionary turnovers, shooting fouls, and nonshooting fouls. In contrast, an OLS linear-linear model best fits the data for discretionary turnovers. Next, we perform a series of Deviance and Pearson goodness-of-fit tests to test whether the Poisson model is appropriate in this context. We can conclude from these tests that the Poisson model is the appropriate model for this dataset. These results help further prove that the authors correctly specified their models to detect referee bias and provide a basis for determining the best model to estimate referee bias in other sports.

Literature Review

The literature review presented in this paper will detail the various model specifications of related papers on the topic of referee bias within the NBA. Price and Wolfers (2010) investigate whether NBA referees exhibit racial biases. They find that referees tend to call fewer fouls on players of the same race, and more fouls are players of the opposite race. The authors use several methodologies to estimate the biases, which use fouls and points scored as dependent variables and estimate the models using OLS, a linear probability model, a probit model, and an IV model with all models including fixed effects. Deutshcer (2015) investigates whether referees favor home teams, favor star players, and favor losing teams. The author uses a logit model with the dependent variable being a dummy variable indicating that the referee called a foul and finds no evidence of any of these alleged referee biases. In a more recent study, Gong (2022) investigates how a crowds influence on referees impact their biases toward the home team during the COVID-19 pandemic when there were no crowds in attendance. By utilizing a linear probability model with the dependent variables being a dummy variable for whether the foul call was an incorrect call or incorrect non-call, Gong (2022) finds evidence that a crowd’s influence does not cause referees to exhibit biases that favor home teams. This paper will contribute to the economic literature on detecting referee bias within the NBA by determining the appropriate model specifications that best fit the relationship between turnovers and fouls with variables that can be associated with referee bias.

Data and Empirical Strategy

I. Data

The data is play-by-play data from ESPN.com for all regular season and playoff games from the 2002-2003 through 2007-2008 seasons. The play-by-play data is preferred over the game-level box score data, as it distinguishes between the types of turnovers and fouls committed within a game. The turnovers and fouls are split into two categories depending on the type committed. Turnovers are assigned to either discretionary turnovers (DTOs) or nondiscretionary turnovers (NTOs), while fouls are assigned to either shooting fouls or nonshooting fouls. Also included in the dataset is the attendance for each game, whether it was a home game for a given team, scoring differentials for each team at each minute, what quarter the game was in at the current minute, whether it was a playoff game, and how many games a team had won in a playoff series prior to the current game. There are two observations for each minute, one for each team. Games with missing minutes and the final three minutes of games are dropped from the sample due to player behavior changing.

The variables of interest for this analysis are discretionary turnovers (DTOs), nondiscretionary turnovers (NTOs), shooting fouls, and nonshooting fouls. These will be our dependent variables. The turnovers and fouls are separated to distinguish the referee's behavior from the player's behavior. Under the assumption presented by the authors, on average, referees have a varying degree of effect over certain types of turnovers. Following this assumption, referee behavior has more of an effect on DTOs than they do on NTOs. For DTOs, the referees always call these, such as travels or goaltending. For NTOs, the referees do not call these, or their influence on the call is minimal, such as a bad pass that goes out of bounds or shot clock violations. To determine whether referee bias is present, we compare how DTOs are affected relative to how NTOs are affected by variables associated with possible bias. In the case of fouls, it is more difficult to determine which fouls are more affected by referee or player behavior, so we look at the effects individually and not across foul types.

II. Empirical Strategy

Since our four dependent variables are count variables, the authors employ the Poisson estimation technique. For the dependent variables to take on the Poisson distribution, the data for each dependent variable needs to be skewed right, and the mean must be equal to the variance. Based on Table I, II, and Figure 1, which show summary statistics of the different turnover and foul types, we can see that this holds for each of our four dependent variables. By estimating the models under the assumption of the Poisson distribution, the dependent variables will be a log expectation equal to a linear function of the regressors. This allows us to interpret the coefficients as percentage effects. Each model contains dummy variables that indicate home games, playoff games, scoring differentials, attendance variables, interactions between these variables, and quarter and matchup fixed effects. The quarter and matchup fixed effects are substituted out for matchup control variables in the analysis conducted within Table IV from *Subperfect Game: Profitable Biases of NBA Referees*. The formal empirical models are as follows:

$$lnT_D = \widetilde{\beta_0^D} + (\beta_1^D \gamma_1^R + \beta_2^D \gamma_1^P)Z + QFE + MFE + \widetilde{u_1}, \qquad (1)$$

$$lnT_N = \widetilde{\beta_0^N} + (\beta_1^N \gamma_1^R + \beta_2^N \gamma_1^P)Z + QFE + MFE + \widetilde{u_2}, \qquad (2)$$

Where the lnT_D and lnT_N are the log of DTOs and NTOs, respectively. Z is a dummy variable for home games, playoff games, score differential, series differential, etc. The coefficient on Z in the lnT_D (lnT_N) function is the average effect of referee bias plus player behavior of Z on DTOs (NTOs). *QFE* and *MFE* are quarter and matchup fixed effects. The error terms are labeled as $\widetilde{u_1}$ and $\widetilde{u_2}$, respectively. To determine if referees are favoring the home team, teams losing during games to keep games close, and teams losing in a playoff series to extend the series, we test whether the coefficient on *Z* in (1) is statistically significant and in absolute magnitude, greater than the coefficient on *Z* in (2).

As stated in the introduction, the goal of this paper is to determine whether the Poisson estimation technique is the appropriate model to estimate the data and compare the fit of several other estimation techniques relative to the Poisson models. Beginning with the models estimated in Table III and Table IV from *Subperfect Game: Profitable Biases of NBA Referees*, we repeat the estimations using several different methodologies, specifically Negative Binomial and OLS. Then we will compare the AIC and BIC values to find the models with the lowest AIC/BIC values that provide the best fit for the data. Next, for the Poisson models estimated in Table IV from *Subperfect Game: Profitable Biases of NBA Referees*, we will perform Deviance goodness-of-fit and Pearson goodness-of-fit tests to determine if the Poisson model is appropriate in this context.[1] If we find that the Poisson models have the lowest AIC and BIC values and fail to reject the null hypothesis for both goodness-of-fit tests, then the Poisson model was correctly specified for this analysis.

Empirical Results

I. AIC and BIC Comparisons

Table III presents the results of the AIC and BIC comparisons between the models within Table III of *Subperfect Game: Profitable Biases of NBA Referees*, and two OLS specifications.[2]

[1] Due to the models in Table III in *Subperfect Game: Profitable Bias of Referees* being estimated with the quarter and matchup fixed effects, the model is estimated within Stata as “xtpoisson” rather than the “poisson” command. This “xtpoisson” command does not allow us to perform the goodness-of-fit test post-estimation.
[2] The Negative Binomial model was considered for comparison in Table III, however, due to computer limitations, this specification was dropped from the comparison.

The first specification for OLS will be a linear-linear function, and the second specification will be a log-linear function. Displayed in the top (bottom) half of Table III are the AIC and BIC values that correspond to the regressions in the first (second) column of Table III of *Subperfect Game: Profitable Biases of NBA Referees*. The first column models represent the regressions estimated without the home and quarter interaction terms, while the second column regressions include those interactions. Based on the results, we can see that the Poisson specification gives the lowest AIC and BIC values for NTOs, shooting fouls, and nonshooting fouls across both sets of models. This indicates that the Poisson estimation technique provides the best fit for the relationship between these dependent variables and the corresponding data. The Poisson model does not return the lowest AIC and BIC values for DTOs. The OLS specification as a linear-linear function returns the lowest AIC and BIC values across both sets of models for DTOs. Therefore, the OLS linear-linear specification provides the best fit for the relationship between DTOs and the corresponding data.

It is important to note why estimating these dependent variables with OLS, specified as a log-linear function, presents issues. As we can see across both sections of Table III, the models labeled as the "OLS-Log" have significantly fewer observations than the other two model specifications. This is due to taking the log of the dependent variables, which includes zeros. Zeros in this context are helpful and provide essential information about the relationship between the turnovers/fouls and the variables associated with possible referee bias. When taking the log of zero, the answer is undefined, so these observations are automatically dropped from the data. While the OLS log-linear specification returns the lowest AIC and BIC values for each dependent variable in all of the models presented in Table III, it does not correctly capture the true relationship between the dependent and independent variables. Thus, this specification cannot be compared to the other two specifications because it removes valuable information.

Table IV presents the results of the AIC and BIC comparisons between the models from Table IV of *Subperfect Game: Profitable Biases of NBA Referees*, the same two OLS specifications mentioned before, and additionally, a Negative Binomial specification.[3] The first column models do not include series differential and minutes remaining/score differential interactions, while the second column models include these interactions. Again, since the OLS log-linear model drops a significant number of observations, it will not be compared to the other three models. These results show that the Poisson model provides the lowest AIC and BIC values for NTOs and nonshooting fouls. For shooting fouls that include the interactions, the Poisson and Negative Binomial models return the same AIC and BIC values, indicating both provide the best fit for shooting fouls. The OLS linear-linear specification provides the lowest AIC and BIC values across both sets of models for DTOs. This indicates that the OLS model provides the best fit for DTOs, while the Poisson model provides the best fit for NTOs, shooting fouls, and nonshooting fouls.

Additionally, the Negative Binomial results are very similar to the Poisson results. A negative Binomial model is best utilized when the mean does not equal variance (overdispersion or under-dispersion) for a given count dependent variable. If we refer back to Tables I and II, which show summary statistics for the dependent variables, we see that there is a small presence of under-dispersion across all turnovers and foul types. This could explain why the Poisson and Negative Binomial models give similar AIC and BIC values.

[3] The Zero-Inflated Poisson model was also considered for this analysis, but again, it was dropped from the comparison due to computer limitations.

II. Goodness-of-Fit Tests

Table V presents the results of the Deviance goodness-of-fit tests and Pearson goodness-of-fit tests for the Poisson models from Table IV of *Subperfect Game: Profitable Biases of NBA Referees*. These goodness-of-fit tests test whether the Poisson model is appropriate in this context. The null hypothesis of both tests is that the Poisson model is appropriate while rejecting the null hypothesis indicates that an alternative model would be more appropriate. Since the p-values returned from the calculations for each test are all greater than the 1%, 5%, and 10% levels of significance, we fail to reject the null hypothesis for all models. This indicates that the Poisson model is appropriate to model the relationship between DTOs, NTOs, shooting fouls, nonshooting fouls, and the variables associated with possible referee bias.

Conclusion

The primary goal of this paper was to determine whether the Poisson models used in the paper *Subperfect Game: Profitable Biases of NBA Referees* were the correct specification to estimate referee bias in the NBA. First looked at were the AIC and BIC values across models estimated using Poisson, Negative Binomial, and OLS. Overall, the Poisson model was the model that best fit the data when the dependent variable was NTOs, shooting fouls and nonshooting fouls. On the other hand, the OLS linear-linear specification was the model that provided the best fit for when DTOs were the dependent variable. Next, we tested the assumptions for the Poisson model using Deviance and Pearson goodness-of-fit tests. For every model estimated in Table IV from *Subperfect Game: Profitable Biases of NBA Referees*, we found that the Poisson model was appropriate to estimate the models using DTOs, NTOs, shooting fouls, and nonshooting fouls as dependent variables. From these results, we can conclude that the authors correctly specified their models given the type of dependent variables and data used in the analysis.

References:

Deutscher, C. (2015). No referee bias in the NBA: New evidence with leagues' assessment data. *Journal of Sports Analytics*, *1*(2), 91-96.

Gong, H. (2022). The Effect of the Crowd on Home Bias: Evidence from NBA Games During the COVID-19 Pandemic. *Journal of Sports Economics*, 15270025211073337.

Price, J., Remer, M., & Stone, D. F. (2012). Subperfect game: Profitable biases of NBA referees. Journal of Economics & Management Strategy, 21(1), 271-300.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees. *The Quarterly journal of economics*, *125*(4), 1859-1887.

Benjamin Baker
05/05/2022

Tables:

Table I. Summary Statistics			
	Mean	**Variance**	**Standard Deviation**
Discretionary Turnovers (DTOs)	0.0736	0.0718	0.2679
Nondiscrectionary Turnovers (NTOs)	0.1946	0.1851	0.4302
Shooting Fouls	0.2122	0.1925	0.4388
Nonshooting Fouls	0.1847	0.1821	0.4267

Notes: Overall summary statistics for turnover and foul categories.

Table II. Minute-Level Summary Statistics												
	Away			**Home**			**Losing/Tied**			**Winning**		
	mean	**sd**	**var**	**mean**	**sd**	**var**	**mean**	**sd**	**var**	**mean**	**sd**	**var**
Discretionary Turnovers - Overall	0.0769	(0.274)	0.0751	0.0703	(0.262)	0.0686	0.0698	(0.261)	0.0681	0.078	(0.276)	0.0762
Travel	0.0241	(0.155)	0.0240	0.0202	(0.142)	0.0202	0.0210	(0.144)	0.0207	0.0235	(0.153)	0.0234
Three Seconds	0.0066	(0.081)	0.0066	0.0073	(0.085)	0.0072	0.0059	(0.077)	0.0059	0.0082	(0.090)	0.0081
Offensive Foul	0.0451	(0.211)	0.0445	0.0416	(0.203)	0.0412	0.0418	(0.203)	0.0412	0.0452	(0.211)	0.0445
Offensive Goal-Tend	0.0012	(0.034)	0.0012	0.0011	(0.033)	0.0011	0.0011	(0.034)	0.0012	0.0011	(0.033)	0.0011
Nondiscretionary Turnovers - Overall	0.1953	(0.431)	0.1858	0.194	(0.429)	0.1840	0.1918	(0.427)	0.1823	0.1979	(0.433)	0.1875
Bad Pass	0.1261	(0.350)	0.1225	0.1285	(0.353)	0.1246	0.1242	(0.347)	0.1204	0.1308	(0.356)	0.1267
Lost Ball	0.0635	(0.251)	0.0630	0.0603	(0.244)	0.0595	0.0624	(0.248)	0.0615	0.0614	(0.247)	0.0610
Shot Clock	0.0057	(0.075)	0.0056	0.0052	(0.072)	0.0052	0.0052	(0.072)	0.0052	0.0057	(0.075)	0.0056
Nonshooting Fouls - Overall	0.1852	(0.427)	0.1823	0.1843	(0.426)	0.1815	0.175	(0.416)	0.1731	0.1961	(0.439)	0.1927
Personal Foul	0.1556	(0.393)	0.1544	0.1556	(0.393)	0.1544	0.1476	(0.383)	0.1467	0.1649	(0.404)	0.1632
Loose Ball Foul	0.0283	(0.168)	0.0282	0.0275	(0.166)	0.0276	0.0263	(0.162)	0.0262	0.0298	(0.173)	0.0299
Inbounds Foul	0.0002	(0.016)	0.0003	0.0002	(0.015)	0.0002	0.0002	(0.014)	0.0002	0.0003	(0.016)	0.0003
Clearing Foul	0.0007	(0.026)	0.0007	0.0005	(0.023)	0.0005	0.0005	(0.023)	0.0005	0.0007	(0.026)	0.0007
Away From Ball Foul	0.0004	(0.020)	0.0004	0.0004	(0.020)	0.0004	0.0003	(0.018)	0.0003	0.0004	(0.021)	0.0004
Shooting Fouls - Overall	0.2178	(0.444)	0.1971	0.2067	(0.434)	0.1884	0.1992	(0.426)	0.1815	0.2274	(0.452)	0.2043
Nonflagrant Foul	0.2167	(0.443)	0.1962	0.2058	(0.433)	0.1875	0.1981	(0.425)	0.1806	0.2267	(0.452)	0.2043
Flagrant Foul	0.0010	(0.032)	0.0010	0.0009	(0.029)	0.0008	0.0011	(0.033)	0.0011	0.0008	(0.028)	0.0008

Notes: Sample includes all games from 2002-2003 to 2007-2008 seasons with play-by-play data available on ESPN.com; overtime periods and last three minutes from fourth quarters dropped from all games, and games with missing minutes dropped. "Winning" = winning by one or more points at start of quarter; "Losing/Tied" = losing or tied at quarter start. In total there are 632,880 observations: 316,440 minutes, with two observations for each minute (one for each team).

Table III. Comparisons of AIC and BIC

First Column Models	Observations (N)	AIC	BIC
Discretionary Turnovers			
Poisson	630,810	309698.9	309835.1
OLS	632,880	**120166.4**	**120302.7**
OLS - Log	45,450	-78203.23	-78098.54
Nondiscretionary Turnovers			
Poisson	632,880	**629675.6**	**629811.9**
OLS	632,880	720402.7	720539
OLS - Log	114,479	-66407.08	-66291.3
Shooting Fouls			
Poisson	632,835	**662468.3**	**662604.6**
OLS	632,880	744054	744190.3
OLS - Log	107,468	-97272.04	-97155.07
Nonshooting Fouls			
Poisson	632,880	**609043**	**609179.3**
OLS	632,880	709149.6	709285.9
OLS - Log	126,458	-46739.7	-46624.68
Second Column Models	**Observations (N)**	**AIC**	**BIC**
Discretionary Turnovers			
Poisson	630,810	309701.2	309871.5
OLS	632,880	**120166.2**	**120336.6**
OLS - Log	45,450	-78205.54	-78074.67
Nondiscretionary Turnovers			
Poisson	632,880	**629676.2**	**629846.6**
OLS	632,880	720403	720573.4
OLS - Log	114,479	-66401.33	-66256.6
Shooting Fouls			
Poisson	632,835	**662467.4**	**662637.7**
OLS	632,880	744048.5	744218.8
OLS - Log	126,458	-97273.99	-97127.78
Nonshooting Fouls			
Poisson	632,880	**609045.9**	**609216.3**
OLS	632,880	709152.7	709323.1
OLS - Log	107,468	-46734.04	-46590.27

Notes: This table provides AIC and BIC values for the Poisson models presented in Table III from Subperfect Game: Profitable Biases of NBA Referees and two OLS models, one with a non-log dependent variable and the other with a log dependent variable.

Table IV. Comparisons of AIC and BIC

First Column Models	Observations (N)	AIC	BIC
Discretionary Turnovers			
Poisson	42,480	22795.55	22908.09
Negative Binomial	42,480	22795.55	22908.09
OLS	42,480	**9002.695**	**9115.233**
OLS - Log	3,087	-4994.98	-4916.525
Nondiscretionary Turnovers			
Poisson	42,480	**42698.83**	**42811.37**
Negative Binomial	42,480	42738.22	42850.76
OLS	42,480	46725.75	46838.29
OLS - Log	7,246	-3956.122	-3866.576
Shooting Fouls			
Poisson	42,480	**46289.97**	**46402.51**
Negative Binomial	42,480	46301.83	46414.37
OLS	42,480	50388.04	50500.58
OLS - Log	8,367	-5559.886	-5468.469
Nonshooting Fouls			
Poisson	42,480	**46369.65**	**46482.19**
Negative Binomial	42,480	46371.65	46492.85
OLS	42,480	53577.6	53690.14
OLS - Log	7,943	-1512.751	-1422.011
Second Column Models	**Observations (N)**	**AIC**	**BIC**
Discretionary Turnovers			
Poisson	42,480	22791.68	22947.5
Negative Binomial	42,480	22791.68	22947.5
OLS	42,480	**8998.387**	**9154.209**
OLS - Log	3,087	-4992.015	-4883.386
Nondiscretionary Turnovers			
Poisson	42,480	**42694.19**	**42850.01**
Negative Binomial	42,480	42733.27	42889.09
OLS	42,480	46721.13	46876.96
OLS - Log	7,246	-3953.754	-3829.766
Shooting Fouls			
Poisson	42,480	**46218.96**	**46374.78**
Negative Binomial	42,480	**46218.96**	**46374.78**
OLS	42,480	50298.85	50454.67
OLS - Log	8,367	-5568.444	-5441.867
Nonshooting Fouls			
Poisson	42,480	**46295.52**	**46451.34**
Negative Binomial	42,480	46317.22	46481.69
OLS	42,480	53506.98	53662.8
OLS - Log	7,943	-1509.936	-1384.295

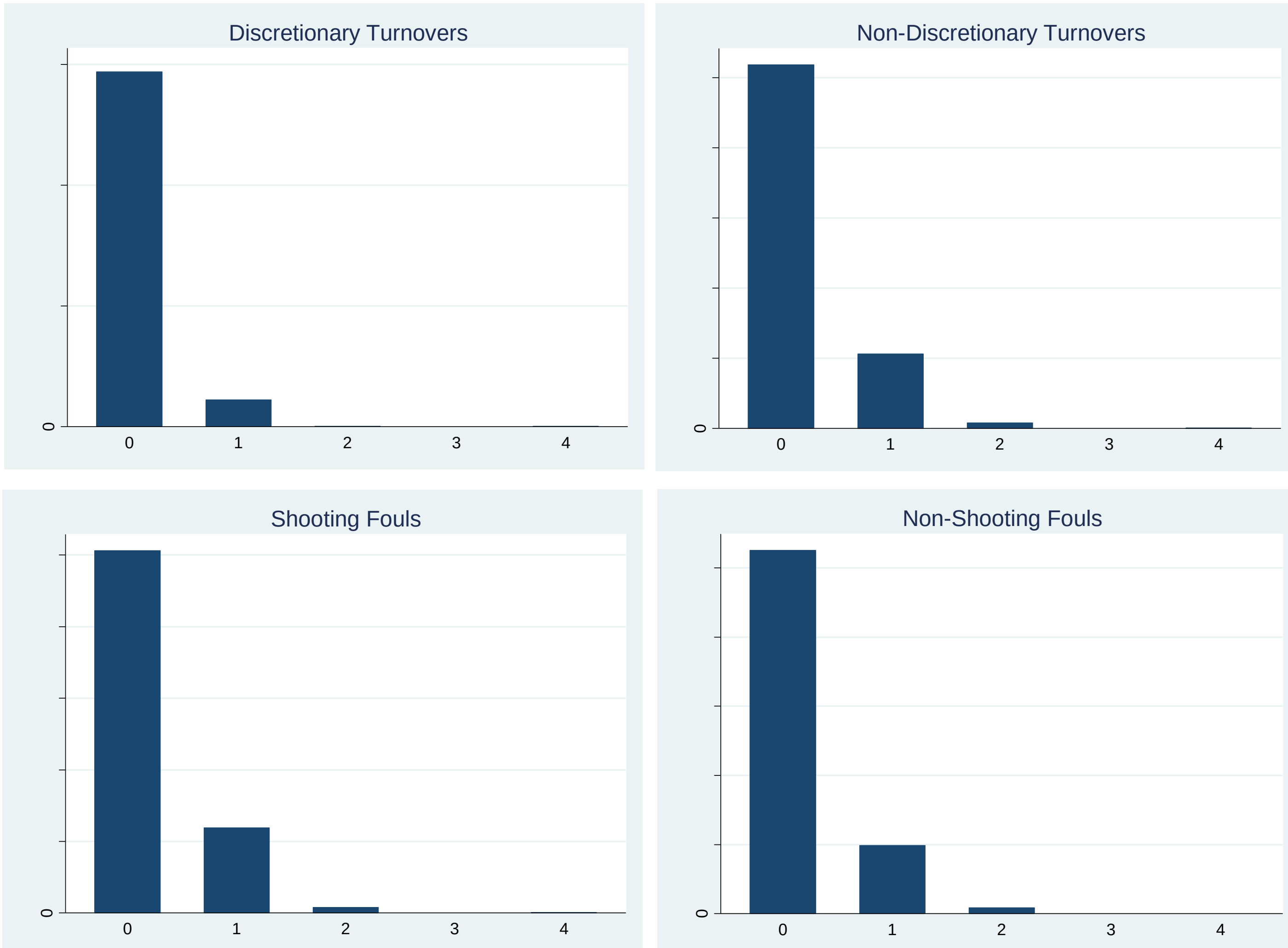
Notes: This table provides AIC and BIC values for the Poisson models presented in Table IV from Subperfect Game: Profitable Biases of NBA Referees, two OLS models, one with a non-log dependent variable and the other with a log dependent variable and a Negative Binomial model.

Table V. Poisson Goodness-of-Fit Tests

First Column Models	Deviance GoF	Prob > chi2	Pearson GoF	Prob > chi2	Conclusion
Discretionary Turnovers	16548.91	1.0000	41292.32	1.0000	Fail to Reject
Nondiscretionary Turnovers	27853.67	1.0000	40723.96	1.0000	Fail to Reject
Shooting Fouls	29191.1	1.0000	38794.18	1.0000	Fail to Reject
Nonshooting Fouls	29944.86	1.0000	42341.81	0.6656	Fail to Reject
Second Column Models	**Deviance GoF**	**Prob > chi2**	**Pearson GoF**	**Prob > chi2**	**Conclusion**
Discretionary Turnovers	16535.04	1.0000	41301.19	1.0000	Fail to Reject
Nondiscretionary Turnovers	27839.03	1.0000	40717.12	1.0000	Fail to Reject
Shooting Fouls	29110.09	1.0000	38729.85	1.0000	Fail to Reject
Nonshooting Fouls	29860.73	1.0000	42289.13	0.723	Fail to Reject

Notes: This table is Deviance and Pearson goodness-of-fit tests for the Poisson models in Table IV from Subperfect Game: Profitable Biases of NBA Referees.

Figure 1:


Discretionary Turnovers
0
0
1
2
3
4
Non-Discretionary Turnovers
0
0
1
2
3
4
Shooting Fouls
0
0
1
2
3
4
Non-Shooting Fouls
0
0
1
2
3
4