Air Quality and Error Quantity: Pollution and Performance in a High-skilled, Quality-focused Occupation*

James Archsmith[†]
University of Maryland, College Park
UC Davis

Anthony Heyes[‡]
University of Ottawa
University of Sussex

Soodeh Saberian[§]
University of Ottawa

Draft Date: December 11, 2017

Please review and cite to the most recent version of this appendix available at:

http://econjim.com/WP1601a

Online Appendix - Not for publication.

*We thank Stefan Ambec, Pierre Brochu, James Bushnell, Janet Currie, David Forrest, Richard Tol, Timo Goeschl, William Greene, Erich Muehlegger, Matthew Neidell, David Rapson, Roberton Williams III and numerous seminar participants at EAERE Venice 2015, University of Sussex and UC Davis for invaluable feedback. Errors are ours.

[†]Department of Agricultural and Resource Economics, University of Maryland, College Park, MD. University of California, Davis, One Shields Ave., Davis, CA 95616 Email: archsmith@econjim.com, URL: https://econjim.com.
He is grateful for financial support from the UC Davis Office of Graduate Studies, College of Letters and Science: Division of Social Sciences, and the UC Davis Department of Economics.

[‡]Department of Economics, University of Ottawa, 120 University Private, Ottawa, Canada. K1N 6N5. He is also a part-time professor of economics at the University of Sussex.
Heyes is Tier 1 Canada Research Chair (CRC) in environmental economics and the financial support of the CRC Program is acknowledged.

[§]Department of Economics, University of Ottawa, 120 University Private, Ottawa, Canada. K1N 6N5.

A.1 Data

A.1.1 PITCHf/x

The PITCHf/x system records the vertical (in feet from ground level) and horizontal (in feet from the center of home plate) position of the center of the ball as it crosses home plate. We consider a pitch to be a strike by MLB rules if any portion of the ball passes through the strike zone. An MLB baseball has a diameter between 2.86 and 2.94 inches. A pitch is recorded as a strike if the center of the ball passes within 1.85 inches of the strike zone.

The strike zone, defined by MLB Rule 2.00, is an area centered over home plate, which is 17 inches wide on its front face, extending vertically from the hollow beneath a batter's kneecap to the midpoint between the top of the shoulders and top of the batter's uniform pants.

Scatter plots of all pitches thrown in 2008 season is shown in Figure A.1. Panel A.1a shows the location of all pitches where the umpires was required to make a decision. Panel A.1b shows the locations of pitches where the umpire made an incorrect call. From this is clear umpires frequently make mistakes, particularly when the pitch is close to the boundary of the strike zone.

As shown in Figure A.1 there are systematic deviations by umpires from the by-rule strike zone. The probability of calling a strike based on pitch location and batter handedness for two specific umpires are shown in Figure A.2. We control for such deviations using umpire-specific strike zone controls.

Data from PITCHf/x are available for all games from 2008 onward. We attribute games with anomalously low rates of correct calls resulted from miscalibration of the PITCHf/x cameras. In our main specification we exclude any game with a mean rate of correct calls less than 0.72. Figure A.3 shows a histogram of the proportion of correct calls by game for 2008 to 2015.

A.1.2 Pollution data

We collect pollution data from the US EPA's AQS, a nationwide network of pollution monitors. To control for heterogeneity in pollution levels over time and across venues our primary regression specification includes venue-month-year controls, so identification of the impact of air quality on decision making comes from within-month deviations at any specific venue in each year. Figure

A.4 shows the distribution of monthly deviations from the venue-specific annual mean for each of the local criteria pollutants. With the exception of PM_{10}, all pollutants show a clear seasonal pattern. CO levels tend to be lowest in the warm months, when automobile engines operate the most efficiently. O_3, which is formed by the combination of NO_X and VOCs in sunlight, has levels peaking in the summer months. SO_2 and $PM_{2.5}$ both peak in the summer.

Umpires are evaluated on their accuracy of their calls for both salary increases and post-season assignments. Moreover, the pitch tracking data is also available to MLB umpires. There has been a steady upward trend in the accuracy of umpires' calls over time. At the same time, improved technology, more stringent regulations, and the financial crisis of 2008 have caused a general downward trend in pollution levels over time. Figure A.5 shows average annual pollution levels during games and the mean probability of umpires making a correct call. Failing to account for these broad trends can lead to a spurious correlation between air quality and the accuracy of decision making.

A.1.3 List of venues

Table A.1 shows the list of venues and number of games played in each venue.

A.1.4 Additional venue data

We compiled additional details of each venue by examining its aerial photographs on Google Earth. We assign the reported latitude, longitude and elevation using the most recent available aerial photographs of the location of home plate.[1] Using this information, we then compute the time to sunset for each game.[2]

A.1.5 Distance calculations

We compute the distance between each monitor and a venue as the Haversine distance between the geographic coordinates of the pollution monitor defined by the EPA and the venue defined in

[1]These time-invariant venue attributes are useful for computing time-varying attributes such as pollution levels, and the position of the Sun relative to the direction the umpire is looking. We are unable to determine the orientation of domed venues as the playing surface is obscured in satellite photographs. However, the position of the sun should be irrelevant in these venues and absorbed by a venue-specific fixed effect.

[2]For purpose of publication Stata code to implement these calculations is available on the author's website.

Google Earth.[3]

A.1.6 Summary statistics

Summary statistics for key variables across the various restrictions placed on the estimation sample are shown in Table A.2. Column (1) show means for all all pitches from 2007 through 2015 where the umpire was required to make a ball/strike call. Column (2) imposes the sample restrictions described in Section 5 including eliminating games where using umpires with less than two full seasons of experience, games where the PITCHf/x equipment appears to be miscalibrated, and pitches where the top or bottom of the strike zone was the operative boundary. Column (3) further limits the sample to observations with non-missing pollution values. Column (4) represents the estimation sample for the primary regression specification, excluding observations with any missing covariates from that specification.

A.2 Quasi-random assignment

An attractive attribute that we have claimed for our setting is that umpires are assigned to venues in a way that is orthogonal to realized pollution levels, conditional on observables. It is however possible that other factors determined after assignment of umpires to games (and potentially after air pollution levels are observed) may be correlated with pollution and the probability of an umpire making a correct call which threaten our causal identification.

In a typical randomized control trial, as a preliminary to presenting results of the experiment, one would generally provide evidence that random assignment to treatment is truly random, namely assignment of the treatment is orthogonal to any unobserved factors affecting the outcome. Clearly, it is impossible to show treatment is uncorrelated with an unobserved variable. However, it is typical in such experiments that the authors present tables of "balance on observables" (e.g., 2014), arguing the lack of correlation of the treatment with observables provides evidence treatment

[3] Detailed descriptions of the calculations are available at http://www.movable-type.co.uk/scripts/gis-faq-5.1.html (accessed on 8/3/2015). We make standard adjustments to the Haversine distance calculations to account for the fact the Earth is an eplisoid, rather than a perfect sphere, however these adjustments are unlikely to make much of a difference over the small distance scales examined here. For purpose of replication, Stata code to implement these calculations is available on the author's website.

is also uncorrelated with unobservables. The null hypothesis of such balance tests is valid randomization, *i.e.*, values of the observed variables are uncorrelated with treatment status. A typical test is a t-test of means for each variable across treatment and control groups. In the context of a continuous treatment variable (D), the null is the conditional mean of Z conditional on treatment is equal to the unconditional mean of Z or:

$$\mathbf{E}\left[Z|D\right] = \mathbf{E}\left[Z\right] \tag{1}$$

Which is identical to a Wald test of $\mathrm{H}_0 : \beta_1 = 0$ in the regression

$$Z_i = \beta_D D_i + \beta_0 + \varepsilon_i \tag{2}$$

We intend each of these tests to provide the observational analog – a test of "conditional balance on observables". Here, we argue that conditional on observables included in our regressions (X) treatment or control status (D) is uncorrelated with any factors not included in our regression. The analogous test would then show treatment is uncorrelated with observable factors conditional on regression covariates or:

$$\mathbf{E}\left[Z|D, X\right] = \mathbf{E}\left[Z|X\right] \tag{3}$$

Which is identical to a Wald test of $\mathrm{H}_0 : \beta_D = 0$ from the regression

$$Z_i = \beta_D D_i + \mathbf{B}_X X_i + \beta_0 + \varepsilon_i \tag{4}$$

One may wish to test more than one observable characteristic for evidence it is conditionally unconfounded with treatment. In this case Z will be a vector of J observables factors. Simply estimating the set of J regressions

$$Z_i^j = \beta_D^j D_i + \mathbf{B}_X^j X_i + \beta_0^j + \varepsilon_i^j \ \ \forall j \in J \tag{5}$$

And performing a Wald test of $\beta_D^j = 0$ for each equation is unsatisfactory for two reasons. First, this is a test with multiple comparisons and one may expect to reject $H_0 : \beta_D^j = 0$ for some portion of J even when H_0 is true. Second, correlation in Z can cause the estimated β_D^j to be correlated across equations causing misleading results when aggregating the series of hypothesis tests. We account for these challenges by estimating all $j \in J$ equations as a system of seemingly unrelated regressions then testing cross-equation equality of $\mathbf{B}^j$.[4] As with our preferred specification, we allow for arbitrary correlation of the errors within clusters defined at the game level.

First, we test conditional balance of full-time umpires over pollution levels as described by regressing indicators for each full-time umpire on pollution measures and controls from our preferred specification not specific to umpires.[5] The results for each pollutant are show in Table A.3. We cannot find evidence that pollution treatments are correlated with the identity of the home plate umpire, conditional on observables. In other words we fail to reject the hypothesis that umpires are randomly assigned to pollution condition.

An additional concern is that players may make decisions after pollution levels are realized which lead to a sample of umpire decisions that is systematically different from the typical MLB game. We test for such threats to external validity by examining balance across the universe of pitches. Specifically we construct a sample of all pitches where the umpire would make a ball/strike decision if the batter did not swing, regardless of the actual decision by the batter.[6]

Since umpires are only responsible for adjudicating a ball versus strike when the batter elects not to swing, we first examine whether the batter's decision to swing leads to a selected subsample by testing conditional balance across pollution treatments on the batter's decision to swing. The results of these tests for each local criteria pollutant are shown in Table A.4. In each case, we fail to reject the null of conditional random assignment of swinging to pollution levels.

Second, we consider whether decisions faced by MLB umpires are more or less difficult when pollution levels are high. As demonstrated in Table A.5, umpires have higher unconditional error rates for pitches traveling close to the strike zone boundary. Table A.6 summarizes results from a

[4] We impose cross-model equality restrictions on all model covariates (X).

[5] Specifically we exclude umpire fixed effects, umpire experience trends, and umpire-specific strike zone controls from the vector of regression controls.

[6] Specifically the dataset includes all pitches with the exception of pitches as part of an intentional walk, batter hit-by-pitch, any bunt or attempted bunt, pitchouts, and "automatic" balls and strikes (which are called when there are rules violations during the pitch).

test of whether the absolute distance from the pitch location to the boundary of the strike zone is balanced across pollution levels. Again, we fail to find evidence that umpires face more or less difficult decisions on polluted days.[7]

A.3 Robustness

A.3.1 Effects by decision difficulty

To investigate whether the effects of air quality on decision making differ with the difficulty of the decision, in Table A.7 we re-estimate our preferred specification on the subsample of three inch bands of varying distances from the strike zone boundary. The effect is larger for CO for pitches within 6 inches of the strike zone. Effects for $PM_{2.5}$ are generally poorly estimated. The effects are attenuated for both pollutants in the 6-9 inch band.

A.3.2 Player controls

We implement a number of additional regressions including fixed effects for the identity of the players involved with each event where an umpire makes a decision. These players are the pitcher, the batter, and the catcher.[8] Estimates from these regressions are shown in Table A.8. Column (1) repeats the primary specification. Column (2) adds fixed effects for the pitcher and the batter. Column (3) adds fixed effects for the catcher to the primary specification. Column (4) uses pitcher and catcher fixed effects. Finally, Column (5) includes fixed effects for the batter, pitcher, and catcher. In each case, point estimates sustain.

A.3.3 Alternative control variables

Table A.9 presents estimates of alternatives to our primary specification. One may be concerned that effects of the sun on umpires are likely adequately captured by the month and hour of day

[7]Using the same methods, we also find other factors determined after the assignment of umpires to games (and potentially after pollution levels are observed), such as the type and velocity of each pitch, do not predict pollution levels. We do not report these results here.

[8]There are cases where the PITCHf/x lack sufficient detail to determine the identity of the catcher particularly if there are player substitutions during the game. We exclude observations where the catcher's identity is ambiguous from models using catcher fixed effects.

fixed effects therefore Column (2) excludes controls for the position of the sun. It is also likely that umpires are more accurate when they have more time to rest between pitches. As such Column (3) includes a continuous measure of the time elapsed since the previous pitch. Column (4) includes venue-specific cubic time trends as non-linear time controls. In Table A.9 we also present results using alternative methods of assigning pollution to venues. Column (5) repeats our preferred specification using measures from closest stations within 5 mile radius.

References

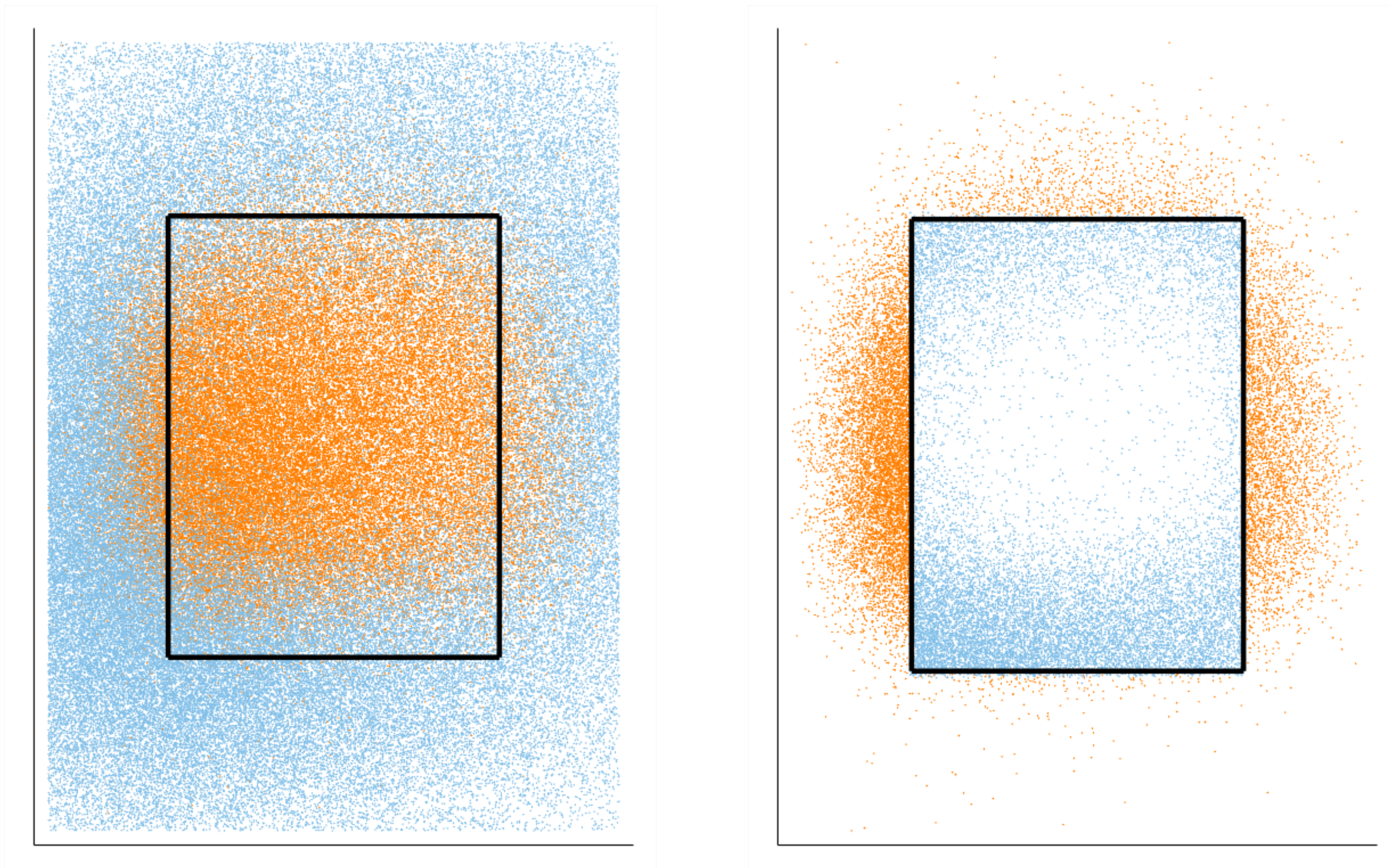
Cameron, A Colin and Douglas L Miller (2015). "A Practitioner's Guide to Cluster- Robust Inference." In: *Journal of Human Resouces* 50(2), pp. 317–372.

Rapson, David and Katrina Jessoe (2014). "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use." In: *The American Economic Review* 104(4), pp. 1417–1438. DOI: 10.1257/aer.104.4.1417.

Figure A.1: Called balls and strikes (from 2008 season)

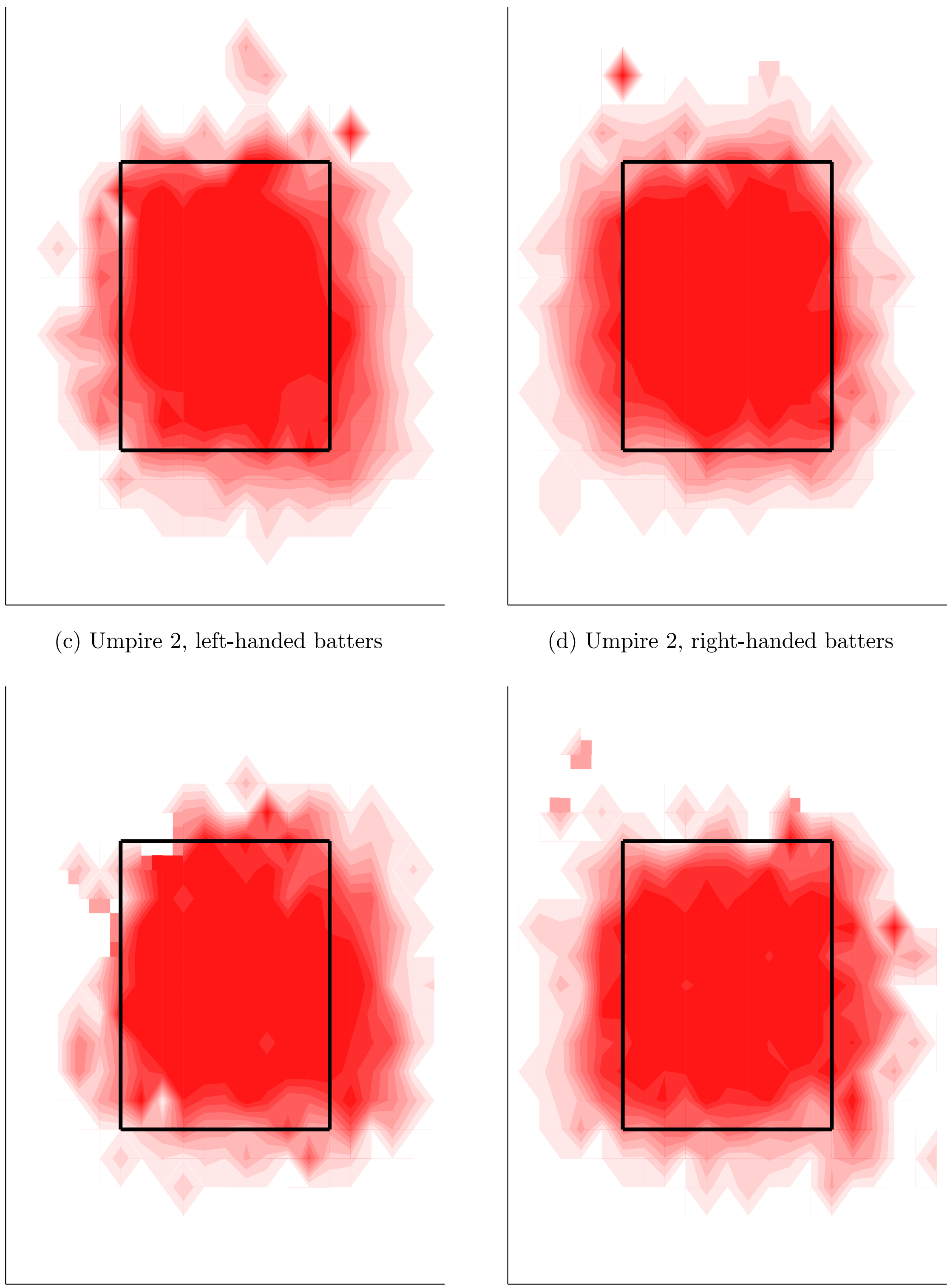
(a) All pitches requiring a decision

(b) Incorrect decisions



Scatter plot of all pitches thrown in 2008 season from the perspective of the pitcher. Orange dots denote called strikes, blue dots denote called balls. Vertical black lines represent the boundaries of the strike zone. Orange dots outside the lines and blue dots inside the lines are incorrect calls. Spherical noise equivalent to 0.1% of the graph size added to the position of each point.

Figure A.2: Umpire-specific strike zones



(a) Umpire 1, left-handed batters

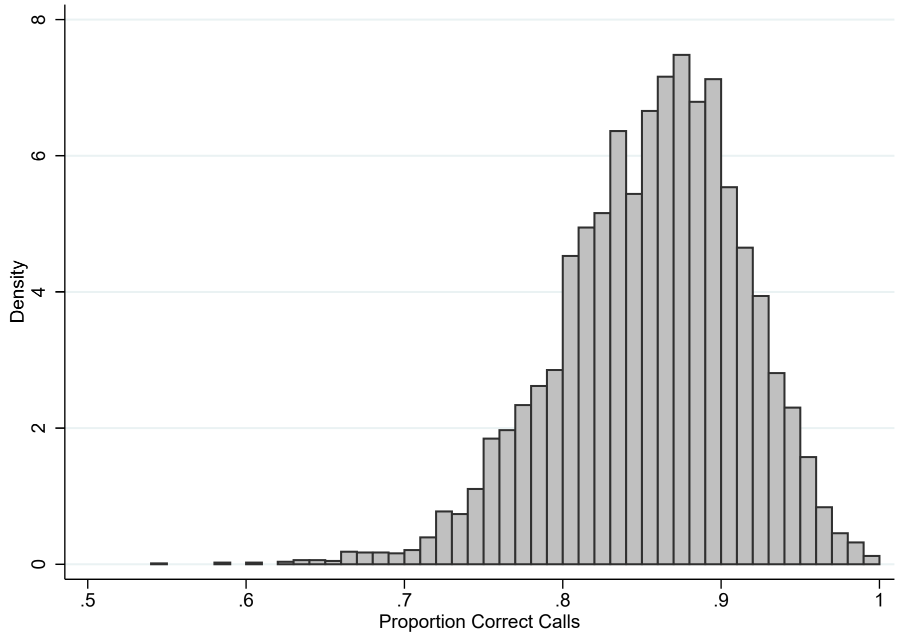
(b) Umpire 1, right-handed batters

(c) Umpire 2, left-handed batters

(d) Umpire 2, right-handed batters

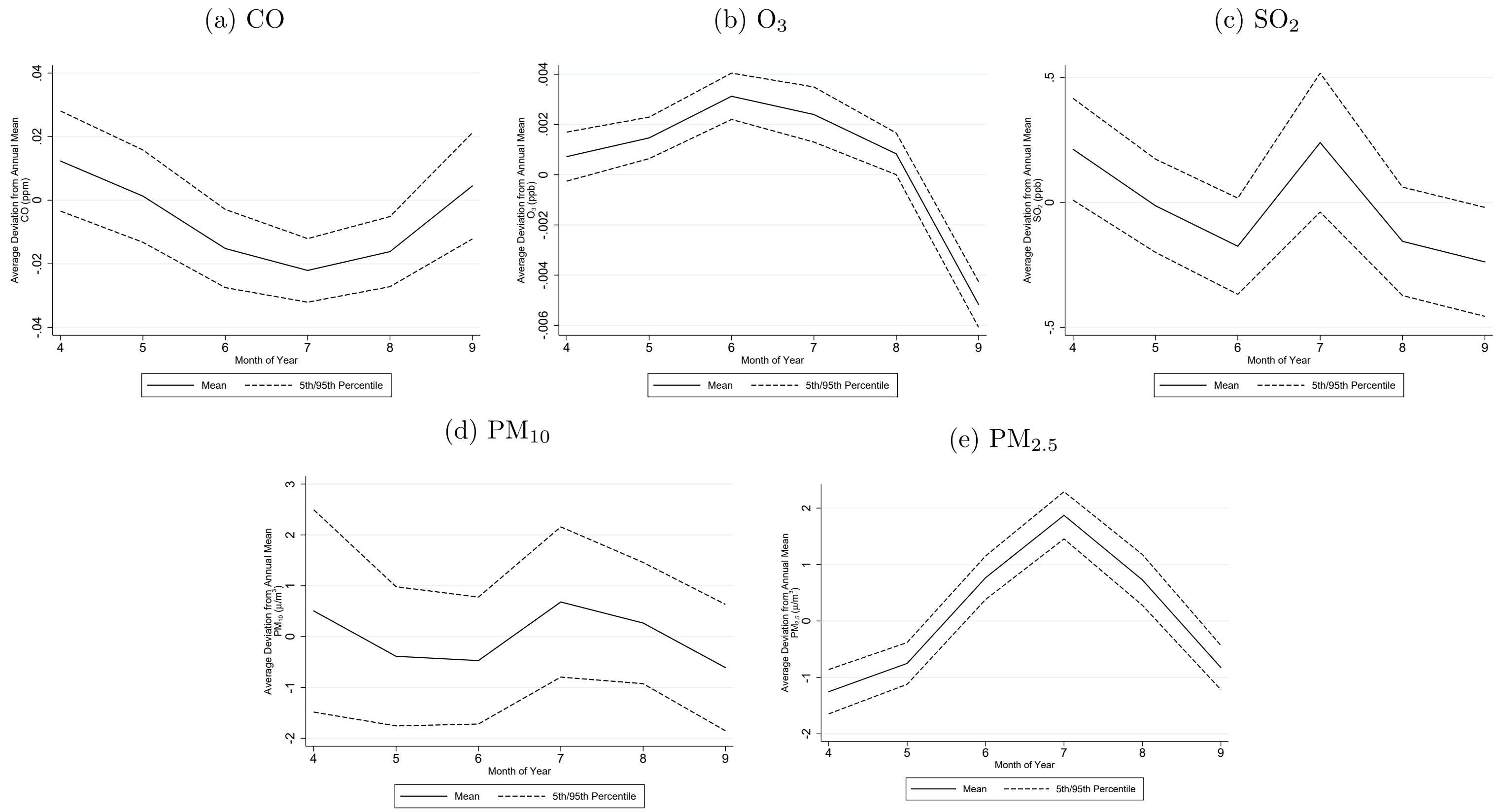
This figure plots probability of umpire calling a strike given the pitch position. The ID of Umpire 1 is 427044 and umpire 2 is 427261. Darker colors indicate a higher probability of calling a strike.

Figure A.3: Distribution of mean of correct call rate



Histogram of the percentage of correct calls per game. This plot includes only games with available pollution data.

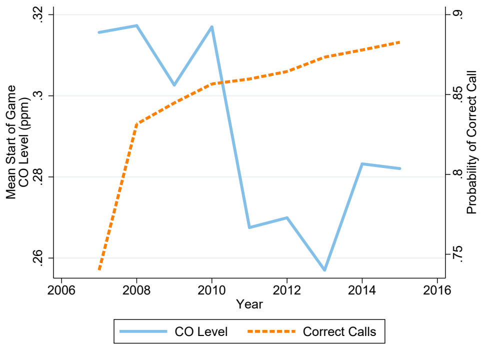
Figure A.4: Monthly deviations from annual mean pollution levels



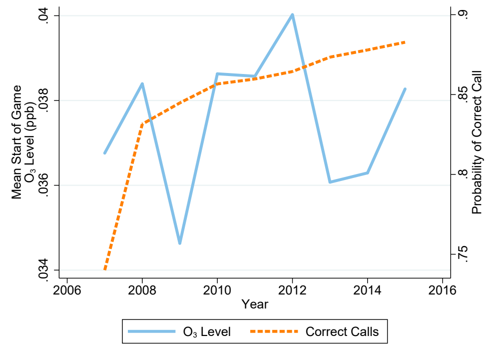
Mean monthly deviations of pollution levels from the venue-specific annual mean. Dotted lines represent the 5th and 95th percentile deviations.

Figure A.5: Annual trends in decision accuracy and pollution levels

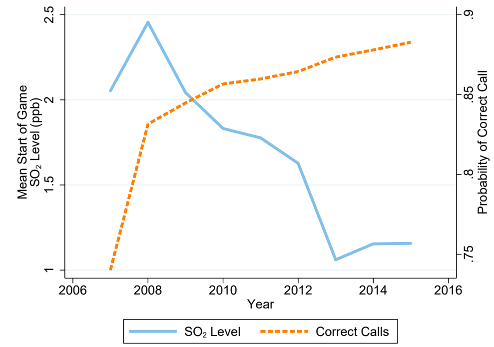
(a) CO



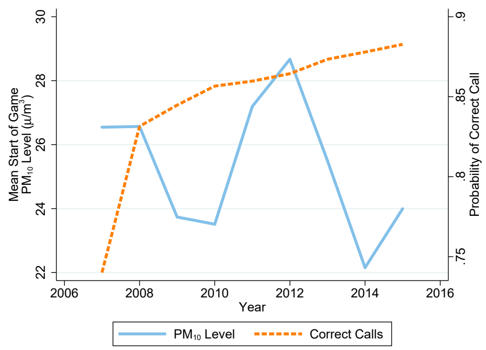
(b) O_3



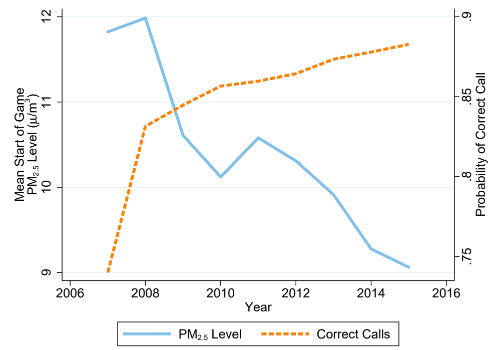
(c) SO_2



(d) PM_{10}



(e) $PM_{2.5}$



Annual mean pollution level and mean probability of correct call by umpires for all pitches where they are required to make a ball/strike decision.

Table A.1: List of venues

	Number of games
AT&T Park	284
Angel Stadium of Anaheim	303
Busch Stadium	292
Chase Field	126
Citi Field	265
Citizens Bank Park	308
Comerica Park	308
Coors Field	295
Dodger Stadium	284
Fenway Park	270
Great American Ball Park	286
Kauffman Stadium	284
Marlins Park	22
Miller Park	180
Minute Maid Park	59
Nationals Park	284
O.co Coliseum	293
Oriole Park at Camden Yards	274
PNC Park	268
Petco Park	288
Progressive Field	297
Rangers Ballpark in Arlington	126
Safeco Field	234
Shea Stadium	34
Sun Life Stadium	153
Target Field	229
Turner Field	284
U.S. Cellular Field	278
Wrigley Field	280
Yankee Stadium	297

Table A.2: Summary Statistics by Sample Restriction

	Mean	Std. Dev.
Correct call	0.827	0.378
Pitch in strike zone	0.541	0.498
Game indoors	0.135	0.342
Attendance	30,977	10,689
Pitch speed (mph)	87.82	6.00
Outdoor temperature (F)	72.31	11.84
Relative humidity (%)	59.07	18.55
Wind speed (mph)	7.567	5.125
Outdoor air pressure (inHg)	29.52	0.74
CO (ppm)	0.295	0.139
$PM_{2.5}$ (10 $\mu g/m^3$)	1.09	0.58
Ozone (ppm)	0.034	0.015
Observations	623,573	
Number of games	12,543	
Number of venues	29	
Number of umpires	86	

Summary statistics for key variables with standard deviations in parentheses. Column (1) show means for all all pitches from 2007 through 2015 where the umpire was required to make a ball/strike call. Column (2) imposes the sample restrictions described in Section 5 including eliminating games where using umpires with less than two full seasons of experience, games where the PITCHf/x equipment appears to be miscalibrated, and pitches where the top or bottom of the strike zone was the operative boundary. Column (3) further limits the sample to observations with non-missing pollution values. Column (4) represents the estimation sample for the primary regression specification, excluding observations with any missing covariates from that specification.

Table A.3: Conditional balance of full-time MLB umpires on observables

	CO	$PM_{2.5}$	O_3
χ^2 Statistic	89.98	67.96	95.85
χ^2 DoF	85	85	85
P-value	0.3351	0.9121	0.1978

χ^2-statistics and p-values represent a joint significance of pollution values from regressions of indicators identifying the home plate umpire for a given game on pollution values and covariates from the primary specification, excluding any umpire-specific effects. Regression excludes umpires who officiate less than two full seasons over the course of the 8 years of data. Statistics robust to arbitrary heteroskedasticity and correlation with game-level clusters.

Table A.4: Conditional effect of pollution levels on swinging

	Coefficient	Std. error	P-value	Number of pitches	Number of clusters
Carbon monoxide	-0.00235	0.00379	0.535	1,663,679	12,543
Ozone	0.04479	0.03960	0.258	1,663,679	12,543
$PM_{2.5}$	0.00132	0.00096	0.168	1,663,679	12,543

Coefficients and p-values from a regression of an indicator of whether the batter elects to swing on pollution level and covariates from the preferred specification. The sample includes all pitches where the umpire would make a ball/strike decision if the batter did not swing. Regression excludes umpires who officiate less than two full seasons over the course of the 8 years of data. Standard errors clustered at the game level, p-values computed using the finite cluster adjustment suggested in Cameron and Miller (2015).

Table A.5: Unconditional rate of correct calls

Distance to Boundary	Correct Call Rate
$[0, 1)$ inches	0.549
$[1, 2)$ inches	0.645
$[2, 3)$ inches	0.743
$[3, 4)$ inches	0.824
$[4, 5)$ inches	0.889
$[5, 6)$ inches	0.938
$[6, 7)$ inches	0.967
$[7, 8)$ inches	0.984
$[8, 9)$ inches	0.992

The unconditional rate of correct calls by umpires by the distance of the pitch from the strike zone boundary. Pitches closer to the boundary are more difficult to adjudicate. Rates computed using all pitches in the primary estimation sample.

Table A.6: Conditional effect of pollution levels on pitch location

	Coefficient	Std. Error	p-Value	N Pitches	N Clusters
Carbon Monoxide	-0.00099	0.00382	0.795	731,709	12,543
Ozone	0.05508	0.04084	0.178	731,709	12,543
$PM_{2.5}$	0.00028	0.00102	0.788	731,709	12,543

Coefficients and p-values from a regression of the absolute distance of a pitch to the strike zone boundary on pollution levels and covariates from the preferred specification. The sample includes all pitches where the umpire would make a ball/strike decision if the batter did not swing. Regression excludes umpires who officiate less than two full seasons over the course of the 8 years of data. Standard errors clustered at the game level, p-values computed using the finite cluster adjustment suggested in Cameron and Miller (2015).

Table A.7: Effect sizes by decision difficulty

	Preferred (1)	Pitch within 0 in - 3 in (2)	Pitch within 3 in - 6 in (3)	Pitch within 6 in - 9 in (4)
CO (>0.5 ppm)	-0.020	-0.040	-0.023	-0.015
	(0.008)**	(0.016)**	(0.011)**	(0.006)**
$PM_{2.5}$ (10 μg/m^3)	-0.004	-0.003	-0.005	0.000
	(0.001)***	(0.002)	(0.002)***	(0.001)
Ozone (ppm)	0.029	-0.031	-0.014	0.002
	(0.048)	(0.087)	(0.072)	(0.037)
N obs	623,573	223,710	210,893	188,958
N clusters	12,543	12,542	12,543	12,541
N venues	29	29	29	29
Pitch location window	All	$[0, 3)$ in	$[3, 6)$ in	$[6, 9)$ in

Linear probability model. Dependent variable is probability of correct call. Standard errors clustered at the game level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include venue, time fixed effects and controls for weather, pitch characteristics, game situation and umpire. See notes in Table 2 for a full description of controls.

Table A.8: Player FEs

	Preferred (1)	Batter & pitcher FEs (2)	Catcher FEs (3)	Pitcher & catcher FEs (4)	Batter, pitcher & catcher FEs (5)
CO (>0.5 ppm)	-0.020	-0.022	-0.019	-0.021	-0.022
	(0.008)**	(0.008)***	(0.008)**	(0.008)***	(0.008)***
$PM_{2.5}$ (10 $\mu g/m^3$)	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Ozone (ppm)	0.029	0.038	0.038	0.048	0.045
	(0.048)	(0.047)	(0.047)	(0.047)	(0.047)
N obs	623,573	623,380	623,379	623,358	623,186
N clusters	12,543	12,543	12,543	12,543	12,543
N venues	29	29	29	29	29
Base controls	Preferred	Preferred	Preferred	Preferred	Preferred
Batter FEs	N	Y	N	N	Y
Pitcher FEs	N	Y	N	Y	Y
Catcher FEs	N	N	Y	Y	Y

Linear probability model. Dependent variable is probability of correct call. Standard errors clustered at the game level shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All regressions include venue, time fixed effects and controls for weather, pitch characteristics, game situation and umpire. See notes in Table 2 for a full description of controls. Column (1) repeats the primary specification. Column (2) adds fixed effects for the pitcher and the batter. Column (3) adds fixed effects for the catcher to the primary specification. Column (4) uses pitcher and catcher fixed effects. Finally, Column (5) includes fixed effects for the batter, pitcher, and catcher. Columns (3) to (5) exclude observations where the identity of the catcher is ambiguous in the PITCHf/x data.

Table A.9: Alternative control variables

	Preferred (1)	Position of the sun (2)	Time since last pitch (3)	Non-linear time ctrl (4)	Closest monitor within 5 miles (5)
CO (>0.5 ppm)	-0.020	-0.020	-0.021	-0.017	-0.018
	(0.008)**	(0.008)**	(0.008)**	(0.007)**	(0.008)**
$PM_{2.5}$ (10 μg/m^3)	-0.004	-0.004	-0.004	-0.003	-0.002
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)
Ozone (ppm)	0.029	0.029	0.031	-0.038	-0.017
	(0.048)	(0.048)	(0.048)	(0.045)	(0.057)
N obs	623,573	623,573	618,122	623,573	446,444
N clusters	12,543	12,543	12,543	12,543	8,981
N venues	29	29	29	29	23
Sun position	Y	N	N	N	N
Time to last pitch	N	N	Y	N	N
Non-linear time controls	N	N	N	Y	N

Linear probability model. Dependent variable is probability of correct call. Standard errors clustered at the game level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include venue, time fixed effects and controls for weather, pitch characteristics, game situation and umpire. See notes in Table 2 for a full description of controls. Column (1) repeats the preferred specification. Column (2) excludes controls from the position of the Sun relative to the Umpire's field of view. Column (3) includes the time elapsed since the previous pitch. Column (4) replaces venue-by-month fixed effects with venue-specific cubic time trends. Columns (5) uses closest monitors within 5 miles of venues.